

01170<sup>3</sup>



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

**DIVISIÓN DE ESTUDIOS DE POSGRADO  
FACULTAD DE INGENIERÍA**

**“SISTEMAS DE INFORMACIÓN ESTRATÉGICA Y  
ALMACENES DE DATOS”**

**T E S I S  
QUE PARA OBTENER EL GRADO DE  
MAESTRA EN INGENIERÍA  
(ELÉCTRICA)**

**P R E S E N T A :  
MARÍA ISABEL GARRIDO GALINDO**

**DIRECTOR:  
DR. JOSÉ ABEL HERRERA CAMACHO**



MÉXICO, D.F.

**TESIS CON  
FALLA DE ORIGEN**

2003

A



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.

## DEDICATORIAS

Al Señor dedico mis obras y Él consolida mis proyectos

Proverbios 16:3

A la División de Estudios de Posgrado de la Facultad de Ingeniería de la UNAM.

Por el orgullo que siento de formar parte de ella y por todo lo que le debo.

A la Dirección General de Asuntos del Personal Académico de la UNAM (DGAPA) y a TELMEX.

Por su invaluable apoyo para la realización de mis estudios de maestría.

A los Doctores Abel Herrera, Ricardo Aceves y Servio Tulio Guillén.

Por toda su comprensión y apoyo para la realización de este trabajo. Gracias.

A Iliana.

Por ser una excelente compañera.

Por tu apoyo y cariño.

Por compartir conmigo tu saludable visión de las cosas.

Por tu invaluable ayuda en el logro de este sueño.

Gracias por todo.

A mi familia.

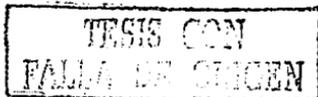
Por todo su amor con todo mi amor.

Al Doc y a Alexandraa Stuttgart.

Gracias por todo su cariño y atención en los momentos difíciles y por enseñarme que con voluntad todo es posible. Kunstwollen.

A mis amigos.

Por su amor y apoyo.



B

## INDICE.

INTRODUCCIÓN.	1
PROBLEMÁTICA.	3
OBJETIVOS.	3
DESCRIPCIÓN DEL TRABAJO DE TESIS.	4

### CAPÍTULO I. MARCO TEÓRICO.

1.1. Antecedentes.	6
1.2. Sistemas de Información Estratégica Automatizados.	7
1.3. Arquitectura de los Sistemas de Información Estratégica Automatizados.	8
1.4. Los Sistemas Transaccionales como fuentes de información de los SIE.	9
1.5. Procesos automáticos para la adquisición de datos.	12
1.6. Almacenes de Datos y Tecnología OLAP.	15
1.7. Data Mart.	22
1.8. Metadatos.	23
1.9. Minería de Datos.	24

### CAPÍTULO II. HERRAMIENTAS Y METODOLOGÍA PROPUESTAS PARA EL ANÁLISIS, DISEÑO Y DESARROLLO DE ALMACENES DE DATOS Y SISTEMAS DE INFORMACIÓN ESTRATÉGICA.

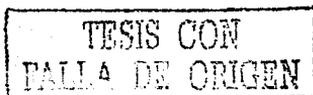
2.1. Herramientas multidimensionales.	38
2.2. Antecedentes de la metodología propuesta.	40
2.3. Diagrama de flujo de datos del ciclo de vida de Aproximaciones Sucesivas para el desarrollo de Sistemas de Información Estratégica.	42
2.4. Descripción de la Metodología de Aproximaciones Sucesivas.	43

### CAPÍTULO III. APLICACIÓN DE LA METODOLOGÍA Y HERRAMIENTAS PROPUESTAS PARA EL ANÁLISIS Y DISEÑO DEL SISTEMA DE INFORMACIÓN ESTRATÉGICA DE LA COMISIÓN FEDERAL DE ELECTRICIDAD.

3.1. Aspectos generales de la Comisión Federal de Electricidad (CFE).	48
3.2. Especificación del modelo de procesos de la CFE.	51
3.3. Especificación de las fuentes de Información con base al modelo de procesos.	51
3.4. Información general de las fuentes de información representativas de cada proceso.	52
3.5. Arquitectura del Sistema de Información Estratégica de la CFE.	52
3.6. Organización de recursos informáticos.	53
3.7. Necesidades de información a nivel ejecutivo.	54
3.8. Frecuencia de actualización de los Sistemas Transaccionales.	55
3.9. Cuentas y Permisos	56

### CAPÍTULO IV. APLICACIÓN DE LA METODOLOGÍA PARA EL DESARROLLO DEL DATA MART DEL ÁREA DE ENERGÉTICOS.

4.1. Recopilación de la información relativa al área de Energéticos.	57
4.2. Análisis de la fuente de información.	57
4.3. Diseño de vistas multidimensionales que respondan a las necesidades	62

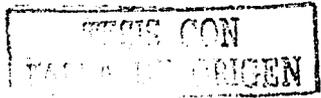


de información corporativa.	.....	63
4.4. Diseño de las estructuras multidimensionales.	.....	66
4.5. Desarrollo de las estructuras multidimensionales.	.....	69
4.6. Implantación de las estructuras relacionales orientadas a la transformación de datos del Data Mart de Energéticos.	.....	70
4.7. Diseño del repositorio relacional (ROLAP).	.....	73
4.8. Diccionario de datos para índices calculados.	.....	75
4.9. Desarrollo del repositorio relacional (ROLAP).	.....	75
4.10. Diseño y desarrollo de programas para la carga y agregación de datos.	.....	79
4.11. Mecanismos para el acopio automático de información, adicionando funciones para la Extracción, Transporte, Transformación e Inserción.	.....	79
4.12. Integración de las aplicaciones ETTI y desarrollo de un tablero de control que permita realizar la gestión y administración centralizada de estos procesos.	.....	80
4.13. Diseño y desarrollo de la aplicación para la explotación de datos del Almacén de Datos de la CFE.	.....	80

**CAPÍTULO V. DESARROLLO DE UNA APLICACIÓN PARA LA MINERÍA DE DATOS.**

5.1. Planteamiento.	.....	85
5.2. Estrategia de desarrollo.	.....	85
5.3. Descripción de elementos involucrados en la aplicación desarrollada.	.....	87
5.4. Análisis de la fuente de datos.	.....	88
5.5. Etapa de preprocesamiento.	.....	93
5.6. Entrenamiento de la neurona lineal adaptiva.	.....	95
5.7. Pruebas a la neurona entrenada.	.....	98
5.8. Presentación de una aplicación para minería de datos.	.....	99

CONCLUSIONES.	.....	103
REFERENCIAS.	.....	104



D

## INTRODUCCIÓN.

Las organizaciones son sistemas que realizan actividades comerciales, industriales y de servicios con o sin fines de lucro.

Estos sistemas están constituidos por combinaciones de recursos humanos, materiales y económicos, coordinados por una autoridad que toma decisiones encaminadas a la consecución de los fines para los que fueron creados.

Para que estos sistemas puedan lograr sus fines u objetivos, se deben encontrar los caminos o medios para alcanzarlos; para encontrar dichos medios se requiere que un especialista (o equipo de especialistas) considere soluciones posibles y elija las que prometen optimización, con máxima eficiencia y mínimo costo en una red de interacciones compleja.

Para encontrar soluciones óptimas, los especialistas deben basarse en su percepción de todos los elementos que interactúan en una determinada situación, es decir, en el conocimiento de la realidad. Dicho conocimiento se logra a partir del análisis de información sustancial; de ahí la importancia y carácter estratégico de la información.

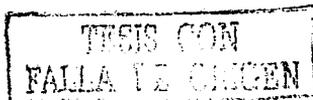
La obtención de información sustancial y estratégica, involucra procesos de gran complejidad. La información es el resultado del esfuerzo cooperativo de diversos elementos. Al conjunto de elementos que tienen como objetivo proporcionar la información sustancial y estratégica se le conoce como Sistema de Información Estratégica.

Los Sistemas de Información Estratégica tienen un solo objetivo: informar. Informar es la presentación periódica de informes para enterar o dar conocimiento. La información, al igual que la materia y la energía, constituye un recurso fundamental, es decir, un bien en el sentido económico del término.

Para que los sistemas perduren, deben tener una gran capacidad de evolución y adaptación; particularmente, los Sistemas de Información Estratégica deben contar con estas capacidades para tener la posibilidad de cumplir el objetivo de informar permanentemente; por ello, generalmente es posible establecer el inicio de estos sistemas pero nunca su fin y en todo caso, debe considerarse una inversión permanente de recursos para su mantenimiento, ya que de otro modo, la consecuencia será el envejecimiento del sistema, lo que derivará en su inutilidad.

Los Sistemas de Información Estratégica están formados por subunidades con ciertas condiciones de frontera, entre las cuales se dan procesos de transporte, de distinta índole; sin embargo, intentar comprender el funcionamiento integral de estos sistemas a partir del estudio de cada una de las partes que lo constituyen, no es la estrategia más adecuada. Es necesario estudiar no sólo partes y procesos aislados, sino también es necesario entender los problemas decisivos hallados en el orden que los unifica, resultantes de la interacción dinámica de partes, lo que provoca un comportamiento diferente de éstas cuando se estudian aisladas o dentro del todo; lo más importante en este caso, es la operación del sistema integral.

A partir de la determinación de las fuentes de información adecuadas para conocer la situación real de los sistemas de las que proceden, los Sistemas de Información Estratégica inician su función con la obtención de información sustancial y, posteriormente, se estructuran informes; sin embargo, tener la posibilidad de almacenar esta información nos permite conservar la memoria de la situación del sistema en distintos periodos de tiempo, esta información histórica puede servir de base para el entendimiento de los factores que determinan ciertos comportamientos del sistema, lo cual permitirá bajo ciertas consideraciones, anticiparse a ciertos eventos.



Los Sistemas de Información Estratégica básicamente se componen de los siguientes subsistemas:

- Fuentes de información internas y externas a las organizaciones.
- Subunidad para la adquisición de información sustancial y estratégica.
- Subunidad para el almacenamiento de información histórica.
- Subunidad para la presentación de informes.
- Subunidad para la búsqueda de patrones ocultos, tendencias y correlaciones.

Debido a la complejidad de estos sistemas y a la gran cantidad de recursos que deben invertirse para su creación y mantenimiento, se debe contar con una adecuada metodología de desarrollo, mantenimiento y actualización.

Una metodología es un conjunto de procedimientos definidos para lograr ciertos objetivos; específicamente, la metodología para la creación de Sistemas de Información Estratégica, debe considerar dos aspectos críticos: primero, la creación de estos sistemas depende del logro de una gran cantidad de objetivos de naturaleza diversa; segundo, debido a que estos sistemas constantemente evolucionan y se adaptan a nuevas circunstancias, su creación no tiene fin.

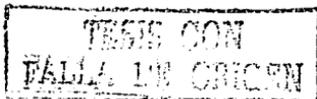
Las metodologías existentes definen un conjunto de procedimientos que abordan la creación de sólo ciertas subunidades de manera independiente; por ejemplo, la metodología de Inmon <sup>1</sup> es específica para la creación de almacenes de datos, las metodologías propuestas por Boehm <sup>2</sup>, Yourdon <sup>3</sup>, Martin y Board <sup>4</sup> definen una serie de procedimientos para la automatización de funciones; en todo caso, las metodologías que proponen la creación aislada de cada subunidad del Sistema de Información Estratégica, para posteriormente proceder a la integración de las mismas, no son efectivas en términos prácticos, ya que esta integración generalmente resulta muy compleja.

La tendencia contemporánea para la creación de Sistemas de Información Estratégica debe ser orientada a la concepción integral y automatizada de estos sistemas sin perder de vista la complejidad que reside en cada subsistema y en su interacción dinámica.

En suma, se puede comprobar la necesidad de nuevas metodologías para la creación de Sistemas de Información Estratégica a partir de la tendencia planteada.

---

1. Inmon: Autor de la metodología para el desarrollo de almacenes de datos (Data Warehouse).  
2. Barry Boehm: Autor del ciclo de vida de cascada.  
3. Edward Yourdon: Autor del ciclo de vida estructurado moderno.  
4. James Martin y Bernard Board: Autores del ciclo de vida de prototipos.



## **PROBLEMÁTICA.**

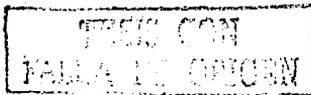
Se carece de metodologías efectivas e integrales para la creación de Sistemas de Información Estratégica que promuevan la automatización de los distintos subsistemas que los componen.

## **OBJETIVO GENERAL.**

Proponer una metodología integral para el análisis, diseño y desarrollo de Sistemas de Información Estratégica automatizados y, posteriormente, probar su efectividad a partir de su aplicación en un caso práctico.

## **OBJETIVOS PARTICULARES.**

- Proponer una arquitectura integral para los Sistemas de Información Estratégica Automatizados.
- Proponer una metodología para la creación de Sistemas de Información Estratégica Automatizados.
- Proponer un conjunto de herramientas para el diseño de almacenes de datos.
- Probar la efectividad de la metodología y herramientas propuestas a partir de su aplicación en un caso práctico.
- Desarrollar una aplicación para la explotación de los datos de un subsistema en particular.
- Desarrollar un sistema para la minería de datos que extraiga información significativa de un subsistema en particular.



## ORGANIZACIÓN DEL TRABAJO DE TESIS

### Descripción del capítulo I.

En este capítulo se abordan los aspectos relativos a los Sistemas de Información Estratégica Automatizados, sus características principales, sus ventajas, así como la presentación de una arquitectura propuesta para la creación de estos sistemas.

En vista de que los sistemas transaccionales constituyen las fuentes de información de los Sistemas de Información Estratégica, se analizan sus características y potencialidades, y se describen los elementos relevantes de la tecnología con la que se desarrollan y, finalmente, se aborda el aspecto de las bases de datos transaccionales típicamente desarrolladas con tecnología relacional.

Se presenta una descripción de los procesos involucrados en la adquisición periódica de datos, tales como: la extracción, el filtrado, el transporte, la limpieza, la transformación y la inserción de datos en el Almacén de Datos.

Se describen los elementos más significativos de los Almacenes de Datos y la tecnología OLAP, sus características y funcionalidades. Se analizan las arquitecturas ROLAP y MOLAP, así como sus modelos de datos.

Se aborda el tema de los Data Mart, los cuales se diseñan para satisfacer las necesidades específicas de grupos comunes como subconjuntos del Almacén de Datos.

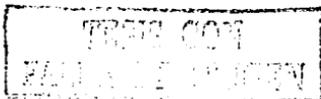
A partir de la concentración masiva de datos en el Almacén, surge la necesidad de administrar y controlar su contenido, y ésto se realiza mediante los metadatos, los cuales también se analizan en este capítulo.

Finalmente, se analizan algunas de las técnicas más frecuentemente empleadas en la minería de datos.

### Descripción del capítulo II.

En este capítulo se presenta la metodología de Aproximaciones Sucesivas propuesta para el análisis, diseño y desarrollo de Sistemas de Información Estratégica, así como las herramientas para el diseño de Almacenes de Datos.

Este es posiblemente el capítulo central de este trabajo de tesis, ya que constituye el resumen de una serie de experiencias en el desarrollo de sistemas de este estilo.



#### Descripción del capítulo III.

El objetivo de este capítulo consiste en presentar un caso práctico en donde se aplica la metodología propuesta.

El caso práctico presentado en este trabajo de tesis se refiere al Sistema de Información Estratégica de la Comisión Federal de Electricidad (CFE).

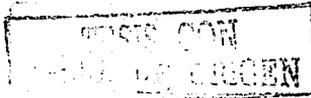
En vista de la complejidad que involucra este sistema, sólo se presenta el análisis y las especificaciones generales del Sistema de Información Estratégica de la CFE.

#### Descripción del capítulo IV.

Reportar en este trabajo el desarrollo completo del Sistema de Información Estratégica de la CFE es un asunto muy complicado ya que se requerirían varios tomos especializados en las diferentes áreas, por lo que en este capítulo se reporta el desarrollo del subsistema (Data Mart) de Energéticos, con el fin de ejemplificar la aplicación de la metodología propuesta y probar su efectividad.

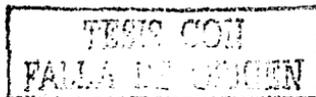
#### Descripción del capítulo V.

En este capítulo se reporta el resultado del desarrollo de un mecanismo para la minería de datos, mediante el uso de la técnica de destilación de patrones y el entrenamiento de una neurona lineal adaptiva. El problema planteado tiene como objetivo realizar la clasificación de cuatro aspectos de la CFE para la determinación de condiciones suficientes o insuficientes para su correcta operación. Evidentemente, este ejemplo no pretende sustituir la experiencia de los ejecutivos en la toma de decisiones, sino que busca mostrar la aplicabilidad de las modernas técnicas de minería en los Almacenes de Datos.



# CAPITULO I

Marco teórico.



## 1.1. Antecedentes.

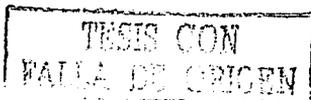
Actualmente casi todas las organizaciones poseen sistemas automatizados que apoyan sus operaciones cotidianas, a los que se les denomina Sistemas Transaccionales, los cuales, se desarrollan con base a la tecnología OLTP (*On Line Transaction Processing*, Procesamiento Transaccional en Línea).

Generalmente, los Sistemas Transaccionales guardan datos derivados de procesos, actividades, clientes, etc., en bases de datos que crecen de manera acelerada. Los datos se almacenan con el fin de hacer un análisis posterior de los mismos; sin embargo, es común apreciar que dichas organizaciones prácticamente se encuentran inundadas en una gran cantidad de datos desaprovechados.

Las organizaciones que suelen analizar los datos almacenados por Sistemas Transaccionales, regularmente lo hacen de manera superficial y bajo la influencia de los resultados que esperan encontrar. Es común el uso periódico de alguna hoja de cálculo o software estadístico para localizar correlaciones entre variables, establecer metas e intentar modelar de esta manera la información.

El objetivo de obtener información a partir de los diferentes sistemas transaccionales de una organización, es el de proporcionar a sus directivos los elementos objetivos para tomar decisiones estructuradas, las cuales están estrechamente ligadas a la planeación. Al conjunto de elementos que obtienen información a partir de los datos almacenados por los Sistemas Transaccionales se les denomina Sistemas de Soporte a las Decisiones (DSS).

Sin embargo, es seguro que los directivos se encuentren en la necesidad de tomar decisiones no estructuradas, también conocidas como estratégicas, las cuales no son rutinarias y no permiten ser planeadas o no se conoce de antemano la información que se requiere analizar para la toma de decisiones. El tipo de información que se requiere en este caso, es de tipo estratégica, la cual se obtiene a partir de Sistemas de Información Estratégica (SIE); los cuales se desarrollan con base a la tecnología OLAP (*On Line Analytical Processing*, Procesamiento Analítico en Línea).



## 1.2. Sistemas de Información Estratégica Automatizados.

La información estratégica es fundamental para conducir a las organizaciones por rumbos exitosos; sin embargo, no se otorga la importancia necesaria a las actividades y recursos que se necesitan para obtenerla; generalmente, esta responsabilidad se asigna a ciertos grupos de empleados que conocen a detalle las actividades, terminología, ubicación e interpretación de datos.

En primera instancia, los empleados asignados buscan información en distintas bases de datos internas y externas a la organización, posteriormente, extraen, concentran y procesan los datos obtenidos en hojas de cálculo y software estadístico y, finalmente, obtienen la información requerida, lo que implica que exista una distracción del tiempo destinado a sus obligaciones cotidianas y el empleo de una gran cantidad de recursos, ya que no es posible anticipar el tipo de información que necesitarán los directivos para sustentar decisiones de tipo estratégicas.

Es evidente la necesidad de automatizar todas las actividades mencionadas, para satisfacer los requerimientos de información en los tiempos y formas requeridos y para que el capital intelectual sea invertido en el análisis de la información y en la toma de decisiones adecuadas y no en la obtención de la información.

La automatización inicia a partir de la selección, extracción, transporte y procesamiento de datos estratégicos contenidos en fuentes internas y externas a las organizaciones para depositarlos, posteriormente, en repositorios cuya organización y estructura facilitan su análisis; a este tipo de estructuras de datos se les conoce como cubos de información y a los repositorios que contienen dichas estructuras de datos se les denomina Almacenes de Datos (Data Warehouse, DWH).

En vista de que el almacén concentra datos automáticamente y de manera periódica, su volumen crece de manera vertiginosa, por lo que es necesario desarrollar elementos automáticos que tengan la posibilidad de introducirse en el repositorio de datos para identificar, descubrir, clasificar, agrupar, presentar y generar información estratégica.

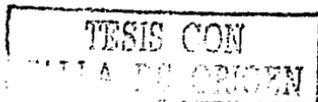
En suma, los Sistemas de Información Estratégica (SIE) son conjuntos de componentes automatizados e interrelacionados que permiten capturar, procesar, almacenar y distribuir información estratégica para apoyar la toma de decisiones y el control de gestión de una organización.

Características principales de los SIE:

- Suelen introducirse después de que se implantan los Sistemas Transaccionales más revelantes de la organización, ya que estos últimos constituyen su plataforma de información.
- Suelen ser intensivos en cálculos y escasos en entradas y salidas de información.
- La justificación económica para su desarrollo es difícil, debido a que inicialmente no se conocen los alcances del proyecto.
- Son interactivos con altos estándares de diseño gráfico y visual.
- Apoyan la toma de decisiones.

Ventajas que proporcionan los SIE:

- ✓ Capacidad para efectuar análisis específico.
- ✓ Control.
- ✓ Aumento en el número de alternativas examinadas.
- ✓ Mejor entendimiento del sistema.
- ✓ Respuesta rápida a situaciones inesperadas.
- ✓ Proporcionan la información base para la toma de decisiones.



### 1.3. Arquitectura de los Sistemas de Información Estratégica Automatizados.

La arquitectura que se propone en este trabajo de tesis, para la creación de Sistemas de Información Estratégica Automatizados, muestra en términos generales los elementos y subsistemas involucrados en su construcción.

En primera instancia, se encuentran los Sistemas Transaccionales (OLTP) como fuentes de datos derivados de diversos aspectos de las organizaciones, así como fuentes externas que tengan la posibilidad de aportar elementos comparativos o complementarios. Desde el punto de vista de los Sistemas de Información Estratégica, estas fuentes son elementos externos e independientes de sólo lectura.

En segunda instancia, se encuentran los Sistemas Orientados al Análisis (OLAP) cuyo objetivo es el de obtener, procesar, almacenar y presentar los elementos relevantes que son proporcionados por los Sistemas Transaccionales y las fuentes externas, con el fin de obtener información estratégica.

Los Sistemas Orientados al Análisis extraen los datos relevantes de las fuentes internas y externas de las organizaciones; posteriormente, estos datos son procesados y concentrados en un Almacén general de Datos (Data Warehouse), o bien, en repositorios de propósito específico denominados Data Mart.

Finalmente, la información obtenida es presentada en mecanismos que facilitan su análisis, donde generalmente es organizada y almacenada en bases de datos multidimensionales que permiten la creación de variables dimensionadas.

En la figura 1.1 se muestra la arquitectura general sugerida para Sistemas de Información Estratégica, considerando modernas tecnologías de información que abren la posibilidad de integraciones posteriores de nuevas fuentes de datos internas y externas, presentación de datos y manejo de información eficiente, flexibilidad para la creación de nuevas vistas, elementos que facilitan la actualización periódica de la información, posibilidad de generación de nuevas relaciones y cálculos de datos, almacenamiento de datos significativos para conservar el comportamiento de la organización a lo largo del tiempo, almacenamiento de la información mediante el concepto de variables dimensionadas, etc.

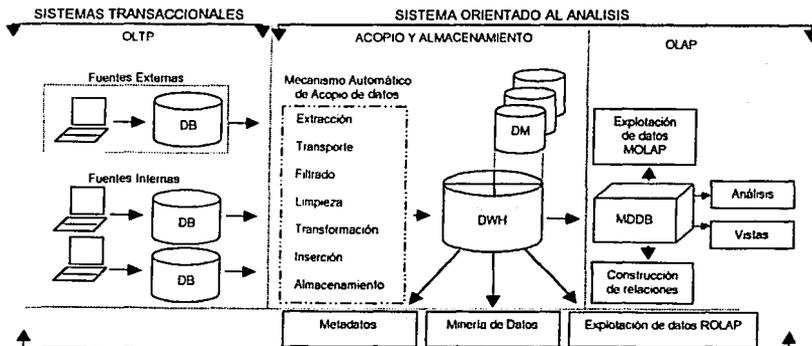


Figura 1.1. Arquitectura general de los Sistemas de Información Estratégica.

#### 1.4. Los Sistemas Transaccionales como fuentes de información de los SIE.

Los sistemas de bases de datos con procesamiento transaccional en línea (OLTP) son óptimos para el manejo de datos que constantemente cambian; generalmente son manejados por un gran número de usuarios que ejecutan transacciones de manera simultánea y que actualizan o modifican los datos en tiempo real. Estos sistemas se diseñan con el fin de promover:

- ✓ Una buena colocación de datos.
- ✓ Transacciones que minimizan los bloqueos generados por la concurrencia.
- ✓ Backup en línea.
- ✓ Alta normalización de bases de datos.
- ✓ Uso cuidadoso de índices.
- ✓ Óptima configuración del hardware.
- ✓ Ausencia de datos históricos o agregados.

La buena colocación de los datos se refiere a los embotellamientos de entrada/salida, que ocurren en los sistemas OLTP debido a que múltiples usuarios modifican simultáneamente datos en las bases de datos. Para resolver este problema, se debe determinar el patrón de acceso a los datos y poner juntos aquellos que se acceden con más frecuencia. Se puede disponer, además, del uso de grupos de archivos y de sistemas RAID (Redundant Array of Independent Disks) para ayudar en este proceso.

Se deben concebir transacciones cortas para minimizar largos períodos de bloqueo y de esa forma mejorar la concurrencia. Siempre que sea posible, se deben ejecutar procesos atómicos (por ejemplo los llamados "stored procedure") para procesar transacciones completas. Por otro lado, el orden en el cual se hace referencia a las tablas dentro de las transacciones, afecta la concurrencia, por lo que se deben poner referencias a las tablas frecuentemente accedidas al final de la transacción, para minimizar la duración del bloqueo de las mismas.

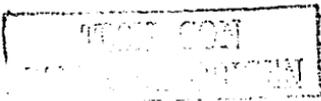
Los sistemas OLTP a menudo son caracterizados por operaciones continuas (24 horas al día, 7 días a la semana) por lo que el tiempo de caída de los sistemas debe ser mínimo. Los backup en línea permiten salvar las bases de datos mientras éstas son usadas.

Generalmente, las bases de datos transaccionales se desarrollan con tecnología relacional, aunque en algunas organizaciones todavía es posible encontrar sistemas que hacen uso de archivos planos o indexados como medio de almacenamiento. Los archivos de texto indexados almacenan los datos sobre archivos interrelacionados por algún elemento particular. El manejo de este tipo de almacenamiento presenta graves dificultades en el proceso de recuperación de datos.

Podemos definir a las bases de datos relacionales como colecciones no redundantes de datos estructurados e interrelacionados de acuerdo con un modelo capaz de recoger el máximo contenido semántico y donde dichos datos son compartidos por diferentes sistemas de aplicación y múltiples usuarios. Las bases de datos se describen y se manipulan apoyándose en un modelo de datos relacional.

Los Sistemas Transaccionales involucran bases de datos que tienden a ser organizadas alrededor de procesos específicos de los cuales sólo se almacenan datos actuales y atómicos. Estas bases de datos relacionales son diseñadas a partir de la consideración de alta normalización, lo que permite reducir la información redundante tanto como sea posible para incrementar la velocidad de actualización y así mejorar la concurrencia.

Los índices deben ser actualizados cada vez que un registro es adicionado o modificado, por lo que se debe evitar la creación y uso indiscriminado de índices sobre aquellas tablas que se actualizan con más frecuencia.



También, se debe contar con una configuración óptima de hardware para manipular un gran número de usuarios concurrentes y agilizar los tiempos de respuesta requeridos por un sistema OLTP.

No es común que los sistemas OLTP utilicen datos históricos o agregados. Los datos que son raramente referenciados pueden ser almacenados en bases de datos separadas, o ponerlos de manera consistente fuera de las tablas actualizadas; o sea, en tablas que contengan sólo datos históricos. Esta reducción de las tablas minimiza el tiempo de backup y la ejecución de consultas.

Por otro lado, los Sistemas Transaccionales generalmente trabajan como unidades independientes que no conservan prácticamente ninguna relación con otros sistemas de una organización ni tampoco suelen compartir datos. La nula relación entre estos sistemas puede ser tal, que pueden incluso ser desarrollados con diferentes tecnologías, conceptos, objetivos y plataformas. Es muy probable la duplicación de datos con consecuencias tales como las de poca confiabilidad, conceptos iguales con denominaciones diferentes, altos costos, etc.

### **Características de las bases de datos relacionales.**

En vista de que las bases de datos transaccionales constituyen la plataforma de información de los Sistemas de Información Estratégica y de que éstas generalmente se desarrollan con tecnología relacional, a continuación se presenta una revisión general de sus principales características.

#### **1. Redundancia controlada.**

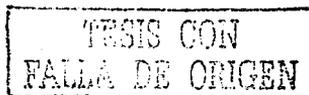
Existen dos tipos de redundancia, la redundancia lógica manejada por los usuarios, y la física, manejada por el sistema. La redundancia lógica de datos debe ser controlada, de forma que no existan duplicidades perjudiciales ni innecesarias, y las redundancias físicas muchas veces son convenientes ya que responden a objetivos de eficiencia; este tipo de redundancias son tratadas por el mismo sistema, de manera que no puedan producirse inconsistencias. En suma, en las bases de datos relacionales no debe existir redundancia lógica, aunque sí se admite cierta redundancia física por motivos de eficiencia; por lo tanto, un dato se actualizará lógicamente por el usuario de forma única, y el sistema se preocupará de cambiar físicamente todos aquellos campos en los que el dato estuviese repetido; a la redundancia física se le conoce como redundancia controlada por el sistema.

#### **2. Múltiples usuarios y diferentes aplicaciones.**

Las bases de datos pretenden ser útiles a toda la organización en su conjunto mediante el manejo de datos como recurso estratégico. Las bases de datos atienden a múltiples usuarios y diferentes aplicaciones, en contraposición a los sistemas que manejan archivos, en los que cada uno de ellos se diseña para responder a las necesidades de una aplicación específica.

#### **3. Independencia.**

Otros aspectos importantes tienen que ver con la independencia, tanto física como lógica, entre los datos y tratamientos. Esta independencia, objetivo fundamental de las bases de datos, es una característica esencial que distingue las bases de datos de los archivos y que ha tenido una enorme influencia sobre los SGBD (Sistema de Gestión de Bases de Datos).



## Tipos de bases de datos transaccionales.

Podemos distinguir tres tipos de bases de datos transaccionales:

- ✓ Base de Datos de producción,
- ✓ Base de Datos de respaldo,
- ✓ Base de Datos para manipulación especial de datos.

Las bases de datos de producción son aquellas que manejan los datos que proceden de la operación real de las organizaciones y que se actualizan según requiera el manejo del proceso específico que proporcione los datos; mientras que las bases de datos de respaldo sirven como un almacén temporal en donde se guardan datos en forma redundante, a fin de asegurar que se preserve una copia de los mismos en caso de que la base de datos de producción presente algún problema. Finalmente, las bases de datos para la manipulación especial de datos, concentran ciertos datos derivados de la operación real a los cuales se requiere realizar alguna transformación, cálculo, procesamiento o preprocesamiento particular; generalmente estas bases de datos no se almacenan en el mismo lugar donde se alojan las bases de datos de producción, ya que se pretende no alterar el desempeño y, por otro lado, se pretende que los datos que residen en la base de datos de producción no sean modificados.

En los datos contenidos en las distintas bases de datos de una organización existen claves ocultas en espera de ser reveladas. La idea es tomar aquellas piezas que sean relevantes y con ellas generar información estratégica. Véase figura 1.2.

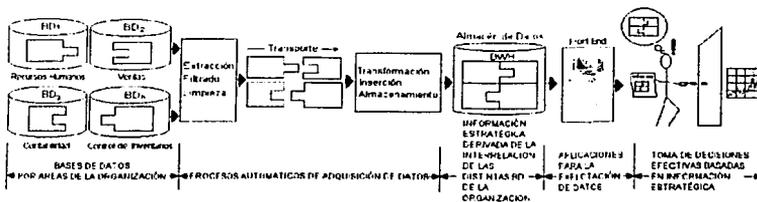


Figura 1.2. Decisiones exitosas con base a información estratégica derivada de datos clave de distintas BD.

TESIS CON  
FALLA DE ORIGEN

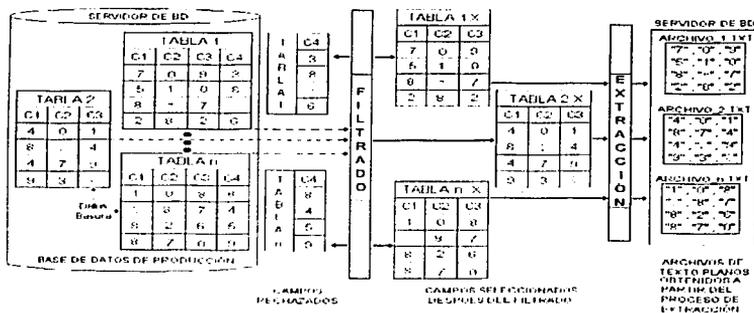
### 1.5. Procesos automáticos para la adquisición de datos.

Las bases de datos de producción contienen datos actualizados, ya que generalmente operan en línea; lo que significa que los datos son modificados, borrados, actualizados, etc., según se efectúe en el proceso específico que proporciona los datos.

Los servidores que contienen las bases de datos de producción, realizan procesos complejos para consulta, modificación, borrado, inserción, actualización y ordenamiento de datos, por ello, generalmente estos equipos tienen como función exclusiva el manejo de las bases de datos que contiene, por lo que cualquier tarea adicional repercutirá directamente en su desempeño.

En vista de que los datos para el Almacén (DWH), generalmente se toman de bases de datos de producción y de que se recomienda no realizar ninguna operación adicional sobre los servidores que los contienen, es necesario seleccionar, mediante reglas, los datos relevantes y críticos que servirán de base para generar información estratégica, a este proceso se le conoce como filtrado de datos. Los datos seleccionados se extraen de la base de datos de producción y, generalmente, se depositan en archivos de texto planos que contienen los campos seleccionados separados por algún carácter, por ejemplo el "pipe"; a este proceso se le conoce como extracción. Véase la figura 1.3.

Hay varias alternativas de extracción que balancean el desempeño y restricciones de tiempo de almacenamiento. El desempeño puede caer si las transacciones en línea compiten con la extracción. Una solución alternativa es crear una vista, desde la cual extraer los datos. El inconveniente aquí, es el espacio de disco adicional para guardar esa copia de la base de datos. Debe asegurarse que no se actualicen los datos mientras se hace la extracción para no generar inconsistencias.



1.3. Ilustración de los procesos de Filtrado y Extracción de datos.

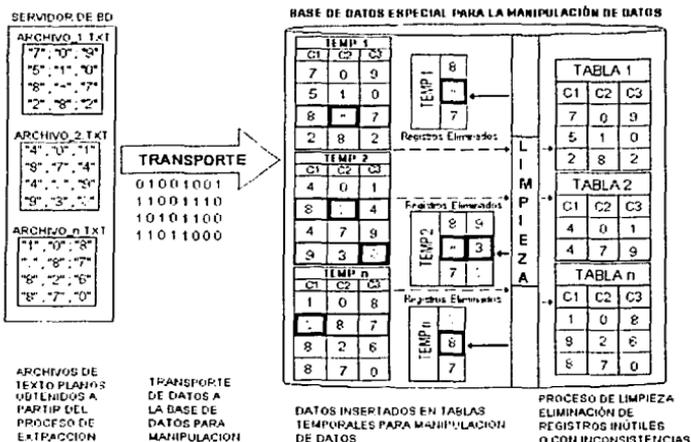
TESIS CON  
DE ORIGEN

Generalmente, los datos que provienen de los sistemas transaccionales se encuentran almacenados en diferentes fuentes de datos que incluso pudieran estar dispersas geográficamente, por lo que se requiere transportar los datos hacia una base de datos destino.

Los archivos se transportan a un servidor específico para manipulación de datos, en caso de que ambos servidores se encuentren interconectados, se puede efectuar este procedimiento por medio de ftp (File Transport Protocol), a este proceso se le conoce como transporte.

El contenido de estos archivos es depositado en tablas temporales contenidas en bases de datos específicas para manipulación de datos. Posteriormente, se eliminan posibles registros inútiles o con inconsistencias, es decir que se efectúan procesos de limpieza.

La limpieza de los datos puede ser manejada de diferentes maneras. Si los errores son inherentes a las aplicaciones fuente, los datos pueden ser limpiados sistemáticamente como parte del proceso de transformación. Desafortunadamente muchos errores ocurren porque las aplicaciones fuente sólo tienen una mínima validación de dominio, que permite la aparición de datos inválidos. La única manera de solucionarlos es corriendo rutinas pesadas de validación a nivel fuente; por ejemplo, los errores que surgen de tipos incorrectos, son muy difíciles de detectar y corregir. Véase la figura 1.4.



1.4. Ilustración de los procesos de Transporte y Limpieza

Los datos relevantes son procesados para obtener estimados, agregaciones, índices, relaciones, etc, requeridos por los directivos para sustentar sus decisiones, a este proceso se le conoce como transformación de datos.

La transformación involucra reglas de conversión de valores de aplicaciones locales a valores globales integrados.

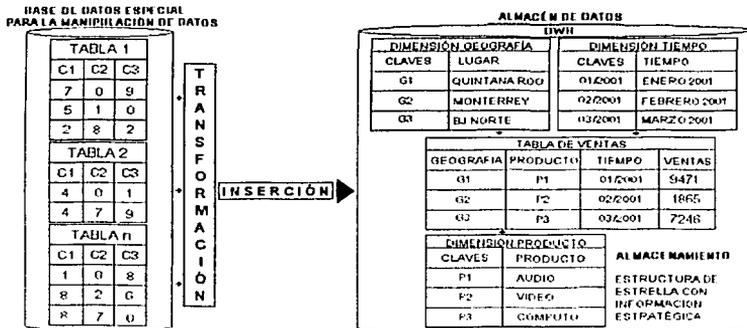
Las reglas de transformación de datos, integran los elementos atómicos y dispersos en las bases de datos transaccionales en piezas de información del Almacén de Datos; sin embargo, para lograr esto, es necesario abordar ciertos problemas que se derivan de la naturaleza no homogénea de las fuentes, por ejemplo:

- Diferentes representaciones para el dato fecha: algunos sistemas pueden almacenar las fechas con formato dd/mm/aa, o bien, dd/mmm/aaaa, etc.
- Diferentes tipos de datos representando el mismo concepto: el dato sexo puede ser representado como F y M, H y M, o bien Femenino y Masculino.
- Diferentes claves para representar el mismo elemento: un empleado puede tener como clave de identificación su curp para el Sistema de Nómina y un número consecutivo para el sistema de Recursos Humanos.
- Diferentes niveles de precisión al representar un dato: los números reales no siempre se almacenan de la misma forma, y es posible que esto genere algún problema.

En el proceso de transformación de datos se intentan resolver los problemas derivados de las distintas representaciones de fechas, tipos de datos, claves, etc., en vista de que se requiere integrar toda la información en el Almacén de Datos (DWH).

Los datos transformados, son transportados a un servidor específico, donde finalmente son insertados al Almacén de Datos; a este proceso se le denomina inserción.

Se pretende que estos procesos sean adaptados según las necesidades de cada organización y que sean ejecutados periódicamente a fin de actualizar con frecuencia los datos contenidos en el Almacén; la frecuencia de actualización se dicta según las políticas y necesidades de cada organización. Véase la figura 1.5.



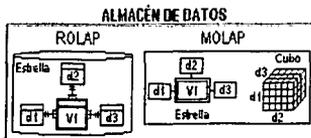
1.5. Ilustración de los procesos de Transformación e Inserción.

TESIS CON  
EN ORDEN

## 1.6. Almacenes de Datos y Tecnología OLAP.

El Almacén de Datos (Data Warehouse) es el centro de la arquitectura de los Sistemas de Información Estratégica. Surgen como respuesta a la problemática de almacenar información sintética a partir de datos atómicos almacenados en las bases de datos de producción.

Aunque a menudo se habla del Almacén de Datos como si fuera único, sus datos pueden estar distribuidos en múltiples bases manejadas por diferentes administradores de bases de datos. Véase la figura 1.6.



1.6. Esquema de Almacén de Datos

### Características de los Almacenes de Datos (DWH).

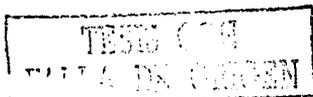
- Ofrecen información bajo demanda (análisis libre mediante el uso de herramientas de generación de informes).
- Reflejan el modelo del negocio en contraposición al modelo del proceso.
- Almacenan grandes volúmenes de datos (información histórica e integración de datos de múltiples aplicaciones).
- Poseen variables dimensionadas para consultas largas y elaboradas.
- Ejecutan actualizaciones controladas y no permiten la eliminación de datos (por lo que se conserva la información relativa a la historia de la organización).

### Funcionalidades.

- Integración de bases de datos tecnológicamente heterogéneas (relacionales, documentales, geográficas, archivos, etc.) con bases de datos geográficamente dispersas (base de datos ubicada en el Distrito Federal, base de datos ubicada en Monterrey, base de datos ubicada en Guadalajara, etc.) y con las bases de datos que se derivan de las distintas áreas o procesos de una organización (base de datos de Recursos Humanos, base de datos de Ventas, etc.).
- Ejecución de consultas complejas no predefinidas, visualizando el resultado en forma de gráficas o tablas, y en diferentes niveles de agrupamiento y agregación de datos.
- Agrupamiento y desagrupamiento de datos en forma interactiva.
- Análisis de problemas en términos de dimensiones. Por ejemplo, permite analizar datos históricos a través de una dimensión de tiempo.
- Control de calidad de datos para asegurar no sólo la consistencia del Almacén de Datos, sino también la relevancia de los datos con base a los cuales se toman las decisiones.

### Tecnología de Procesamiento Analítico en Línea (OLAP)

Los sistemas OLAP tienen la habilidad de responder a las preguntas quién y qué; sin embargo, los Almacenes de Datos añaden la posibilidad de responder las preguntas qué tal sí y por qué. Los sistemas OLAP posibilitan la toma de decisiones sobre acciones futuras. Un cálculo típico OLAP es más complejo que una simple sumariazión de datos.



Las aplicaciones OLAP están orientadas a diversas áreas organizacionales. Los departamentos de finanzas usan sistemas OLAP para aplicaciones específicas tales como análisis de estados financieros y modelos financieros. En el caso del departamento de ventas el análisis de ventas es una aplicación OLAP típica y otras aplicaciones pueden ser análisis de investigación de mercados, análisis de promociones, análisis de clientes, etc.

Aunque los sistemas OLAP se aplican en diversas áreas, todas ellas buscan lo siguiente:

#### **a. Vistas multidimensionales.**

Las vistas multidimensionales son las representaciones inherentes de los actuales modelos de datos; es raro el modelo que se limita a menos de tres dimensiones. Un directivo generalmente requiere datos financieros organizados por escenarios por ejemplo, gasto actual contra gasto presupuestado analizados por organización, concepto (por ejemplo, proyectos) y tiempo, o bien el análisis de las ventas organizadas por producto, geografía, canal y tiempo.

Los directivos deben tener la capacidad de analizar los datos a través de cualquier dimensión y a cualquier nivel de agregación con igual funcionalidad. El software OLAP debe soportar este tipo de vistas de datos en una forma natural y funcional, proporcionando a los usuarios las facilidades necesarias para obtener la información crítica sin necesidad de utilizar sintáxis para el desarrollo de queries complejos. Después de todo, los directivos no tienen por qué conocer dicha sintáxis ni elaborar concatenaciones de tablas ni sumalizaciones.

#### **b. Cálculos complejos.**

La característica más significativa de las bases de datos OLAP es su habilidad de realizar cálculos complejos ya que deben ser capaces de hacer más que simples agregaciones.

Por ejemplo, el análisis de las finanzas incluye algoritmos de tendencias tales como promedios y porcentajes de crecimiento. El análisis de las ventas y promociones de una determinada organización y de sus competidores, requiere del modelado de relaciones complejas entre los involucrados. El mundo real es complicado, por lo que la habilidad de modelar relaciones complejas es el aspecto más importante en las aplicaciones que incluyen procesos de análisis.

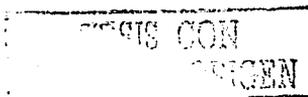
El software OLAP debe proveer de un conjunto variado de herramientas poderosas que permitan lograr que los desarrolladores sean más eficientes y los usuarios más autosuficientes.

#### **c. Manejo inteligente del tiempo**

El tiempo es un componente integral de casi cualquier aplicación analítica. El tiempo es una dimensión única por su carácter secuencial (enero siempre se encuentra antes de febrero). Un sistema real OLAP entiende la naturaleza secuencial del tiempo.

El éxito de los negocios generalmente se juzga a través del tiempo, por ejemplo, este mes contra el mes anterior, o el mes actual contra el mismo mes del año anterior.

La jerarquía del tiempo no siempre se maneja de la misma manera que las otras jerarquías. Por ejemplo, un directivo podría requerir la información respecto a las ventas en mayo o las ventas de los primeros cinco meses del año de un año determinado. El mismo directivo podría requerir la información de las ventas de tornillos (por ejemplo), pero nunca consultará las ventas para los cinco primeros tornillos. En suma, las comparaciones por periodos de tiempo son fácilmente definidas en un sistema OLAP.



## **Arquitecturas de los sistemas OLAP.**

Los Sistemas de Procesamiento Analítico en línea pueden ser desarrollados mediante la aplicación de dos tipos de arquitecturas diferentes que se describen a continuación.

### **a. Sistemas de Procesamiento Analítico en Línea Relacional (ROLAP).**

La arquitectura ROLAP accede a los datos almacenados en un Almacén de Datos mediante el uso de un manejador de base de datos relacional, con el fin de proporcionar los medios para realizar análisis dinámico de datos; en esta arquitectura se manejan los requerimientos para el acceso, recopilación y almacenamiento de datos. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios, y el motor ROLAP se integra con niveles de presentación, a través de los cuales los usuarios realizan consultas.

Los usuarios finales ejecutan análisis multidimensional a través del motor ROLAP, que transforma dinámicamente las consultas en sentencias SQL, las cuales se ejecutan en los Almacenes de Datos, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados solicitados.

Los Almacenes de Datos son óptimos para las consultas de datos, ya que éstos no se modifican; en ellos las tablas se encuentran fuertemente indexadas y las filas de datos (artículos) son a menudo preprocesadas y organizadas para soportar los diversos tipos de consultas que serán ejecutadas. Debido a que los usuarios no modifican los datos, los problemas de concurrencia y atomicidad no constituyen una preocupación; los datos son modificados o adicionados sólo por períodos y los grandes volúmenes de actualización son hechos fuera de horario, o sea, en el tiempo de mínimo tráfico en el Almacén de Datos.

Teniendo en cuenta lo anterior, y sabiendo que estos sistemas tienen bajos requerimientos de actualización y grandes volúmenes de datos, los Almacenes de Datos deben ser diseñados para promover un indexado fuerte que permita agilizar la ejecución de las consultas; además, la denormalización del Almacén de Datos satisface aquellos requerimientos de consultas más comunes y agiliza el tiempo de respuesta de las mismas, introduciendo datos preagregados o sumarizados.

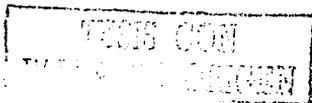
### **Modelo de datos ROLAP.**

Los Almacenes de Datos desarrollados con tecnología relacional contienen estructuras de datos denominadas esquemas estrella, las cuales se componen de un tabla central conocida como tabla de hechos, la cual contiene atributos. Técnicamente, la tabla de hechos es una intersección de entidades cuya llave primaria es llave compuesta; el dominio de cada componente de la llave está formada por la unión de los dominios de los diferentes niveles de las dimensiones.

Las dimensiones son tablas que contienen criterios de explotación de las variables contenidas en las tablas de hechos. Las tablas de dimensiones se relacionan con las tablas de hechos a partir de la herencia de su llave primaria.

En el caso de que dos o más tablas de hechos compartan una tabla de dimensión, se dice que ese conjunto de estrellas forma una arquitectura de estrella compuesta o constelación.

En vista de que los Almacenes de Datos responden a necesidades de consultas complejas, la duplicación de información con fines de eficiencia es permitido en este modelo de datos; por ello, es común encontrar dimensiones que contienen criterios tales como localidades geográficas, municipios y entidades federativas. En términos de las bases de datos relacionales, esto sería un error de redundancia, ya que a partir de los datos contenidos a nivel localidad pueden calcularse



los datos a los siguientes niveles de agregación; sin embargo, aunque la arquitectura ROLAP permite el desarrollo del Almacén de Datos a partir de tecnología relacional, esta duplicidad no sólo es permitida sino que es necesaria por motivos de eficiencia en las consultas. Al proceso que permite calcular datos sumariados a distintos niveles de detalle se le conoce como consolidación. Véase Tabla 1.

#### b. Sistemas de Procesamiento Analítico en Línea Multidimensional (MOLAP).

Esta arquitectura utiliza Bases de Datos Multidimensionales (BDMD) para almacenar, organizar, navegar y analizar los datos y requiere un acoplamiento fuerte entre las capas de aplicación y presentación.

El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP, mientras que el nivel de presentación proporciona una interfaz, a través de la cual los usuarios finales realizan análisis OLAP. Una arquitectura Cliente/Servidor permite a varios usuarios solicitar informes a través de la interfaz y la lógica de la aplicación, y también pueden acceder a la Base de Datos Multidimensional (BDMD) para obtener los datos.

El sistema MOLAP contiene datos precalculados, así como los resultados de posibles consultas, por lo que éstas son muy rápidas; sin embargo, esto requiere gran espacio de almacenamiento.

#### Modelo de datos multidimensionales.

La explotación de los datos contenidos en los Almacenes de Datos generalmente son diseñados tomando en cuenta un modelo de datos llamado arquitectura de estrella o cubo de información.

##### Arquitectura de estrella.

Una arquitectura de estrella válida es aquella que consta de una intersección de entidades simples (tabla de hechos) unida a una o más dimensiones apropiadas, las cuales constituyen los criterios de explotación de las variables.

Para ejemplificar un esquema estrella, supóngase que se desean conocer las ventas de ciertos productos por región; en este caso, el modelado en estrella tendría como parte central o variable a las ventas y los criterios de explotación o dimensiones serían los productos y las regiones. Véase la figura 1.7.



Figura 1.7. Esquema estrella

##### Arquitectura de estrella compuesta.

La arquitectura de estrella compuesta consta de dos arquitecturas de estrellas válidas que se encuentran conectadas a través de la estructura dimensional o por la conexión de las tablas de hechos. Véase la figura 1.8.

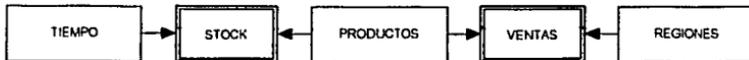
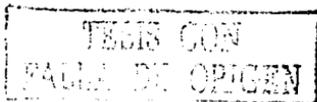


Figura 1.8. Esquema de estrella compuesta



Supóngase que se tienen las variables stock y ventas, tal y como se muestra en la figura 1.8, ambas variables tienen como dimensión común a los productos; en este caso se dice que la dimensión producto establece una conexión entre ambas variables, y por lo tanto, podemos afirmar que ésta es una estrella compuesta. Algunos autores utilizan el término constelación para referirse al tipo de estructura de estrella compuesta.

Los hechos son las variables que se desean explotar y algunos ejemplos son: el número de clientes de una empresa, el número de empleados, el monto de las ventas, el monto de las pérdidas, etc.

Las dimensiones son los criterios, mediante los cuales, se desean explotar las variables. Para ejemplificar este concepto, podríamos considerar la variable de monto de ventas y los criterios de explotación de esta variable podrían ser: en un período de tiempo o a una fecha determinada, y por ubicación geográfica y por producto. En este caso se dice que la variable monto de las ventas está dimensionada por los criterios de: tiempo, ubicación geográfica y producto.

La notación para expresar lo anterior es:

ventas (tiempo, geografía, producto)

En las bases de datos multidimensionales la información se almacena dimensionalmente y no relacionalmente. Las dimensiones determinan la estructura de la información almacenada y definen adicionalmente caminos de consolidación. La información almacenada se presenta como variables que a su vez están caracterizadas por una o más dimensiones. Un ejemplo de esto se muestra en la figura 1.9.

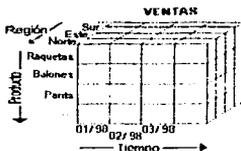


Figura 1.9. Cubo de información con tres dimensiones

De este modo, la información puede analizarse dentro del cubo formado por la intersección de las dimensiones de la variable particular. Por ejemplo, el gerente de producción podría estar interesado en analizar la información en el plano horizontal que se muestra en la figura 1.10 sección [b] y que representa el producto del que es responsable. De manera similar, el gerente de área puede seccionar el cubo en un plano vertical, el cual representa la región o área a su cargo, como se muestra en la figura 1.10 sección [a].

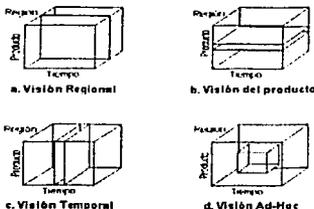


Figura 1.10. Vistas Multidimensionales

En todo momento, el gerente financiero puede comparar la información de diferentes años, o bien, hacer análisis de tendencias de las ventas, cortando el cubo en su profundidad. figura 1.10 sección [c].

Las dimensiones pueden estructurarse jerárquicamente, de modo que se puedan construir caminos de consolidación. Estos caminos de consolidación permiten analizar la información bajo un mecanismo "drill-down", es decir, de lo más general a lo más específico. Un ejemplo de esto es el análisis de las ventas de toda la compañía para observar que ha habido un crecimiento del 2% menor que el esperado, de acuerdo a la proyección realizada con la historia de años anteriores. Una visión más detallada puede decirnos cuál de todas las regiones experimentó un crecimiento menor, y bajar aún más, para observar la división que dió un menor rendimiento.

### Consolidación.

La consolidación es la clave para las consultas rápidas y consistentes. La consolidación es la presumarización lógica de subtotalet y totales. Ejemplo de consolidación en la Tabla 1.

	Este	Oeste	Central	Total
Clavos	50,000	60,000	100,000	210,000
Tornillos	40,000	70,000	80,000	190,000
Tuercas	90,000	120,000	140,000	350,000
Total	180,000	250,000	320,000	750,000

Tabla 1. Consolidación de la información en totales

Los datos que tienen las celdas originales son llamados inputs. Los totales calculados son llamados outputs. Este, Oeste y Central son miembros de la dimensión Región. El Total Región es un miembro output de la dimensión Región. Similarmente Clavos, Tornillos y Tuercas y Total son miembros de la dimensión Producto. Los valores que en este caso son celdas que representan una variable. Esta variable se encuentra dimensionada por Producto y Región.

La jerarquía es una instancia particular de la estructura de la dimensión. Una dimensión puede tener más de una jerarquía.

Las jerarquías representan los diferentes niveles de sumarización de la información y básicamente podemos diferenciar dos tipos de jerarquías: la jerarquía simple y la jerarquía múltiple.

### Jerarquia Simple.

En este tipo de jerarquía podemos apreciar que un elemento padre puede contener diversos elementos hijos y la restricción más importante consiste en que un elemento hijo puede tener un solo un elemento padre. Véase la figura 1.11.

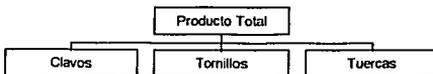


Figura 1.11. Jerarquía Simple

### Jerarquia Múltiple.

En la jerarquía múltiple un elemento hijo puede tener múltiples padres. Véase figura 1.12.

Drill Down: Significa bajar a un nivel detallado de la información.

Drill Up: Significa subir a un nivel agregado de la información.

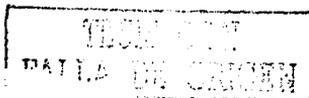


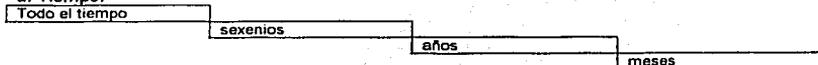


Figura 1.12. Jerarquía Múltiple

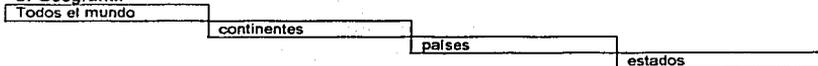
Para ejemplificar este concepto, tomemos en consideración el ejemplo anterior de las ventas por tiempo, geografía y producto: ventas(tiempo, geografía, producto)

Supóngase que se desea que las dimensiones tengan los siguientes niveles de abstracción:

**a. Tiempo:**



**b. Geografía:**



**c. Productos.**



**Series de Tiempo.**

El conjunto de valores de una variable en distintos instantes de tiempo es denominado Series de Tiempo. En el siguiente ejemplo, se muestran las series de tiempo para la variable ventas. (Véase la tabla 2)

Semanal	Mensual
1,500,000	
2,782,000	
2,568,000	
3,254,000	10,104

Tabla 2. Cinco series de tiempo

**Operaciones típicas que OLAP realiza con los datos**

- Swap: Alterar las filas por columnas (permutar dos dimensiones de análisis).
- Down: Bajar el nivel de visualización en las filas a una jerarquía inferior.
- Drill down: Explotación a niveles más detallados, informar para una fila en concreto, de datos a un nivel inferior (Pre-agregaciones).
- Expand: Operación idéntica al detalle, sin perder la información a nivel superior para éste y el resto de los valores.
- Collapse: Operación inversa de expandir.
- Drill up: Exploración de datos consolidados.
- Slice: Corte Bidimensional de la información.
- Dice : Rotación de las dimensiones.

## 1.7. Data Mart.

El uso efectivo de los Data Mart en un ambiente de almacenamiento de datos, es un factor importante para la efectividad del almacén, y puede también ser determinante en el éxito del proyecto de desarrollo. Los Data Mart son diseñados para satisfacer las necesidades específicas de los grupos comunes de usuarios (divisiones geográficas, divisiones organizacionales, etc.). Los Data Mart son generalmente subconjuntos de los Almacenes de Datos (DWH), pero también pueden integrar un número de fuentes heterogéneas e inclusive ser más grandes, en volumen de datos que el propio almacén central. Como los Data Mart son un factor crítico para el éxito del proyecto, también lo son su creación y su mantenimiento.

Para desarrollar un Almacén de Datos se requiere un considerable equipo de desarrolladores, hardware, software, tiempo y dinero. Las necesidades de las diferentes áreas de las organizaciones, a veces conflictivas, deben ser sobrellevadas en su conjunto. En algunas ocasiones, los usuarios consideran que es difícil construir Almacenes de Datos, y por lo tanto de usar. En consecuencia, las organizaciones están construyendo Data Mart, en lugar de o de manera complementaria a los Almacenes de Datos.

### Crecimiento de los Data Mart.

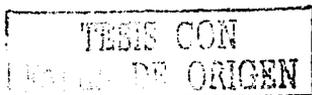
Los Data Mart pueden llegar a tener tamaños semejantes a los Almacenes Corporativos de menor escala. Aunque hoy en día es difícil diferenciar a los Data Mart y a los Almacenes de Datos por su tamaño, algunas distinciones entre ellos son todavía importantes, por ejemplo:

- Un Data Mart está enfocado a una sola área o grupo de usuarios, mientras que un Almacén de Datos contiene información de diferentes sujetos y áreas de la corporación.
- Una organización puede tener un solo Almacén de Datos, pero varios Data Mart.
- Los Data Mart no contienen información almacenada como datos operacionales, pero si la tienen los Almacenes de Datos.
- Los Data Mart contienen menos información y son más fáciles de entender y usar, que los almacenes corporativos. Un Almacén de Datos puede contener tanta información, que puede resultar más difícil su manejo.

El reciente crecimiento de los Data Mart, ha generado también, muchos problemas a los usuarios, para acceder a la información de la organización, ya que:

- Se pierde performance a medida que aumenta el tamaño de los Data Mart. Los usuarios esperan mejor respuesta de los Data Mart, que de los Almacenes de Datos.
- Los usuarios requieren acceso a datos de muchos Data Mart. Los datos pueden ser replicados entre los Data Mart, pero se requieren mejores soluciones.
- Las organizaciones no pueden administrar fácilmente muchos Data Mart. Mientras que sólo se tiene un Almacén de Datos, se pueden tener muchos Data Mart.
- Las organizaciones tienen dificultades para construir los Data Mart. Aunque es aceptable que la construcción de un Almacén se lleve varios años, los Data Mart requieren de un ciclo de desarrollo muy corto, y una inversión moderada.

Las soluciones de los Data Mart, requieren una arquitectura de 3 capas: Los Almacenes de Datos son la primera capa, los Data Mart son la segunda capa, y las estaciones de trabajo de usuarios son la tercera.

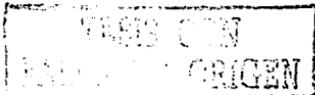


## 1.8. Metadatos.

Los diccionarios de datos o metadatos describen los datos almacenados con el objeto de facilitar el acceso a los mismos a través de las herramientas de explotación del Almacén de Datos. El diccionario de datos establece correspondencias entre los datos almacenados y los conceptos que éstos representan para facilitar la extracción de información por parte del usuario final.

El conocimiento de los metadatos es tan esencial como el conocimiento de los datos del Almacén. Deben incluir dominio, reglas de validación, derivación y transformación de los datos extraídos. También deben describir las bases de datos del Almacén, incluyendo reglas de distribución y control de migración hacia los Data Mart. Los procesos que monitorean a su vez los procesos del Almacén de Datos, tales como extracción, carga y uso deben crear metadatos que pueden usarse para determinar qué tan bien se comporta el sistema.

Los metadatos deben estar disponibles a los usuarios, con el fin de que puedan usarlos en sus análisis y para que los administradores puedan manejar y proveer el acceso a los mismos a través de los servicios del repositorio.



## 1.9. Minería de Datos.

La minería de datos (Data Mining) es el proceso de extracción de información significativa de grandes bases de datos, información que revela inteligencia del negocio a través de patrones ocultos, tendencias y correlaciones para permitir al usuario realizar predicciones que resuelvan problemas.

Las herramientas de minería de datos predicen nuevas perspectivas y pronostican situaciones futuras de las empresas, lo que evidentemente ayuda a tomar decisiones adecuadas.

Las técnicas y algoritmos de minería de datos son una poderosa herramienta de trabajo para el soporte a la toma de decisiones.

Tradicionalmente, se han utilizado dos tipos de análisis estadístico para la minería de datos:

### a. Análisis Confirmatorio.

En un análisis confirmatorio, se tiene una hipótesis y ésta se confirma o se refuta; sin embargo, el problema que puede presentarse es la carencia de hipótesis.

### b. Análisis Exploratorio.

En el análisis exploratorio, se encuentran hipótesis convenientes y éstas se confirman o se refutan. En este análisis, el que toma la iniciativa es el sistema y no el usuario.

## Actividades de minería de datos.

Desde un punto de vista orientado a procesos, hay tres clases de actividades de minería de datos como se muestra en la figura 1.13.

### a. Descubrimiento de patrones.

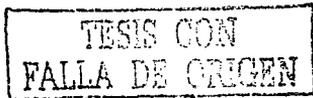
El descubrimiento es el proceso de búsqueda en una base de datos para encontrar patrones ocultos sin una idea predeterminada o sin hipótesis sobre patrones que pudieran encontrarse.

En otras palabras, el sistema toma la iniciativa de encontrar qué patrones son interesantes, sin que el usuario piense inicialmente sobre cuestiones relevantes. En grandes bases de datos hay muchos patrones sobre los cuales los usuarios prácticamente nunca piensan. Dos aspectos importantes son la riqueza de patrones que pueden ser descubiertos y la calidad de la información, ambos determinan el poder y facilidad de uso de las técnicas de descubrimiento.

#### *Ejemplo de descubrimiento de patrones.*

Supóngase que se tiene la base de datos demográfica de algún país, los usuarios pueden tomar la iniciativa de hacer preguntas a la base de datos tales como: cuál es el promedio de edad de los usuarios de los servicios bancarios y el sistema podría contestar que la edad promedio es de cuarenta y siete años. El mismo usuario podría después pedir al sistema que éste tome la iniciativa y encuentre algo interesante respecto a la edad y por sí mismo el sistema actuaría como lo haría un analista humano. El sistema podría encontrar algunos datos característicos, distribuciones, etc. y trataría de encontrar algunas densidades de datos que salen de lo ordinario.

Suponiendo que el sistema realiza un análisis de la base de datos demográfica para establecer relaciones entre profesión y edad, el sistema podría establecer:



IF profesión = atleta  
THEN edad < 30  
con un grado de certeza del 71%

Esta regla significa que si se encuentran 100 atletas, en la base de datos, 71 de ellos son menores de 30 años.

El sistema podría establecer:

IF profesión = atleta  
THEN edad < 60  
con un grado de certeza del 97%

Esta regla significa que si se encuentran 100 atletas en la base de datos, 97 de ellos son menores de 60 años.

Esta información puede obtenerse mediante la técnica de destilación de patrones de datos.

b. Modelado Predictivo.

El modelo predictivo le permite al usuario establecer valores de campos desconocidos a partir de patrones descubiertos previamente en la base de datos.

*Ejemplo de Modelado Predictivo.*

Usando el ejemplo anterior, una vez que sabemos que los atletas son usualmente menores de 30 años, podemos adivinar la edad de alguien en particular si sabemos que su profesión es atleta. Por ejemplo, si se muestra el registro de Sergio Garrido, cuya profesión es atleta, mediante la aplicación de las reglas encontradas previamente, podemos estar un 70% seguros de que tiene 30 años y podemos también tener la certeza de que tiene menos de 60 años. Nótese que el descubrimiento nos ayuda a encontrar conocimiento general, pero la predicción sólo adivina el valor de la edad para un individuo específico. Nótese también que en este caso, la predicción es transparente, (por ejemplo, es clara la razón por la que adivinamos que la edad de ese individuo es menor de 30 años). En algunos sistemas, la edad pudiera ser adivinada, pero la razón de la adivinanza pudiera no ser clara, lo que hace que el sistema sea opaco.

c. Análisis Forense.

El análisis forense es el proceso de aplicación de extracción de patrones para encontrar elementos de datos anómalos o inusuales. Para descubrir los datos inusuales, primero se deben encontrar las normas y, después, aquellos elementos que se desvían de dichas normas.

*Ejemplo de Análisis Forense.*

Usando nuevamente el ejemplo anterior, una vez que sabemos que el 97% de los atletas es menor de 60 años, podemos sorprendernos con el 3% que es mayor de 60 años y que aún practican el atletismo. Estos elementos son inusuales, pero aún no sabemos por qué. Ellos podrían ser inusualmente saludables o practicar deportes donde la edad es poco importante (por ejemplo el golf) o bien, la base de datos puede contener errores. Nótese que el descubrimiento ayuda a encontrar conocimiento común, pero el análisis forense encuentra elementos inusuales.

TEMA CON  
FALLA DE ORIGEN

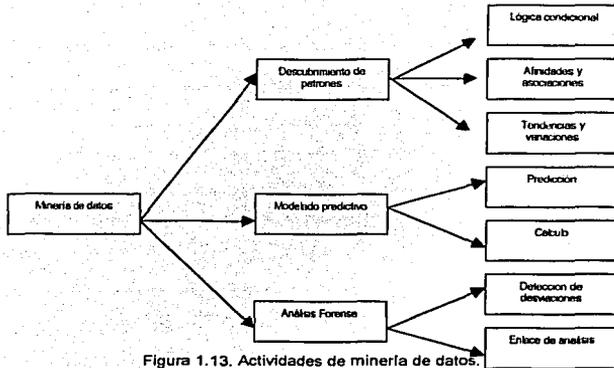


Figura 1.13. Actividades de minería de datos.

### Técnicas de minería de datos.

Para analizar las tecnologías empleadas para la minería de datos, podemos considerar una estructura de árbol, cuyas ramas reflejan las distintas familias de tecnologías. Véase la figura 1.14.

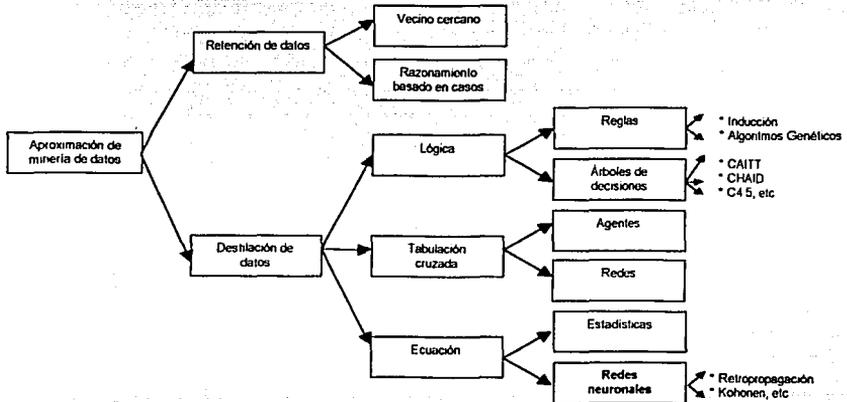


Figura 1.14. Árbol de técnicas de minería de datos

TESIS CON  
FALLA DE ORIGEN

La minería de datos está fundamentada en dos tecnologías, la retención de datos y la destilación de patrones. La rama de retención de datos se basa en la necesidad de mantener los datos analizados, es decir, aquellos a los que se les han aplicado técnicas de minería de datos para futuras asociaciones de patrones. Las técnicas basadas en la retención, sólo se aplican a las tareas de modelado de predicción y análisis forense.

## 1. Retención de datos.

En las aproximaciones con retención los datos, éstos se mantienen para realizar posteriores comparaciones de patrones; lo que implica que cuando se presentan nuevos datos, éstos son comparados con el conjunto de datos previo.

A continuación se describen dos métodos basados en la técnica de retención de patrones:

### a. Método del Vecino Cercano.

En este método se mantiene un conjunto de datos (generalmente en memoria), para la comparación con nuevos elementos de datos. Cuando se presenta un nuevo registro para realizar el proceso de predicción, se encuentra la distancia entre éste y los registros similares en el conjunto de datos, a fin de identificar los más similares o vecinos cercanos.

#### *Ejemplo del Vecino Cercano.*

Dado un prospecto de cliente para servicios bancarios, los atributos de los prospectos son comparados con los de todos los clientes de los bancos existentes. (por ejemplo las edades y montos de ingreso de los clientes). Posteriormente se selecciona un conjunto de vecinos cercanos al prospecto (basado en el ingreso, edad, etc. más cercana).

El término K-vecino más cercano es usado para especificar que tenemos los primeros K vecinos prospecto (por ejemplo los 10 primeros), como se muestra en la figura 1.15. A continuación, se realiza una comparación más cercana para seleccionar cual producto nuevo se adapta más al prospecto, basados en los productos usados por los primeros K vecinos.

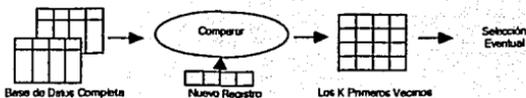


Figura 1.15. Ejemplo del vecino más cercano.

### b. Método del Razonamiento Basado en Casos.

En vista de que resulta caro retener datos, algunas veces solo se retiene un conjunto de casos típicos, por ejemplo, se pueden seleccionar y retener cien clientes típicos como una base de comparación. En este caso el problema principal consiste en seleccionar casos de clientes típicos; por ejemplo, ¿cómo podemos esperar seleccionar los casos típicos?, y si la base de clientes cambia, ¿cómo cambiamos el concepto de cliente típico?.

Otro problema se presenta cuando se tienen valores no numéricos, ya que resulta complejo determinar casos con base a este tipo de valores.

## 2. Destilación de Patrones.

En la tecnología de destilación de patrones se analizan datos, se extraen patrones y después se desechan. Por la naturaleza de esta tecnología, surgen dos preguntas: ¿qué tipos de patrones pueden extraerse? y ¿cómo pueden ser representados éstos patrones?. A continuación se presentan algunas aproximaciones que intentan responder a las preguntas planteadas.

*Aproximaciones de la tecnología de destilación de patrones.*

Obviamente los patrones necesitan ser expresados con un formalismo y un lenguaje. Esta elección es la raíz de tres distintas aproximaciones: Lógica, Tablas y Ecuaciones. Cada una de estas aproximaciones marcan su raíz histórica con un origen matemático distinto.

No todas las aproximaciones basadas en la destilación de patrones proveen conocimiento, los patrones pueden ser destilados con un lenguaje opaco o con un formalismo no muy accesible para las humanas, por ejemplo, ecuaciones complejas. Por lo tanto, algunas de estas aproximaciones producen transparencia y patrones comprensibles que proporcionan conocimiento, otras solo producen patrones usados para predicción opaca.

*Tipos de datos usados por las aproximaciones.*

La lógica puede tratar tanto con datos numéricos como no numéricos, mientras que las ecuaciones requieren que todos los datos sean numéricos y en la tabulación solo se trabaja con datos no numéricos.

*Aprendizaje Supervisado y no Supervisado.*

Mientras que en el aprendizaje supervisado hay un maestro que enseña al sistema, y que indica cuándo la predicción es correcta o incorrecta, en el aprendizaje no supervisado, el sistema no tiene maestro pero se trata de encontrar grupos de patrones interesantes en la base de datos.

### a. Aproximaciones por lógica.

La lógica forma las bases de muchos de los lenguajes escritos y es esencial para el proceso del pensamiento que se lleva en el lado izquierdo del cerebro. Los patrones que se expresan en lenguajes lógicos son distinguidos por dos elementos importantes: son legibles y entendibles.

El operador central en los lenguajes lógicos son una variación de los bien conocidos enunciados if / then (si / entonces). Por ejemplo si está lloviendo, entonces está nublado; sin embargo, nótese que mientras que la forma más común de la lógica es la lógica condicional, frecuentemente se requiere el uso de otros tipos de lógica tales como asociación lógica con reglas como when / also (cuando / también). Por ejemplo, cuando se compra pintura, también se compran brochas.

La lógica proposicional y de predicados (por ejemplo, lógica condicional) son las mejor conocidas; sin embargo, la lógica variacional y de la de tendencias son también usadas en análisis de datos de negocios.

\* Reglas.

Las relaciones lógicas generalmente se representan como reglas. Los tipos más simples de reglas expresan condicionales o relaciones de asociación. Una regla condicional es un enunciado de la forma:

IF	Condición 1
THEN	Condición 2

### Ejemplo de Reglas.

En la base de datos demográfica, tenemos la regla IF profesión = atleta THEN edad < 30. Aquí se comparan los valores de los campos de una tabla dada. Aquí la profesión es el atributo y atleta es el valor. Otro ejemplo de expresión atributo-valor es País = México, donde País es el atributo y el valor es México.

Las reglas condicionales generalmente trabajan en tablas con atributos (por ejemplo campos) y valores, como se muestra a continuación:

Nombre	Profesión	Edad
Sergio Garrido	Atleta	27
...	...	...

Tabla 3

Las reglas pueden ir más allá de las representaciones de atributos - valores. En los atributos lógicos comparamos los valores de dos campos, sin la denominación explícita de ningún valor.

#### \* Reglas de Asociación.

Mientras que las reglas discutidas anteriormente son de naturaleza IF-THEN, las reglas de asociación se refieren al agrupamiento de elementos (por ejemplo, alguien compra un producto en la tienda y podría comprar otro producto al mismo tiempo; este proceso es comúnmente llamado análisis de carro de supermercado). La potencialidad de un sistema de descubrimiento se mide por tipos y generalidades de los patrones, los cuales pueden encontrarse y expresarse en un lenguaje conveniente.

#### \* Afinidad Lógica vs. Lógica Condicional.

La afinidad lógica es diferente de la lógica condicional en términos del lenguaje de expresión y de la estructura de datos que usa. El análisis de afinidades (o análisis de asociaciones) es la búsqueda de patrones y condiciones que describen cómo varios elementos se agrupan o tienden a agruparse con una serie de eventos o transacciones. Una regla de afinidad tiene la forma:

WHEN Elemento 1
ALSO Elemento 2

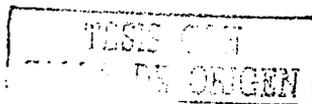
#### Ejemplo de Afinidad.

Un ejemplo de afinidad es, WHEN pintura ALSO brocha. En una análisis simple de afinidad, los sistemas usan una tabla de transacción como la siguiente:

No. Transacción	Elemento
123	Pintura
123	Brocha
123	Clavos
124	Pintura
124	Brocha
124	Madera
125	...

Tabla 4

El campo número de transacción es usado para agrupar elementos, mientras que el campo elemento contiene las entidades agrupadas. En este ejemplo, la afinidad para la transacción 123 y 124 es el par (pintura, brocha).



### Ejemplo de Afinidades multidimensionales OLAP.

Las afinidades planas necesitan ser extendidas a afinidades multidimensionales, es decir OLAP para mejores resultados. Una afinidad dimensional tiene la forma:

Confianza = 95 %		
IF	día	= sábado
WHEN	elemento = brocha	
ALSO	elemento = pintura	

Aquí se combinan las condiciones lógicas y las asociaciones. Esta forma de estructura híbrida proporciona a la lógica un poder real y transparencia.

Las reglas tienen la ventaja de ser capaces de tratar con datos numéricos y no numéricos de una manera uniforme. Cuando se trata con datos numéricos, algunas aproximaciones intercambian los campos numéricos por códigos con valores específicos. Esto puede remover todas las consideraciones numéricas de los códigos, lo que repercute en la pérdida de patrones. Por ejemplo, el campo edad puede fragmentarse en tres rangos (1-30), (31-60), (61-100), correspondientes a jóvenes, edad mediana y ancianos. Por supuesto, los datos pueden contener patrones que se traslapen en cualquiera de estos tres rangos (por ejemplo, el rango (27-34) podría ser muy significativo para algunos patrones y cualquier aproximación basada en asignación de código podría omitirlo).

Las reglas pueden también trabajar con datos multidimensionales y tipo OLAP porque ellos pueden tratar con rangos de datos numéricos y sus formatos lógicos permiten que sus patrones se fusionen en múltiples dimensiones.

A continuación se revisan dos aproximaciones para la generación de reglas, llamadas Inducción y Algoritmos Genéticos.

#### \* Inducción.

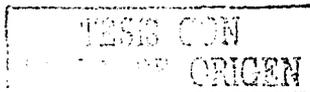
La inducción es el proceso de búsqueda en un conjunto de datos para generar patrones, por medio de la exploración automática del conjunto de datos, tal y como se muestra en la figura 1.16, la inducción forma las hipótesis que guían a los patrones.



Figura 1.16. Reglas de inducción.

#### \* Reglas Borrosas.

Es necesario distinguir entre reglas borrosas y reglas inexactas. Las reglas inexactas frecuentemente tienen asociados a ellas factores de confianza mezclados, por ejemplo, cada regla tiene asociado un entero o un porcentaje que representa su validez. Sin embargo, el grado de confianza en reglas borrosas puede variar en términos de valores numéricos en el cuerpo de la regla; por ejemplo, la confianza puede ser proporcional a la edad de la persona y así como la edad varía lo hace la confianza. De esta manera las reglas borrosas pueden producir expresiones más compactas de conocimiento y guiar hacia comportamientos más estables.



**\* Algoritmos Genéticos.**

Los algoritmos genéticos solo generan reglas a partir de los conjuntos de datos, pero no siguen el protocolo orientado a la exploración de reglas de inducción. En vez de ello, apuestan a la idea de mutación para hacer cambios en patrones hasta que una forma de patrones confiables emerge vía la selección de especies. Véase la figura 1.17.

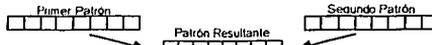


Figura 1.17. Algoritmos genéticos.

**\* Operación Genética Cross-Over.**

La operación genética denominada como cross-over es muy similar a la operación que se usa para mezclar especies de plantas o animales. El intercambio de material genético por cromosomas está basado en el mismo método; en el caso de reglas, el material intercambiado es una parte del patrón que describe las reglas.

**\* Reglas de Inducción vs. Algoritmos Genéticos.**

El foco principal en algoritmos genéticos es la combinación de patrones procedentes de reglas que han sido descubiertas previamente, mientras que en la inducción el, foco principal de la actividad está en el conjunto de datos.

**\* Árboles de decisiones.**

Los árboles de decisiones expresan en forma simple la lógica condicional. En un sistema de árboles de decisiones se realizan particiones simples a una tabla para obtener tablas de menor tamaño mediante la selección de subconjuntos.

Fabricante	Estado	Ciudad	Color del Producto	Productividad
Garrido	Veracruz	Córdoba	Azul	Alto
Garrido	Oaxaca	Salina Cruz	Verde	Bajo
Widrow	Nayarit	Tepic	Azul	Alto
Widrow	Oaxaca	Salina Cruz	Rojo	Bajo
Luyando	Nayarit	Tepic	Verde	Promedio
Luyando	Veracruz	Córdoba	Rojo	Promedio

Tabla 5. Ejemplo de árboles de decisiones

La representación del árbol de decisión de la tabla anterior se muestra en la figura 1.18.

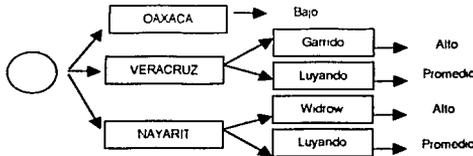
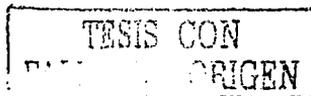


Figura 1.18. Árbol de decisión.



En este árbol de decisiones se selecciona en primera instancia al atributo "Estado" para iniciar la operación de partición, y posteriormente se coloca el atributo fabricante. Por supuesto, si hay 100 columnas en la tabla, la pregunta es cuál es el atributo que deberá ser seleccionado en primera instancia. En efecto, en muchos casos, incluyendo la tabla anterior, cualquier atributo que se seleccione del árbol derivaría en pérdida de información; nótese que las reglas son mucho más que árboles de decisiones.

Considérense las siguientes afirmaciones:

- a. Los productos azules presentan alta productividad.
- b. Oaxaca es el Estado menos productivo.

Dichas afirmaciones no pueden derivarse directamente de los datos contenidos en la tabla anterior y del árbol de decisión planteado; tampoco podemos inferir lo anterior considerando únicamente el árbol, ya que éste selecciona un atributo específico que divide cada estado. Por otro lado, las reglas y la tablas pueden descubrir ambas afirmaciones.

#### b. Tablas.

Las tabulación cruzada es una forma muy básica de análisis de datos, muy conocida en estadística, y ampliamente usada para generación de reportes. Una tabla de dos dimensiones es equivalente a una hoja de cálculo.

*Ejemplo de tablas.*

Si ordenamos los valores de la tabla 5 en términos de la productividad, los datos aparecen como se muestra en la tabla 6:

	Veracruz	Oaxaca	Nayarit	Azul	Verde	Rojo
Productividad alta	1	0	1	2	0	0
Productividad promedio	1	0	1	0	1	1
Productividad baja	0	2	0	0	1	1

Tabla 6

Aquí no se incluyen los campos de "manufactura" y "ciudad" porque la tabla se apreciaría muy grande. Sin embargo, como se puede notar el grado de co-ocurrencias entre "azul" y "alto", lo que indica una alta relación.

\* Agentes.

El término "agente" es frecuentemente usado para referirse a tablas que son desplegadas gráficamente en una red y que permiten realizar algunas conjunciones (por ejemplo el operador "and"), el término "agente" es equivalente al término "par campo-valor".

*Ejemplo de agente.*

Por ejemplo, si consideramos la tabla 6 podemos definir seis "agentes" para el concepto de productividad alta, lo cual se muestra gráficamente en la figura 1.19.

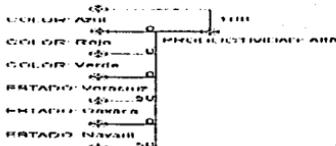
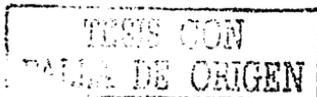


Figura 1.19. Ejemplo de agente



Nótese que aquí los pesos 100 y 50 son simplemente los porcentajes de valores que están asociados con el objetivo, los cuales, representan el impacto de las proporciones, y no el de las probabilidades.

En otra técnica de tabulación, los valores numéricos se transforman en una mezcla de códigos; por ejemplo, transformar edad en tres clases de edades: [1-30], [31-60], [61-100]. Por supuesto, los datos pueden contener patrones que traslapan cualquiera de estos rangos, por ejemplo el rango [27-34], no podría ser detectado por el agente. Si los rangos seleccionados son muy pequeños, entonces se genera una gran cantidad de patrones. Esta incapacidad para tratar con los datos numéricos causan problemas con datos multidimensionales.

Generalmente las redes son representadas mediante la gráfica de distribuciones probabilísticas (derivadas de conteos). Una red es por consiguiente un grafo dirigido, que consta de nodos (representación de variables) y arcos (representación probabilística de dependencias) entre las variables.

En la figura 1.20 se muestra un ejemplo de red, donde solo ha sido dibujado el atributo del color por razones de simplicidad.



Figura 1.20. Ejemplo de red.

Cada nodo contiene una distribución condicional de probabilidades que describen la relación entre nodos y sus padres; la gráfica de la red es acíclica.

### c. Aproximaciones con ecuaciones.

En este tipo de sistemas, el método de expresión de patrones empleado es el de "construcción de superficies" en vez de la expresión lógica o del conteo de ocurrencias. Cada sistema normalmente usa un conjunto de ecuaciones para definir una superficie en un espacio numérico. El mejor ejemplo conocido de una superficie es la línea recta en un espacio bidimensional, como se muestra en la figura 1.21 a. Esta es la ecuación  $Y=(a \cdot X)+b$  lo que conduce a una buena aproximación en regresión lineal en estadística. Como el parámetro  $a$  varía en esta ecuación, la pendiente de la línea cambia.

La regresión trabaja bien cuando los puntos se encuentran situados de manera cercana a una línea recta. Como se muestra en las figuras 1.21b y 1.21c, también es posible usar ecuaciones no lineales para aproximar superficies suaves.

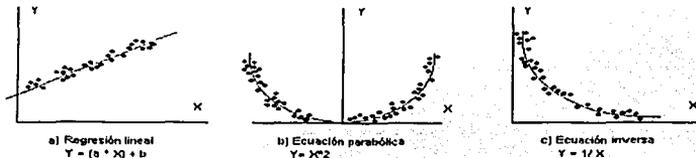


Figura 1.21. Aproximaciones por ecuaciones.

Cuando las superficies son más complejas (por ejemplo  $Y = (X^{2.1x} + (1/X))$ ), o cuando hay muchas dimensiones, la habilidad de los humanos de entender las ecuaciones y superficies decrece rápidamente, por lo que los sistemas se convierten en opacos o "cajas negras".

Las aproximaciones con ecuaciones requieren que el conjunto de datos sean todos de tipo numéricos, e incluso, los datos no numéricos necesitan ser codificados en números, lo cual frecuentemente causa un gran número de problemas.

\* Redes Neuronales Artificiales.

Las redes neuronales son un caso particular de aproximaciones con ecuaciones, las cuales, son sistemas de modelado predictivo que trabajan con ajustes de parámetros. Estructuralmente, las redes neuronales constan de un número de elementos interconectados (llamadas neuronas) organizadas en capas, las cuales aprenden por modificación de pesos en las conexiones. Véase la figura 1.22.

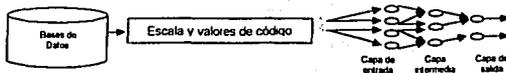


Figura 1.22. Red neuronal artificial.

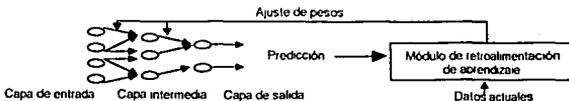
Las redes neuronales generalmente construyen ecuaciones complejas de superficies a través de iteraciones repetidas; con cada iteración se ajustan los parámetros que definen las superficies. Después de muchas iteraciones, una superficie puede ser internamente definida para aproximar muchos de los puntos de los conjuntos de datos.

Las funciones básicas de cada neurona son: (a) evaluar los valores de entrada, (b) cálculo del total de combinaciones de valores de entrada, (c) comparar el total de valores de umbrales y (d) determinar los valores de salida. Mientras que la operación de cada neurona es bastante simple, el comportamiento complejo puede ser creado por la conexión de cierto número de neuronas. Típicamente las neuronas de entrada son conectadas a la capa intermedia (o múltiples capas intermedias), la cual es posteriormente conectada a la capa de salida, como se puede ver en la figura 1.22.

Para construir un modelo neuronal, primero se entrena la red con un conjunto de entrenamiento, y posteriormente se utiliza la red para hacer predicciones. Se puede usar un conjunto de datos de monitoreo durante la fase de entrenamiento para revisar el progreso del entrenamiento.

Generalmente, cada neurona tiene un conjunto de pesos que determina cómo se evalúa la combinación de la fuerza de las señales de entrada. Las entradas de las neuronas pueden ser positivas (excitatorias) o negativas (inhibitorias). El aprendizaje consiste en el cambio de los pesos usados por las neuronas de acuerdo a la clasificación de errores que la red en su conjunto ha incurrido. Las entradas son generalmente escaladas y normalizadas para producir comportamientos suaves.

Durante la etapa de entrenamiento, el conjunto de pesos de la red determina el comportamiento de las capas intermedias. Un tipo de aproximación popular se denomina retropropagación, en la cual los pesos son ajustados con base a qué tan cercana ha hecho sus conjeturas la red. Las conjeturas incorrectas reducen el umbral para conexiones apropiadas, como en la figura 1.23.



TESIS CON  
ORIGEN

### Origen.

Al parecer, el algoritmo de retropropagación fue ideado a principios de los 70's por Werbos, y redescubierto a principios de los 80's por Parker y Rumelhart independientemente, sin embargo, no se hizo popular hasta 1986, cuando Rumerhart, Hinton y Williams presentaron una descripción clara y concisa del mismo.

Desde la fecha clave de 1986 han surgido nuevas versiones que han tratado de aumentar la velocidad de convergencia del algoritmo y han tratado de superar algunos de sus inconvenientes, como la tendencia a alcanzar mínimos locales y no globales.

### Retropropagación.

Al hablar de redes de retropropagación o redes de propagación hacia atrás hacemos referencia a un algoritmo de aprendizaje más que a una arquitectura determinada. La retropropagación consiste en propagar el error hacia atrás, es decir, de la capa de salida hacia la capa de entrada, pasando por las capas ocultas intermedias y ajustando los pesos de las conexiones, con el fin de reducir dicho error. Hay distintas versiones o reglas del algoritmo de retropropagación y distintas arquitecturas conexionistas a las que pueden ser aplicados.

Durante mucho tiempo no se dispuso de algoritmos para entrenar redes multicapa, y como las redes de una capa estaban muy limitadas en cuanto a lo que eran capaces de representar, el campo de las redes neuronales artificiales estaba estancado. La invención y perfeccionamiento del algoritmo de retropropagación dio un gran impulso al desarrollo de este campo. Tiene un buen fundamento matemático y a pesar de sus limitaciones ha expandido enormemente el rango de problemas donde se aplican las redes neuronales artificiales.

### Descripción matemática del algoritmo de retropropagación.

Se explica una versión del algoritmo (Hinton, 1992) para redes con las siguientes características:

- No recurrentes
- Función de activación sigmoideal
- Capas totalmente interconectadas
- Operación totalmente síncrona

### Algoritmo.

1. Aleatorizamos los pesos de las conexiones.
2. Presentamos un patrón de entrada y calculamos la salida.
3. Dada una unidad  $j$ -ésima de la capa de salida  $y_j$  y unidades  $i$ -ésimas de la capa oculta inmediatamente anterior, calculamos la entrada total ponderada y la salida o activación de la misma. Véase la figura 1.24.

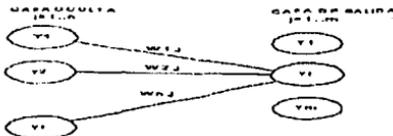


Figura. 1.24. Red con  $n$  neuronas en la capa oculta y  $m$  en la de salida.

$$x_j = \sum_{i=1}^n (y_i \cdot w_{ij})$$

$$y_j = \frac{1}{1 + e^{-x_j}}$$

4. Una vez computadas las actividades de todas las unidades de salida se calcula una estimación del error, generalmente una función cuadrática de los errores individuales cometidos por cada unidad, siendo cada error individual, la diferencia entre la salida deseada y la obtenida.

$$E = \frac{1}{2} \sum_j (d_j - y_j)^2 \quad \text{siendo } d_j \text{ la salida deseada para la unidad } j\text{-ésima}$$

Nota: Se indican por un lado las expresiones matemáticas y por otro lado la explicación intuitiva de cada paso. Conviene recordar que nuestro objetivo es calcular cómo varía el error al variar el peso de cada conexión (tasa de variación del error respecto al peso de una conexión, EP).

5. Cómputo de la rapidez de variación del error al cambiar la actividad de cada unidad de salida.

$$EA_j = \frac{\partial E}{\partial y_j} = d_j - y_j \quad \text{EA, error respecto a la actividad}$$

Es justamente la diferencia entre la salida deseada y la salida real obtenida, es decir, la diferencia entre la actividad deseada y la actividad real.

6. Cómputo de la rapidez de variación del error al cambiar la entrada total que recibe cada unidad de salida.

$$\delta_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial x_j} = EA_j y_j (1 - y_j) = y_j (1 - y_j) (d_j - y_j)$$

Es igual a la tasa de variación del error al variar su activación multiplicado por la tasa de variación de la activación al cambiar su entrada ( que es justamente la derivada de la función sigmoideal).

7. Cómputo de la rapidez de variación del error al ser modificada un peso de la conexión referente a una unidad de salida.

$$EP_{ij} = \frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial x_j} \frac{\partial x_j}{\partial w_{ij}} = \delta_j y_j$$

Es igual a la tasa de variación del error al variar su entrada, por la tasa de variación de la entrada al variar ese peso.

Hasta ahora sabemos calcular el EA sólo para las unidades de salida, ¿qué pasa con las unidades ocultas?. En este caso no tenemos una estimación directa del error aportado por cada unidad oculta; aquí es donde interviene la retropropagación o propagación hacia atrás del error:

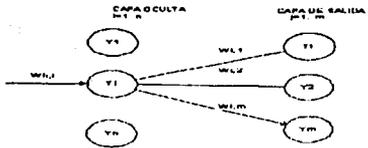
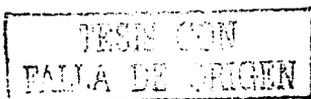


Figura 1.25. La unidad i-ésima de la capa oculta afecta a todas las unidades de salida.

La unidad i-ésima de la capa oculta afecta a todas las unidades de salida, por lo tanto, para estimar como varía el error al variar la actividad de esa unidad oculta, habrá que sumar los efectos individuales de su actividad sobre todas las neuronas de salida. Cada efecto individual sobre la variación del error, será igual a la tasa de variación del error de la unidad de salida al cambiar su entrada total, multiplicado por la tasa de variación de su entrada al variar la actividad de la unidad oculta. Véase la figura 1.25.



$$EA_i = \frac{\partial E}{\partial x_i} = \sum_{j=1}^n \left( \frac{\partial E}{\partial x_j} \frac{\partial x_j}{\partial x_i} \right) = \sum_{j=1}^n (e_j \cdot w_{ji})$$

8. Conociendo EA para las unidades de cualquier capa podemos calcular d y EP con las expresiones ya conocidas.

$$d_i = EA_i \cdot (1 - y_i) - y_i \cdot (1 - y_i) \sum_{j=1}^n (e_j \cdot w_{ji})$$

$$EP_h = e_h \cdot y_h$$

9. Disponiendo de la tasa de variación del error respecto al peso de una conexión (EP), podemos usar distintas reglas para modificar ese peso en aras a reducir dicho de error. Una de las primeras reglas que aprovechó este algoritmo, es la regla delta generalizada, que calcula el incremento a aplicar a cada peso como una proporción directa de la tasa de variación del error.

$$\Delta w_{ij} = \eta e_j y_i$$

siendo  $\eta$  el coeficiente de aprendizaje, típicamente con valores comprendidos entre 0.01 y 1.0

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}$$

*Extensiones de la regla delta generalizada.*

La regla DBD (delta-bar-delta) (Jordan, 1988) consiste en usar un coeficiente de aprendizaje propio y variable para cada conexión.

$$\Delta w_{ij} = \eta e_j y_i$$

Una extensión propuesta por Rumelhart, Hinton y Williams (1986) consiste en añadir un término proporcional a la cantidad del último cambio realizado sobre un peso. Al coeficiente que pondera dicha cantidad se le llama momentum.

$$\Delta w_{ij}(t+1) = \eta e_j y_i + c [\Delta w_{ij}(t)]$$

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t+1)$$

La propuesta EDBD (extended delta-bar-delta) (Minaí y Williams, 1990) consiste en añadir el momentum a la regla DBD.

TESIS CON  
FALLA DE ORIGEN

## **CAPITULO II**

Herramientas y metodología propuestas para el análisis, diseño y desarrollo de Almacenes de Datos y Sistemas de Información Estratégica.

TESIS CON  
FALLA DE ORIGEN

## 2.1. Herramientas multidimensionales.

Para que un proyecto de informática pueda ser desarrollado exitosamente, se debe contar con las metodologías y las herramientas de análisis, diseño y desarrollo adecuadas.

El área del análisis y diseño de tipo relacional ha sido claramente definida y desarrollada por diferentes autores que han propuesto elementos útiles para realizar proyectos de este estilo; sin embargo, en el área multidimensional existen pocas opciones y gran parte de ellas no han sido probadas en proyectos reales; por ello, en este trabajo se propone un conjunto de herramientas gráficas para el análisis y diseño multidimensional y una metodología que evidentemente deberá ser probada mediante el desarrollo de un caso práctico.

### Representación de las variables o hechos.

Los hechos o las variables se representan con un rectángulo con doble línea que contiene una descripción en plural de la variable que se desea explotar; por ejemplo, la variable ventas de la figura 2.1.

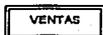


Figura 2.1. Representación gráfica de variables o hechos para un repositorio multidimensional.

### Dimensiones.

Para representar las dimensiones se utiliza un rectángulo con línea simple que contiene una descripción en plural del criterio de explotación que se desea expresar; por ejemplo, la dimensión productos de la figura 2.2.



Figura 2.2. Representación gráfica de dimensiones para un repositorio multidimensional.

### Uniones.

La asociación de las tablas de hechos con las tablas de dimensiones se realizará mediante el uso de flechas cuyo sentido apuntará hacia la tabla de hechos; dichas uniones se representan gráficamente como se muestra en la figura 2.3.



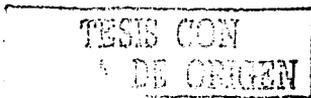
Figura 2.3. Representación gráfica de uniones para asociar tablas de hechos y dimensiones.

### Arquitectura de Estrella.

La arquitectura de una estrella tiene en su parte central las variables y, asociadas a ellas, se encuentran las dimensiones. Un ejemplo de estrella tridimensional se muestra en la figura 2.4.



Figura 2.4. Esquema estrella.



### Arquitectura de estrella compuesta o constelación.

La arquitectura de una estrella compuesta o constelación consta de dos arquitecturas de estrellas válidas que se encuentran conectadas a través de una o más dimensiones. Véase la figura 2.5.

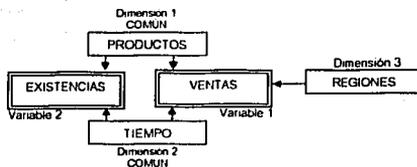


Figura 2.5. Constelación.

### Jerarquía lineal.

La representación de la jerarquía lineal se muestra en la figura 2.6.

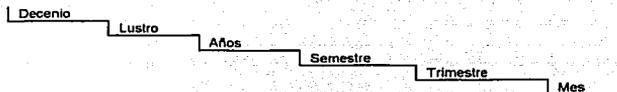


Figura 2.6. Jerarquía lineal.

### Multijerarquía.

La representación de la jerarquía múltiple se muestra en la figura 2.7.

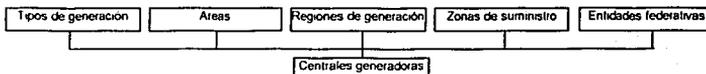


Figura 2.7. Dimensión multijerárquica.

TRABAJOS CON  
ORIGEN

## 2.2. Antecedentes de la metodología propuesta.

El desarrollo de Sistemas de Información Estratégica y Almacenes de Datos, involucra la inversión de una gran cantidad de recursos, por lo que se hace indispensable el uso de metodologías que aseguren el éxito del proyecto.

Actualmente existen ciertas metodologías, como la propuesta por Keneth y Jane Loudon<sup>[1]</sup>, que únicamente sugieren desarrollar Sistemas de Información Estratégica mediante la aplicación de una simple estrategia que propone ir de lo particular a lo general sin dar más detalles de las actividades que deben desempeñarse, ni de los riesgos potenciales que pueden presentarse a lo largo del proyecto. Otras metodologías, como la propuesta por Ralph Kimball y W.H. Inmon<sup>[2]</sup>, que es específica para el análisis, diseño y desarrollo de Almacenes de Datos y, particularmente, aborda los aspectos relativos al modelado de datos multidimensional.

Estando tantos recursos y esfuerzos en riesgo, parece ilógico confiar el éxito del proyecto a estrategias de este estilo.

En este trabajo de tesis, se propone una metodología para el desarrollo de Sistemas de Información Estratégica, la cual se basa en el desarrollo por etapas. La idea principal es que cada etapa integre información relativa a aspectos de la organización no consideradas en las versiones anteriores, que se generen nuevas relaciones, tendencias, etc., que se atiendan nuevas necesidades de los directivos, que se resuelvan posibles problemas de versiones anteriores y que se mejore la eficiencia de todos los procesos involucrados.

Cada etapa o versión debe generar resultados aplicables a la organización, es decir, productos terminados que se pueden complementar y mejorar en las siguientes versiones.

La condición ineludible en esta metodología es la permanente comunicación con los usuarios de los Sistemas de Información Estratégica, que en este caso son los directivos de las organizaciones, quienes en todo momento deben tener la posibilidad de probar lo creado y de proponer nuevos elementos que mejoren el sistema.

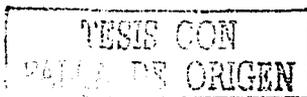
Esta metodología propone que con cada etapa se abra la posibilidad de hacer revisiones tanto de los procesos involucrados, de la información y los medios de acceso a los mismos a fin de refinar sucesivamente todos los aspectos involucrados. La idea es que con cada etapa, el SIE se aproxime cada vez más a ser un reflejo más completo de la organización; por ello, esta metodología se denomina "Aproximaciones Sucesivas".

Los Sistemas de Información Estratégica guardan "fotografías" del estado de las organizaciones a lo largo del tiempo; sin embargo, las organizaciones cambian y, evidentemente el SIE debe evolucionar para continuar siendo útil a la organización. En vista de que la metodología propuesta sugiere el desarrollo por etapas y con refinamientos sucesivos a todos los aspectos involucrados, (como un desarrollo que no concluye nunca), estos constantes cambios son abordados de manera natural por la metodología. Es evidente que cada posible cambio debe ser analizado a detalle y se debe evaluar la conveniencia de hacer los cambios necesarios al sistema actual contra la posibilidad de hacer un nuevo sistema, en todo caso, el grado de evolutividad de un sistema está en función de su diseño y no de la metodología empleada para crearlo.

En términos generales, esta metodología propone que una vez recopilados los requerimientos de información estratégica, se asocian dichos requerimientos al modelo de procesos de la organización y a las fuentes de datos relacionadas con cada aspecto del mismo y que con base a ello se diseñen las vistas que respondan a dichos requerimientos.

[1] "Administración de Sistemas de Información" LOUDON y LOUDON. Ed. Prentice Hall. 1998.

[2] "The Data Warehouse Toolkit" KIMBALL E INMON. Ed. Wiley. 1996.



A partir de las vistas aprobadas se pueden desarrollar las estructuras multidimensionales que muestren el tipo de explotaciones que pueden efectuarse.

Hasta aquí, las estructuras multidimensionales pueden contener datos reales obtenidos a partir de procesos que no se encuentren automatizados; casi siempre en esta etapa se hace uso de archivos planos tomados de las fuentes relacionales de datos.

El siguiente paso consiste en convertir las estructuras multidimensionales en estructuras de estrellas relacionales que junto con otras estructuras forman parte del Almacén de Datos.

A partir de la creación del Almacén se propone el desarrollo de mecanismos para la administración y control de datos, es decir los metadatos.

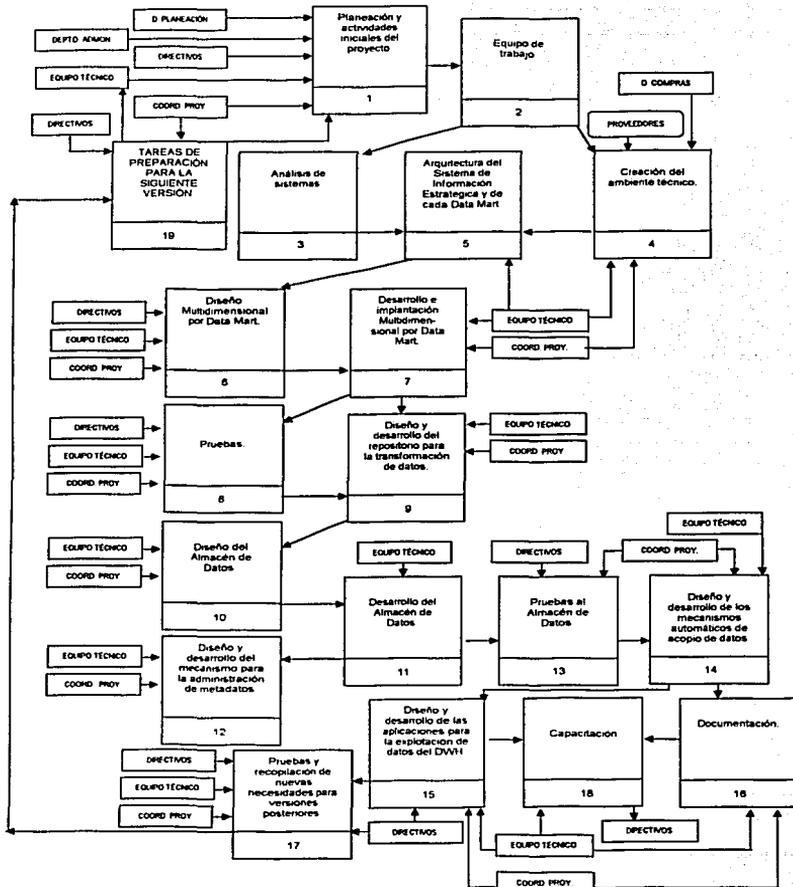
Con mucha frecuencia las fuentes no contienen los datos en los términos, formatos, etc., que el Almacén requiere, por lo que se hace necesario el desarrollo de bases exclusivas para la transformación de datos, lo cual se determina a partir de la asociación de las fuentes con los datos esperados en el Almacén. La siguiente etapa en la metodología sugiere el desarrollo de la base de datos y estructuras para la transformación de datos.

La metodología propone el desarrollo de mecanismos automáticos para la actualización de datos, de la fuente a las estructuras de transformación, de las estructuras de transformación al Almacén y del Almacén a las estructuras multidimensionales.

Posteriormente propone el desarrollo de aplicaciones para explotación de datos y, finalmente, la capacitación y documentación del sistema.

TESIS CON  
FALLA DE ORIGEN

2.3. Diagrama de flujo de datos del ciclo de vida de Aproximaciones Sucesivas para el desarrollo de Sistemas de Información Estratégica.



TESIS CON  
FALLA DE ORIGEN

## 2.4. Descripción de la Metodología de Aproximaciones Sucesivas.

No Actividad	Descripción de Actividad	Actividad que le antecede (obligatoria)	Resultados	Logros
--------------	--------------------------	---	------------	--------

<b>1-Planeación del proyecto y actividades iniciales del proyecto. (NRDM)</b>				
---	--	--	--	--

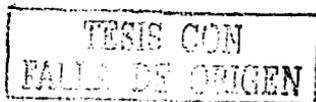
1.0	Actividades de inicio del proyecto			
1.1	Desarrollo del documento de alcances del proyecto	Ninguna	Documento de alcances del proyecto	
1.2	Desarrollo del plan del proyecto	Ninguna	Plan	
1.3	Creación/visión de estándares, reglas y políticas	Ninguna	Estándares, reglas y políticas	
1.4	Identificación de usuarios por grupos	1.2	Lista de distribución	
1.5	Programación de reuniones de trabajo	1.4	Calendario de juntas	
1.6	Revisión del documento del plan y alcances con el grupo de directivos	1.1, 1.2	Plan final	Plan aprobado
1.7	Identificación de los requerimientos generales del sistema	1.6	Requerimientos del sistema	
1.8	Determinación y medición de los requerimientos del sistema	1.7	Determinación y Medición de los requerimientos del sistema	
1.9	Identificación de los requerimientos de seguridad	1.3	Requerimientos de seguridad	
1.10	Desarrollo del documento de requerimientos del sistema	1.7, 1.8, 1.9	Borrador del documento de especificaciones	
1.11	Revisión del documento de especificaciones con el grupo de directivos	1.10	Documento de especificaciones	Documento de especificaciones aprobado
1.12	Desarrollo de las fases del proyecto	1.6, 1.11	Borrador de las fases del proyecto	
1.13	Revisión de las fases con los usuarios finales	1.12	Fases del proyecto	
1.14	Determinación de los perfiles de los integrantes del equipo de trabajo	1.6, 1.11		

<b>A2-Equipo de trabajo.</b>				
------------------------------	--	--	--	--

2.0 NR DM	Conformación del equipo de trabajo y su involucramiento con los aspectos estratégicos de la organización			
2.1 NR DM	Selección y reclutamiento de los integrantes del equipo de trabajo según los perfiles determinados	1.14	Equipo de trabajo seleccionado y reclutado	Equipo de trabajo
2.2	Involucramiento del equipo de trabajo con los aspectos estratégicos de la organización	2.1	Equipo de trabajo involucrado con los aspectos estratégicos de la organización	Equipo de trabajo involucrado

A3-Análisis de sistemas.				
3.1	Identificación de las Especificaciones del modelo de procesos de la organización.		Documento de especificaciones del modelo de procesos	Modelo de procesos
3.2	Identificación de las fuentes de datos asociadas con cada elemento del modelo de procesos.	3.1	Fuentes seleccionadas	Fuentes según el modelo de procesos
3.3	Investigación de las fuentes de datos relativas a cada elemento del modelo de procesos	3.2	Investigación de las fuentes según Data Mart	Ubicación, Plataforma, Responsable, manejador de bd, etc
3.4	Partición del Sistema de Información Estratégica en subsistemas (Data Mart) referentes a cada elemento del modelo de procesos	3.1	Fragmentación de requerimientos según el modelo de procesos	Data Mart identificados
3.5	Selección de los subsistemas (Data Mart) que deberán ser desarrollados en la primera versión según su nivel de importancia en la organización	3.4	Selección de Data Mart para la primera versión	Data Mart para primera versión
3.6	Recopilación de la información relativa al Data Mart a desarrollar	3.5	Información relativa a los Data Mart	
3.7	Análisis de la fuente de datos relativa al Data Mart a desarrollar	3.6	Fuentes analizadas	Formas de acceso, frecuencia de actualización, etc. Diagramas entidad-relación y diccionarios de datos
3.8	Análisis de las necesidades de información del Data Mart seleccionado	3.7	Necesidades de información por Data Mart	Información por Data Mart

A4-Creación del ambiente técnico.				
			NO REQUERIDO EN DM	
4.0	Documento de requerimientos técnicos			
4.1	Identificación del personal de soporte técnico		Lista de distribución	
4.2	Documento de hardware / software/ requerimiento de herramientas	1.7	Requerimientos técnicos	
4.3	Evaluación y selección de vendedores	4.2	Lista de vendedores	
4.4	Negociación de licencias	4.3	Licencias	
4.5	Ordenar hardware / software /herramientas	4.4	Contrato	Orden de Compra
4.6	Instalación de las herramientas de hardware / software /herramientas	4.4	Hardware, software y herramientas instaladas	
4.7	Pruebas técnicas al ambiente	4.6	Pruebas al equipo	Ambiente desarrollado
4.8	Entrenamiento en la tecnología seleccionada	4.7	Equipo entrenado	


  
 TESIS CON  
 FALLAS DE ORIGEN

<b>A5-Arquitectura del Sistema de Información Estratégica y de cada Data Mart.</b>				
--	--	--	--	--

5.1	Documento del diseño y especificaciones de la arquitectura	3.8	Arquitectura del Sistema de Información Estratégica	Documento de Arquitectura
5.2	Documento del diseño y especificaciones de la arquitectura del Data Mart	3.8	Arquitectura del Data Mart	Documento de Arquitectura

<b>A6-Diseño Multidimensional por Data Mart.</b>				
--	--	--	--	--

6.0	Diseño por Data Mart de las vistas multidimensionales que respondan a las necesidades de información	3.8	Borrador del diseño de vistas multidimensionales por Data Mart	
6.1	Sesiones con usuarios finales para la validación de vistas	6.0	Diseño final aprobado de vistas multidimensionales	Vistas multidimensionales por Data Mart
6.2	Diseño por Data Mart de las estructuras multidimensionales según el diseño de vistas	6.1	Diseño de estructuras multidimensionales por Data Mart	Especificación de variables, dimensiones y niveles jerárquicos, modelos, fórmulas, etc.
6.3	Diseño de los repositorios multidimensionales con base a las estructuras diseñadas	6.2	Diseño de repositorios multidimensionales por Data Mart	Diseño de bases de datos multidimensionales

<b>A7-Desarrollo e Implantación Multidimensional por Data Mart.</b>				
---	--	--	--	--

7.0	Creación de bases de datos multidimensionales	6.3	Bases de datos multidimensionales	
7.1	Creación de dimensiones	7.0	Dimensiones	
7.2	Cargas de datos en dimensiones	7.1	Dimensiones con datos	
7.3	Creación de variables	7.2	Variables	
7.4	Carga de datos en variables por el método de creación de archivos planos	7.3	Variables con datos	
7.5	Creación de otros objetos	7.0	Otros objetos	
7.6	Creación de programas de agregación de variables	7.3	Programas de agregación (roll up)	
7.7	Ejecución de programas de agregación de variables	7.2, 7.3	Variables agregadas	Variables relacionadas con dimensiones apropiadas y agregadas

<b>A8-Pruebas a la base de datos multidimensional.</b>				
--	--	--	--	--

8.0	Reuniones con usuarios para la revisión de las bases de datos multidimensionales	7.7	Reuniones completadas	
8.1	Adecuaciones a las bases de datos según a las observaciones de usuarios	8.0	Adecuaciones efectuadas	
8.2	Firma de documento de validación de las bases de datos multidimensionales	8.1	Bases de datos validadas	

<b>A9- Diseño y desarrollo del repositorio relacional para la transformación de datos (OPCIONAL).</b>				
---	--	--	--	--

9.0	Creación de la base de datos relacional para la transformación de datos	8.1, 8.3	Base de datos relacional para transformación de datos	
9.1	Creación de las estructuras relacionales para la transformación de datos	8.4, 9.0	Estructuras de datos sobre la base de datos relacional	

<b>A10-Diseño del Almacén de Datos.</b>				
---	--	--	--	--

10.0	Diseño del Almacén de Datos.	7.7	Diseño del Almacén de Datos	
10.1	Diseño de las estructuras relacionales según el diseño propuesto	10.0	Diseño de estructuras	Diagramas entidad-relación y diccionarios de datos

<b>A11- Desarrollo del Almacén de Datos.</b>				
--	--	--	--	--

11.0	Creación del Almacén de Datos	10.0, 10.2	Almacén de Datos	
11.1	Creación de las estructuras relacionales para el Almacén de Datos	11.0	Estructuras de datos sobre la base de datos relacional	Almacén de Datos

<b>A12-Diseño y desarrollo del mecanismo para la administración de metadatos.</b>			NO REQUERIDO EN DM	
---	--	--	--------------------	--

12.0	Diseño del mecanismo para el control de metadatos	11.0	Diseño del mecanismo para la administración de metadatos	
12.1	Desarrollo del mecanismo para el control de metadatos	12.0	Desarrollo del mecanismo para la administración de metadatos	
12.2	Diseño del repositorio para el almacenamiento de metadatos	11.0	Diseño del repositorio de metadatos	
12.3	Creación del repositorio para el almacenamiento de metadatos	12.2	Repositorio de metadatos	Metadatos
12.4	Pruebas al mecanismo de administración de metadatos	12.1, 12.3	Mecanismo para la administración de metadatos aprobado	Mecanismo probado

TESIS CON  
 DE ORIGEN

<b>A13-Diseño y desarrollo de los mecanismos automáticos para acopio de datos.</b>				
13.0	Diseño de programas para la transferencia de datos	7.7, 10.1	Programas diseñados	
13.1	Diseño de los programas para la conexión y carga de datos de ROLAP a MOLAP y agregación de datos	7.7, 10.1	Programas diseñados	
13.2	Desarrollo de los programas para la conexión y carga de datos de ROLAP a MOLAP y agregación de datos	7.7, 10.1	Programas desarrollados	
13.3	Diseño de los mecanismos automáticos para acopio de datos por Data Mart	3.3, 3.7, 7.7, 10.1	Diseño de mecanismos automáticos para acopio de datos	
13.4	Desarrollo de los mecanismos automáticos para acopio de datos por Data Mart	13.3	Mecanismos automáticos para acopio de datos	Mecanismos automáticos para acopio de datos
13.5 NR DM	Diseño del tablero integrador de control para la gestión y administración centralizada de procesos	13.4	Diseño del tablero integrador de control	
13.6 NR DM	Desarrollo del tablero integrador de control para la gestión y administración centralizada de procesos	13.5	Tablero integrador de control	Tablero integrador de control

<b>A14-Diseño y desarrollo de las aplicaciones para la explotación de datos del Almacén de Datos del Sistema de Información Estratégica .</b>				
---	--	--	--	--

14.0	Aplicaciones para la explotación de datos.			
14.1	Diseño de las aplicaciones para la explotación de datos.	7.7, 10.1	Aplicaciones para la explotación de datos diseñadas	
14.2	Desarrollo de las aplicaciones para la explotación de datos.	14.0	Aplicaciones para la explotación de datos	

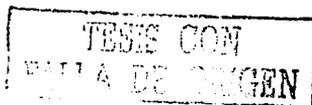
<b>A15-Capacitación.</b>			NO REQUERIDO EN DM	
--------------------------	--	--	--------------------	--

15.0	Capacitación a desarrolladores	14.4	Desarrolladores capacitados	
15.1	Capacitación a analistas	14.4	Analistas de datos capacitados	
15.2	Capacitación a directivos	14.4	Directivos capacitados	

<b>A16-Documentación.</b>			NO REQUERIDO EN DM	
---------------------------	--	--	--------------------	--

16.0	Especificaciones de documentación	14.4	Especificaciones de documentación	
------	-----------------------------------	------	-----------------------------------	--

Nota: NRDM o No requerido en Data Mart: Aspecto que puede no ser abordado en el caso específico del desarrollo de un Data Mart.



## **CAPITULO III**

**Aplicación de la metodología y herramientas propuestas para el análisis y diseño del Sistema de Información Estratégica de la Comisión Federal de Electricidad.**

TESIS CON  
FALTA DE ORIGEN

### 3.1. Aspectos generales de la Comisión Federal de Electricidad (CFE).

#### Antecedentes.

La empresa se creó en 1937 con el objeto de generar energía eléctrica para abastecer a un mercado en crecimiento, satisfacer la demanda de los consumidores, planear e integrar el servicio eléctrico en México y preparar un esquema que le diera a la nación el control sobre sus recursos energéticos.

#### Misión.

La misión de esta empresa es satisfacer la demanda de energía eléctrica del país tomando en cuenta normas de calidad que le permitan asegurar la continuidad en el servicio.

#### Objetivos estratégicos de la empresa.

Para desarrollar sus tareas, la empresa cumple una clara misión buscando alcanzar objetivos estratégicos sustentados en valores corporativos, esto es:

- Asegurar el suministro de energía eléctrica en el país, en condiciones adecuadas de cantidad, calidad y precio.
- Proteger el ambiente, promover el desarrollo social y respetar los valores de las poblaciones donde se ubican las obras de electrificación.
- Proporcionar atención esmerada a sus clientes.

#### Funciones de la empresa.

Esta empresa paraestatal tiene como funciones principales generar, transmitir, distribuir y comercializar la energía eléctrica que consume el país.

Su misión ha motivado que esta empresa procure estar siempre en la búsqueda de los mecanismos necesarios para ser cada vez más productiva, con el fin de satisfacer la creciente demanda de energía eléctrica.

#### Organigrama de la empresa.

Para el desarrollo de sus funciones, la empresa cuenta con la estructura organizacional mostrada en la figura 3.1 para el Control, Evaluación y Administración de la Generación de energía.



3.1. Organigrama.

TESIS CON  
FALLA DE ORIGEN

### *Elementos funcionales de la empresa.*

La empresa esta formada por una serie de elementos que le permiten cumplir con su misión; al conjunto de estos elementos se le conoce como sistema de energía eléctrica.

A continuación se presentará una descripción de los elementos que forman los sistemas de energía eléctrica.

#### *Generación.*

Se realiza en plantas generadoras que aprovechan la energía de la naturaleza, tales como: caídas de agua, mareas, viento, ríos y luz solar; el uso de combustibles fósiles tales como: petróleo, gas natural, carbón o la fusión nuclear. En la mayoría de las ocasiones estas plantas se encuentran distantes de los centros de carga, por lo que se requieren sistemas de transmisión en alta tensión para transportar la energía.

#### *Transmisión.*

A la salida de los generadores, la tensión de la energía eléctrica es elevada mediante el uso de subestaciones transformadoras elevadoras, lo que permite que las líneas de transmisión puedan llevar grandes potencias a enormes distancias con pérdidas relativamente pequeñas. A lo largo del sistema se disponen estaciones de maniobra que hacen posible la interconexión de distintas líneas para formar sistemas más grandes y seguros. Finalmente las subestaciones transformadoras reductoras, que representan la carga del sistema de transmisión, alimentan a los sistemas de distribución.

#### *Distribución.*

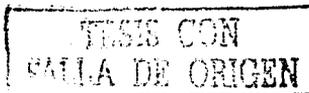
Los sistemas de distribución se representan de manera similar a los de transmisión, pero la tensión que manejan es mucho más pequeña, transportan la energía hasta las acometidas de los usuarios. Estos sistemas cuentan con líneas o redes primarias y secundarias así como bancos de transformadores.

Los sistemas de distribución están constituidos por un gran número de cargas individuales de diferentes clases (comercial, industrial, doméstica, etc.). La potencia suministrada en cada instante por un sistema eléctrico completo es la suma de todas las cargas individuales en ese momento y de las pérdidas que se presenten en el sistema.

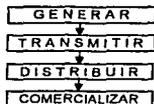
La generación de energía eléctrica, se realiza por medio de diferentes tipos de tecnologías, desde las plantas hidroeléctricas y termoeléctricas, hasta modernas plantas de energía solar, eólica y nuclear.

Las plantas generadoras de energía eléctrica se encuentran distribuidas en todo el país. El control de todas las plantas se realiza de manera centralizada en el Centro Nacional de Control de Energía, (CENACE). Dicho control se efectúa con el fin de regular la generación y los intercambios de energía de las plantas generadoras con base a la demanda.

De todos los procesos que se llevan a cabo en la empresa generadora se desprenden datos valiosos que pueden ser de utilidad a los directivos para fundamentar sus decisiones en datos objetivos derivados de la operación de la empresa; sin embargo, la recopilación de la información para realizar informes ejecutivos en las circunstancias actuales todavía implica la inversión de una gran cantidad de recursos.

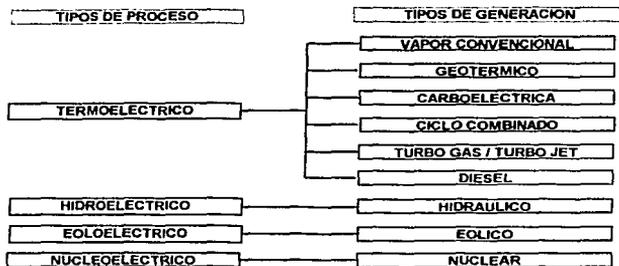


**Funciones.**



3.2. Funciones.

**Procesos y Tipos de Generación de la empresa.**

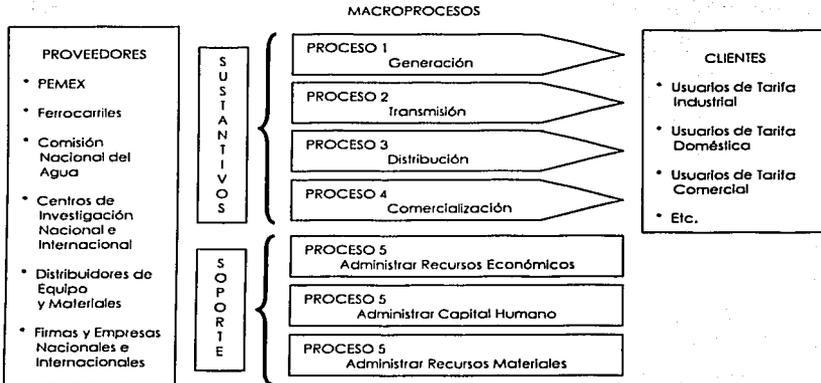


3.3. Procesos y tipos de generación.

En nuestro país, las plantas más importantes son de tipo termoeléctrico e hidroeléctrico, aunque existe el caso de la planta de Laguna Verde que genera energía eléctrica mediante el uso de energía nuclear.

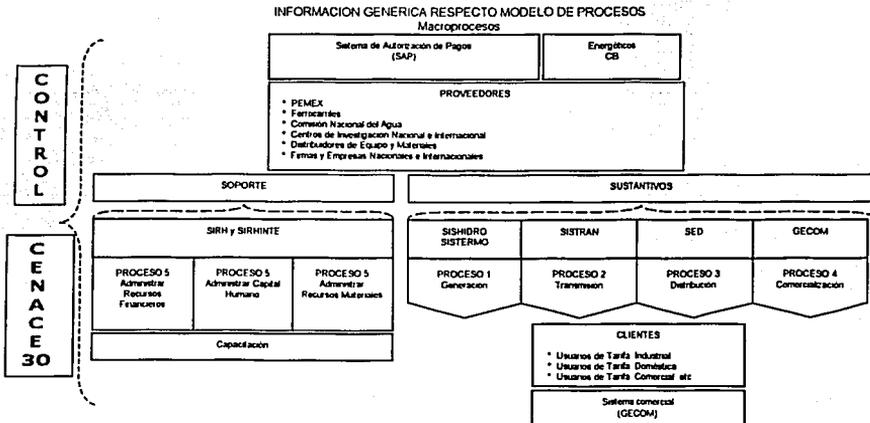
TESIS CON  
FALLA DE ORIGEN

### 3.2. Especificación del Modelo de Procesos de la CFE.



3.4. Gráfica del Modelo de Procesos de la empresa

### 3.3. Especificación de las fuentes de información con base al modelo de procesos.



3.5. Gráfica de requerimientos con base al modelo de procesos de la empresa.

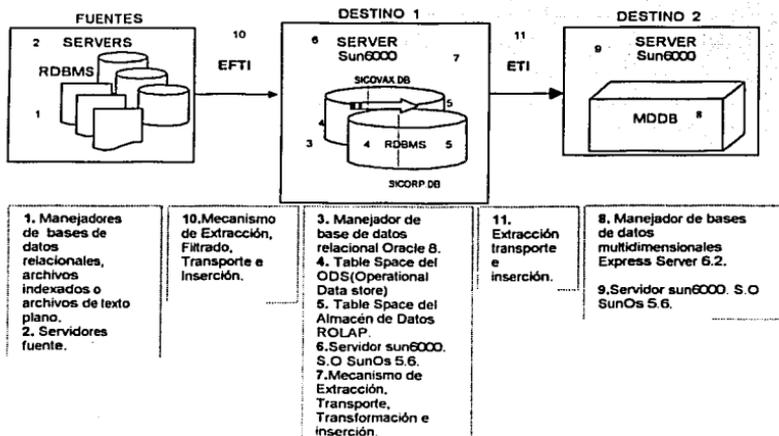
TRABAJO CON  
 FALLA DE ORIGEN

### 3.4. Información general de las fuentes de información representativas de cada proceso.

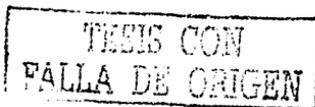
ASPECTOS	FUENTE	BASE DE DATOS	SISTEMA OPERATIVO
Datos relativos a energeticos	cb	BD informix 7 0 SE	SOLARIS
Datos relativos a la comercialización de energía eléctrica	geom	BD informix 7 0 ON LINE	SOLARIS
Datos del padrón de usuarios	geom	BD informix 7 0 ON LINE	SOLARIS
Datos relativos al control de generación de energía eléctrica	cenace30	BD informix 7 0 SE	SOLARIS
Datos relativos a la distribución de energía eléctrica	sed	BD informix 7 0 SE	SOLARIS
Datos relativos a las plantas hidroeléctricas	sishidro	Archivos indexados	VMS
Datos relativos a las plantas termoelectricas	sistermo	Archivos indexados	VMS
Datos relativos al uso de agua en plantas generadoras	agua	Archivos indexados	VMS
Datos relativos a los pagos por transporte de combustible (transporte marítimo y terrestre)	sap	Archivos indexados	VMS
Datos relativos a la transmisión de energía eléctrica	Sisiran	BD informix 7 0 SE	SOLARIS
Datos relativos a Finanzas y Administración	sirhinte, sirh	BD informix 5 0 SE	AIX
Datos relativos a Capacitación del personal de la empresa	capacitacion	BD Fox	Windows 98

Tabla 1. Reporte de las fuentes de información representativas de cada proceso

### 3.5.Arquitectura del Sistema de Información estratégica de la CFE.



### 3.6.Arquitectura del sistema de información estratégica.



### 3.6. Organización de recursos informáticos.

A continuación se presenta una revisión de los recursos informáticos:

- Organización de bases de datos.  
Se crean dos bases de datos relacionales: sicorp.db y sicovax.db. En la primera se almacena la información que procede de fuentes de informix y fox y en la segunda, la información que procede de archivos indexados.
- Organización de Table Spaces, Data Files y Rblogs.

Table Space	Parámetros	Datafiles	Tamaño	Tipo
Control	Default	/sicorp/sicorpcontrol.dbf	124000K	Permanent
Controlidx	Default	/indices/sicorp/controlidx.dbf	204800K	Permanent
Energéticos	Default	/sicorp/energéticos.dbf	1024M	Permanent
Energéticosidx	Default	/indices/sicorp/energéticos.dbf	1024M	Permanent
Comercial	Default	/sicorp/comercial.dbf	3G	Permanent
Comercialidx	Default	/indices/sicorp/comercialidx.dbf	3G	Permanent
Padron	Default	/sicorp/sicorp/Padron.dbf	1536M	Permanent
Padronidx	Default	/indices/sicorp/Padronidx.dbf	10M	Permanent
Distribucion	Default	/sicorp/sicorp/distribucion.dbf	2G	Permanent
Distribucionidx	Default	/indices/sicorp/distribucionidx.dbf	2G	Permanent
Capacitaciones	Default	/sicorp/sicorp/capacitaciones.dbf	1536M	Permanent
Capacitacionesidx	Default	/indices/sicorp/capacitacionesidx.dbf	150M	Permanent
Temp1	Initial=2064K Next=2064K Increase=0% Min number=1 Max number=512	/temporal/sicorp/temp1.dbf	1536000K	Temporal
rhs	Initial=512K Next=512K Increase=0% Min number=5 Max number=1024	/rhsbgs/sicorp/rhs1.dbf	256000K	Permanent

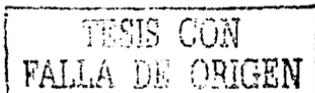
Tabla 2. Table Spaces y Data Files de la base de datos sicorp.bd

Table Space	Parámetros	Datafiles	Tamaño	Tipo
Sap	Default	/sicorp_2/sicovax/Sap.dbf	1024M	Permanent
Sapidx	Default	/indices/sicovax/Sapidx.dbf	1024M	Permanent
Sshdro	Default	/sicorp_2/sshdro.dbf	3G	Permanent
Sshdroidx	Default	/indices/sicovax/sshdroidx.dbf	3G	Permanent
Sistema	Default	/sicorp_2/sistema.dbf	300M	Permanent
Sistemidx	Default	/indices/sicovax/sistemidx.dbf	1536M	Permanent
Agua	Default	/sicorp_2/sicovax/agua.dbf	150M	Permanent
Aguaidx	Default	/indices/sicovax/aguaidx.dbf	1536M	Permanent
Temp1	Initial=2064K Next=2064K Increase=0% Min number=1 Max number=512	/temporal/sicovax/temp1.dbf	1536000K	Temporal
rhs	Initial=512K Next=512K Increase=0% Min number=5 Max number=1024	/rhsbgs/sicovax/rhs1.dbf	256000K	Permanent

Tabla 3. Table Spaces y Data Files de la base de datos sicovax.bd

Valores default	
sicorp	Initial=80K Next=80K Increase=50% Min number=1 Max number=1017
sicovax	Initial=40K Next=40K Increase=50% Min number=1 Max number=505

Tabla 4. Valores de default de ambas bases de datos.



Nombre	Talbespace	Status	Parámetros
RB1	rbs1	Online	Initial= 512K, Next= 512K, optimal= null, Min= 5, Max= 512
RB2	rbs1	Offline	Initial= 512K, Next= 512K, optimal= null, Min= 5, Max= 512
RB3	rbs1	Offline	Initial= 512K, Next= 512K, optimal= null, Min= 5, Max= 1024
RB4	rbs1	Offline	Initial= 512K, Next= 512K, optimal= null, Min= 5, Max= 512

Tabla 5. Rblogs la base de datos sicorp db

Nombre	Talbespace	Status	Parámetros
RB1	rbs1	Online	Initial= 512K, Next= 512K, optimal= null, Min= 5, Max= 512
RB2	rbs1	Online	Initial= 512K, Next= 512K, optimal= null, Min= 5, Max= 512
RB3	rbs1	Online	Initial= 512K, Next= 512K, optimal= null, Min= 5, Max= 1024

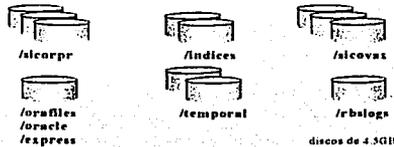
Tabla 6. Rblogs la base de datos sicovax db

- Especificación del servidor que contendrá el Almacén de Datos.

Servidor	sun6000
Sistema operativo	SunOs 5.6
Número de discos	12
Capacidad de disco	4.5 Gb
Espacio total en disco	54 Gb

Tabla 7. Datos del servidor del DWH.

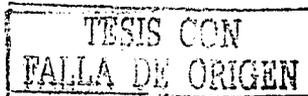
- El número de discos a utilizar por cada filesystem, se muestra en la figura 3.7 y 3.8



3.7. Número de discos por filesystem.



Figura 3.8. Distribución de discos duros en filesystem.



### 3.7. Frecuencia de actualización de los Sistemas Transaccionales.

- Cargas históricas de datos.

Las fuentes de datos han operado durante algunos años, por lo que es necesario considerar la recuperación de dicha información para ser guardada en el Almacén de Datos.

A continuación se presenta una tabla que reporta la antigüedad de los sistemas:

SISTEMAS FUENTE	DATOS DESDE
ENERGÉTICOS	1970
COMERCIAL	1980
PADRON DE USUARIOS	1980
CONTROL	1975
DISTRIBUCION	1985
CAPACITACION	1990
SISHIDRO	1970
SISTERMO	1970
AGUA	1970
AUTORIZACION DE PAGOS	1978
RECURSOS HUMANOS	1992

Tabla 8. Antigüedad de los Sistemas Transaccionales.

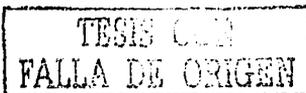
En el caso del Data Mart de Energéticos se consideran datos desde 1970.

- Cargas automáticas de datos mensuales.

Una vez cargados los datos históricos, la actualización de la información tendría que hacerse de forma mensual y automática. Las fechas necesarias para la actualización de los datos son:

SISTEMAS FUENTE	FECHA DE ACTUALIZACIÓN
ENERGÉTICOS	DÍA 10 DE CADA MES
COMERCIAL	DÍA 15 DE CADA MES
PADRON DE USUARIOS	DÍA 15 DE CADA MES
CONTROL	DÍA 20 DE CADA MES
DISTRIBUCION	DÍA 10 DE CADA MES
CAPACITACION	DÍA 10 DE CADA MES
SISHIDRO	DÍA 15 DE CADA MES
SISTERMO	DÍA 15 DE CADA MES
AGUA	DÍA 10 DE CADA MES
AUTORIZACION DE PAGOS	DÍA 20 DE CADA MES
RECURSOS HUMANOS	DÍA 15 DE CADA MES

Tabla 9. Fechas de actualización de los Sistemas Transaccionales.



### 3.8. Cuentas y permisos.

- Usuario de consulta.

Este tipo de cuenta le permite a los usuarios realizar consultas a los datos mediante el uso de PL-SQL o bien mediante alguna aplicación específica.

- Usuario para ejecución de mecanismos automáticos para adquisición de datos.

Los mecanismos automáticos para adquisición de datos requieren cuentas con privilegios de lectura y escritura sobre los tablespaces específicos de cada Data Mart.

- Usuario con permisos de administración.

Se recomienda crear un rol que permita a los administradores controlar los recursos asignados a cada subsistema.

Relación de cuentas creadas por Data Mart y por tipo de usuario.

DATA MART	MANEJADOR DE BD	DB FUENTE	CUENTA ORACLE & INSERTORES	PERMISOS	NOMBRE TABLESPACE	TAMAÑO TABLESPACE EN GB/TES	CUENTA ORACLE & EXPRESS	PERMISOS
SICORP								
ENERGÉTICOS	INFORMIX	CB	L_cb P_cb	Connect y Resource	COMBUSTIBLES	1	L_ex_cb P_ex_cb	Connect
COMERCIAL	INFORMIX	GECOM	L_comercial P_comercial	Connect y Resource	COMERCIAL	3	L_ex_comercial P_ex_comercial	Connect
PADRON DE USUARIOS	INFORMIX	GECOM	L_padron P_padron	Connect y Resource	PADRON	15	L_ex_padron P_ex_padron	Connect
CONTROL	INFORMIX	INF CLNAL	L_control P_control	Connect y Resource	CONTROL	1	L_ex_control P_ex_control	Connect
DISTRIBUCIÓN	INFORMIX	SED	L_sed P_sed	Connect y Resource	DISTRIBUCION	2	L_ex_sed P_ex_sed	Connect
CAPACITACION	FOX	SIC	L_capacita P_capacita	Connect y Resource	CAPACITACION	15	L_ex_capacita P_ex_capacita	Connect
SUBTOTAL						10		
SICOVAX								
SISHIDRO	BASIC	SISHIDRO	L_sishidro P_sishidro	Connect y Resource	SISHIDRO	3	L_ex_sishidro P_ex_sishidro	Connect
SISTERMO	BASIC	SISTERMO	L_sistermo P_sistermo	Connect y Resource	SISTERMO	3	L_ex_sistermo P_ex_sistermo	Connect
AGUA	BASIC	AGUA	L_agua P_agua	Connect y Resource	AGUA	15	L_ex_agua P_ex_agua	Connect
AUTORIZACION DE PAGOS	BASIC	SAP	L_sap P_sap	Connect y Resource	SAP	1	L_ex_sap P_ex_sap	Connect
SIMO	INFORMIX DOS	SIMO	L_simo P_simo	Connect y Resource	SIMO	15	L_ex_simo P_ex_simo	Connect
SUBTOTAL						10		

Tabla 10. Información general de Data Mart creados.

TESIS CON  
FALLA DE ORIGEN

## **CAPITULO IV**

**Aplicación de la metodología para el desarrollo del Data Mart del área de Energéticos.**

TESIS CON  
FALLA DE ORIGEN

#### 4.1. Recopilación de la información relativa al área de Energéticos.

Como se mencionó previamente, la información contenida en este trabajo de tesis, será relativa al Data Mart de Energéticos. Para este aspecto, se encontró solo un sistema fuente dentro de la CFE que cuenta con los datos necesarios para conocer los aspectos relevantes de esta área. El sistema fuente se conoce con el nombre de combustibles, el cual es un sistema desarrollado en delphi 4 y cuya base de datos operacional se encuentra en informix y su nombre es el de CB.db.

#### 4.2. Análisis de la fuente de información.

##### *Generalidades.*

A fin de que la Gerencia de Energéticos cuente con mecanismos para mantener la Información sobre el consumo, existencia, recepción, derrame y traspaso de combustibles en las diferentes Centrales de la CFE, fue desarrollado el Sistema de Combustibles (CB).

Una de las ventajas del sistema es que se puede acceder a él prácticamente desde cualquier parte del país donde haya acceso a la Intranet de la empresa a través del Emulador Visual Appel.

CB es un sistema de adquisición, procesamiento y explotación de datos generados en todas las Centrales de Generación termoelectrónica de la CFE.

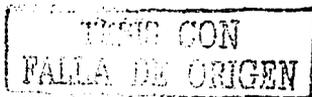
A través del sistema, se integran los resultados de las centrales por Gerencias Regionales y permite la explotación de esta información a la Gerencia de Energéticos. La información del sistema es también utilizada por Contabilidad, Abastecimientos, Energéticos y CENACE. La información también fluye hacia el exterior a las Secretarías de Energía y Minas, y Hacienda.

##### *Introducción al sistema.*

El objetivo principal del sistema CB es, proporcionar un medio para mantener información sobre el consumo, existencia, recepción, derrame y traspaso de combustibles en las diferentes Centrales Generadoras.

Todos los días las Centrales, a través del CB, capturan información relativa al consumo, existencia y recepción de combustóleo, gas, diesel y carbón. Mensualmente, la Gerencia de Energéticos, a través de su Departamento de Información y Estadística, realiza cálculos para agregar y acumular esta información por Gerencias Regionales y Areas de Control y genera reportes para explotar la información de esta base de datos. Este sistema, también lleva el registro de los precios de combustible y transporte, y la generación diaria de energía por cada central.

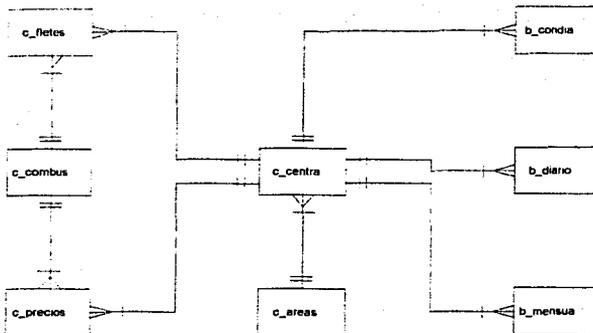
Este acervo de datos ha integrado y sigue integrando una parte importante de la historia del consumo, existencia y recepción de combustible en cada área, por lo que se ha decidido extraer la información contenida en su base de datos para llevarlos al Almacén de Datos.



### Análisis de datos del sistema transaccional de Energéticos.

En esta sección Se realiza una revisión general de los datos contenidos en el sistema transaccional del Área de Energéticos, en vista de que constituyen la plataforma de información del Data Mart de esta Área.

- Diagrama entidad-relación.



#### 4.1. Modelo Entidad-Relación de las estructuras de la base de datos transaccional de energéticos CB.db

- Diccionario de datos.

Tabla: c_precios				
Descripción: Esta tabla contiene los precios de los combustibles				
Atributo	Descripción	Tipo de dato	Domínio	Llave
c_centra	Clave de identificación de la central	Char(3)	Debe existir en c_centra	1, f(c_centra)
fe_vigen	Fecha de vigencia (mm/dd/aaaa)	Date		1
c_combu	Clave de identificación del combustible	Char(1)	Debe existir en c_combu	1, f(c_combu)
cl_proce	Clave de procedencia 1) Nacional 1 Impuesto	Char(1)	N, 1	1
fe_modif	Fecha de modificación (mm/dd/aaaa)	date		
nu_preci	Precio	float		

Tabla c_centra				
Descripción: Catálogo de centros de CIE				
Atributo	Descripción	Tipo de dato	Dominio	Llave
cl_cent	Clave de identificación de la central	Char(3)		1
fe_crea	Fecha de creación (mm/dd/aaaa)	date		
fe_modif	Fecha de modificación (mm/dd/aaaa)	date		
hr_modif	Hora de modificación	Char(4)		
nu_unidad	Numero de unidades	char(1)		
cl_cent	Nombre de la central	Char(40)		
cl_arsa	Clave de identificación del área	Char(5)	Debe existir en c_arsa	f(c_arsa)
cl_zsuna	Clave de identificación de la zona de suministro	Char(5)	Debe existir en c_megan	f(c_region)
cl_regen	Clave de identificación de la región generadora a la que pertenece la planta	Char(5)		
cl_estado	Clave de identificación de la entidad Federativa 01 Agra. 02 BC Nte. 03 BC Sur. 04 Campeche. 05 Chiapas. 06 Chih. 07 Coahuila. 08 Colima. 09 D.F. 10 Durango. 11 Gto. 12 Guerrero. 13 Hidalgo. 14 Jalisco. 15 Mexico. 16 Mich. 17 Morelos. 18 Nayarit. 19 Nvo Leon. 20 Oaxaca. 21 Puebla. 22 Qro. 23 Q Roo. 24 S.L.P. 25 Sonora. 26 Sonora. 27 Tabasco. 28 Tama. 29 Tlaxcala. 30 Veracruz. 31 Yucatán. 32 Zac.	Char(2)	01-32	
ca_tipo	Clave de identificación del tipo de central 1 Vapor. 2 Turbo gas. 3 Carbon. 4 Ciclo combinado 5 Combustion interna	char(1)	1-5	
ca_tcom1	Identificador del manejo o uso de combustóleo. Nulo indica que no se usa	Char(1)		
ca_tcom2	Identificador del manejo o uso de gas. Nulo indica que no se usa	Char(1)		
ca_tcom3	Identificador del manejo o uso de diesel. Nulo indica que no se usa	Char(1)		
ca_tcom4	Identificador del manejo o uso de carbon. Nulo indica que no se usa	Char(1)		
cl_zoom	Clave de identificación de la zona de combustóleo 1 Yucatán. 2 resto	char(1)	1, 2	
cl_zona	Clave de identificación de la zona de diesel	Char(5)		
cl_zogas	Clave de identificación del sector de gas	Char(5)		
nu_potm	Potencia real	integer		
nu_potp	Potencia de placa	integer		
nu_vcom	Volumen total de combustóleo	Fbat		
nu_vcomn	Volumen total de combustóleo	Fbat		
nu_vcomc	Volumen critico de combustóleo	Fbat		
nu_vcomm	Volumen muerto de combustóleo	Fbat		
nu_vdie	Volumen total de diesel	Fbat		
nu_vdiu	Volumen útil de diesel	Fbat		
nu_vdic	Volumen critico de diesel	Fbat		
nu_vmde	Volumen muerto de diesel	Fbat		
nu_vcar	Volumen total de carbon	Fbat		
nu_vcarn	Volumen útil de carbon	Fbat		
nu_vcarc	Volumen critico de carbon	Fbat		
nu_vcaro	Volumen muerto de carbon	Fbat		

Tabla b_centra				
Descripción: Esta tabla contiene el registro diario por unidad de planta				
Atributo	Descripción	Tipo de dato	Dominio	Llave
cl_cent	Clave de la central	Char(3)	Debe existir en c_centra	1, f(c_centra)
fe_regis	Fecha de registro (mm/dd/aaaa)	date		1
nu_unidad	Numero de unidad	char(1)		
nu_cocan	Consumo de combustóleo nacional	integer		
nu_cocoi	Cantidad de combustóleo importado	integer		
nu_cocgn	Consumo de gas nacional	integer		
nu_cocgi	Consumo de gas importado	integer		
nu_cocdi	Consumo de diesel nacional	integer		
nu_cocdi	Consumo de diesel importado	integer		
nu_cocan	Consumo de carbon nacional	integer		
nu_cocoi	Consumo de carbon importado	integer		

Tabla c_arsa:				
Descripción: Esta tabla contiene el catálogo de áreas de control de CIE				
Atributo	Descripción	Tipo de dato	Dominio	Llave
cl_arsa	Clave de identificación del área	Char(5)		1
ds_arsa	Descripción del área	Char(70)		

TESIS CON  
FALLA DE ORIGEN

Tabla c_filios				
Descripción: Esta tabla contiene los filios de los combustibles				
Atributo	Descripción	Tipo de dato	Dominio	Llave
c_centr	Clave de identificación de la central	Char(3)	Debe existir en c_centra	1, f(c_centra)
fu_vigen	Fecha de vigencia (mm/dd/aaaa)	date		1
c_combu	Clave de identificación del combustible	Char(1)	Debe existir en c_combu	1, f(c_combu)
c_tipo	Tipo de filio (M: Marino, T: Terrestre)	char(1)	M, T	1
nu_proco	Proceso del filio	float		

Tabla c_region				
Descripción: Esta tabla contiene el catálogo de regiones de CFE				
Atributo	Descripción	Tipo de dato	Dominio	Llave
c_ries	Clave de identificación de la región	char(5)		1
ct_ritra	Descripción de la región	char(10)		

Tabla c_combus				
Descripción: Esta tabla contiene el catálogo de los combustibles				
Atributo	Descripción	Tipo de dato	Dominio	Llave
c_combu	Clave de identificación del combustible	Char(1)		1
ds_combu	Descripción del combustible	Char(25)		

Tabla b_dimo				
Descripción: Esta tabla contiene el registro diario de consumo, existencia y recepción de combustible por central				
Atributo	Descripción	Tipo de dato	Dominio	Llave
c_centr	Clave de la central	Char(3)	Debe existir en c_centra	1, f(c_centra)
fu_regis	Fecha de registro (mm/dd/aaaa)	date		1
ca_poca1	Poder calorífico del combustible	F:float		
ca_poca2	Poder calorífico del gas	F:float		
ca_poca3	Poder calorífico del diesel	F:float		
ca_poca4	Poder calorífico del carbón	F:float		
nu_cconr	Consumo de combustible nacional	Integer		
nu_coca	Consumo de combustible importado	Integer		
nu_cngan	Consumo de gas nacional	Integer		
nu_coga	Consumo de gas importado	Integer		
nu_codr	Consumo de diesel nacional	Integer		
nu_codi	Consumo de diesel importado	Integer		
nu_cocan	Consumo de carbón nacional	Integer		
nu_coca	Consumo de carbón importado	Integer		
nu_cconr	Existencia de combustible nacional	Integer		
nu_coca	Existencia de combustible importado	Integer		
nu_cngan	Existencia de gas nacional	Integer		
nu_coga	Existencia de gas importado	Integer		
nu_cngan	Existencia de carbón nacional	Integer		
nu_codi	Existencia de carbón importado	Integer		
nu_cconr	Recepción de combustible nacional	Integer		
nu_coca	Recepción de combustible importado	Integer		
nu_cngan	Recepción de gas nacional	Integer		
nu_codi	Recepción de gas importado	Integer		
nu_cngan	Recepción de diesel nacional	Integer		
nu_codi	Recepción de diesel importado	Integer		
nu_cngan	Recepción de carbón nacional	Integer		
nu_codi	Recepción de carbón importado	Integer		
nu_traco	Traspaso de combustible	Integer		
nu_tradi	Traspaso de diesel	Integer		
nu_traca	Traspaso de carbón	Integer		
nu_denco	Overrime de combustible	Integer		
nu_dendi	Overrime de diesel	Integer		
nu_denci	Overrime de carbón	Integer		
nu_gencn	Generación diaria de la central	Integer		

TESIS CON  
FALLA DE ORIGEN



### 4.3. Diseño de vistas multidimensionales que respondan a las necesidades de información corporativa.

A continuación se muestran las vistas aprobadas por los ejecutivos de la CFE.

Vista 1. Consumo de combustibles.

+ TIEMPO			
CONSUMO DE COMBUSTIBLES			
+ CENTRALES	+ TIPO DE COMBUSTIBLES		

Vista 2. Calor en Kcal\*10<sup>10</sup>

+ TIEMPO			
CALOR [Kcal*10 <sup>10</sup> ]			
+ CENTRALES	+ TIPO DE COMBUSTIBLES		

Vista 3. Generación en KWH.

+ TIEMPO			
GENERACION [KWH]			
+ CENTRALES			

Vista 4. Régimen térmico.

+ TIEMPO			
REGIMEN TERMICO			
+ CENTRALES			

Vista 5. Precio ponderado.

+ CENTRALES			
PRECIO PONDERADO			
+ TIEMPO	+ TIPO DE COMBUSTIBLES		

Vista 6. Importe en millones de pesos

+ TIEMPO			
IMPORTE (MILLONES DE PESOS)			
+ CENTRALES	+ TIPO DE COMBUSTIBLES		

TESTE CON  
FALLA DE ORIGEN

#### 4.4. Diseño de las estructuras multidimensionales.

Identificación de las variables (o hechos).

Variable	Descripción
Consumo	Consumo de combustible
Calor	Calor del tipo de combustible por central
Generación	Generación de energía eléctrica
Importe	Importe pagado por central y tipo de combustible
Calor Total	Variable para el cálculo del régimen térmico

Identificación de los criterios de selección de dichas variables, es decir, de las dimensiones.

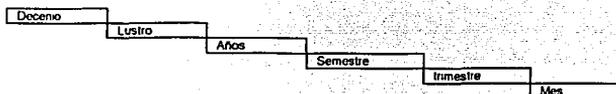
Dimensión	Descripción
Tiempo	Tiempo
Central	Central generadora
Combustible	Tipo de combustible

Identificación de los niveles jerárquicos en las dimensiones.

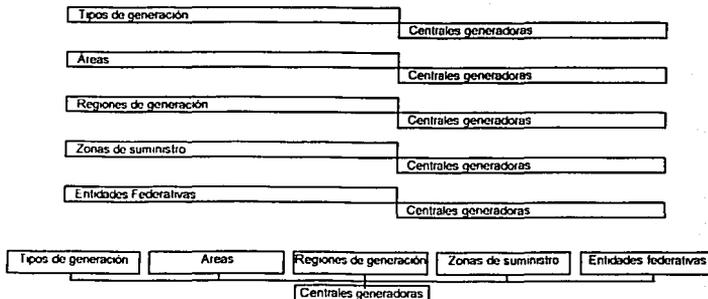
Dimensión: tipo de combustible  
Jerarquía: Sin Jerarquía

Valores		Unidades
Combustible	m <sup>3</sup>	
Gas	miles de m <sup>3</sup>	
Diesel	m <sup>3</sup>	
Carbón	Ton	

Dimensión: tiempo  
Jerarquía: lineal



Dimensión: Central  
Jerarquía: Múltiple.



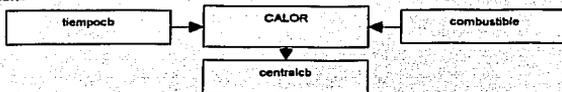
TESIS CON FALLA DE ORIGEN

Diseño del repositorio multidimensional según las vistas diseñadas (MOLAP).

Estrella 1. Consumo de combustibles.



Estrella 2. Calor.



Estrella 3. Generación de energía eléctrica.



Estrella 4. Calor total para el cálculo del régimen térmico.



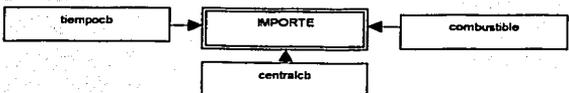
Estrella 5. Régimen térmico.



Estrella 6. Precio ponderado.

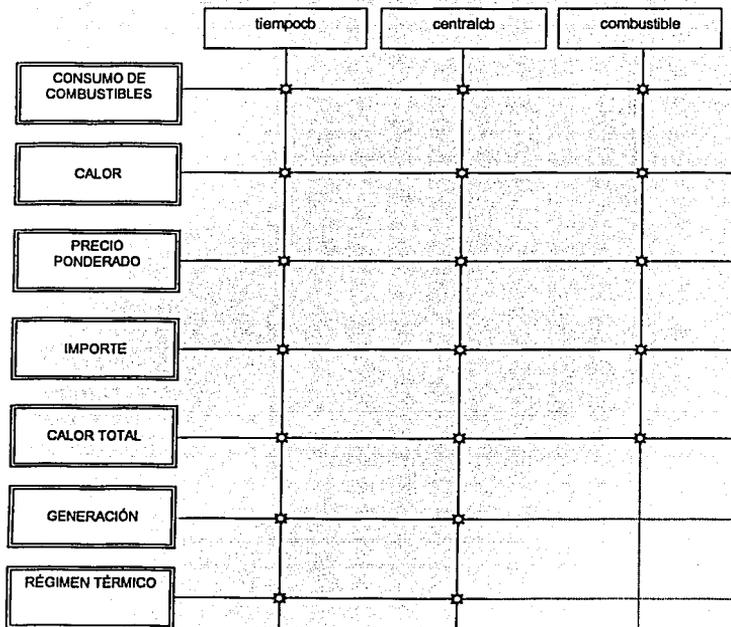


Estrella 7. Importe.



TESIS CON  
FALLA DE ORIGEN

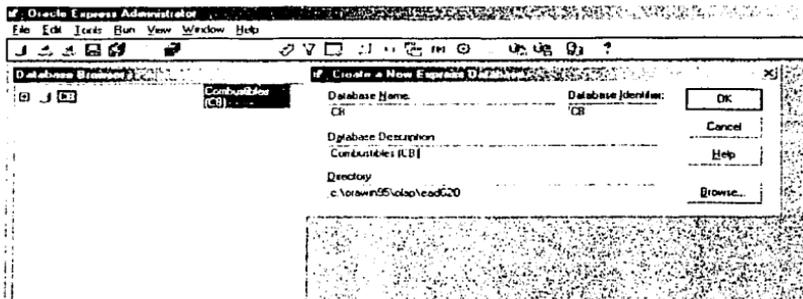
Constelación multidimensional del Data Mart de Energéticos.



TESTE CON  
FALLA DE ORIGEN

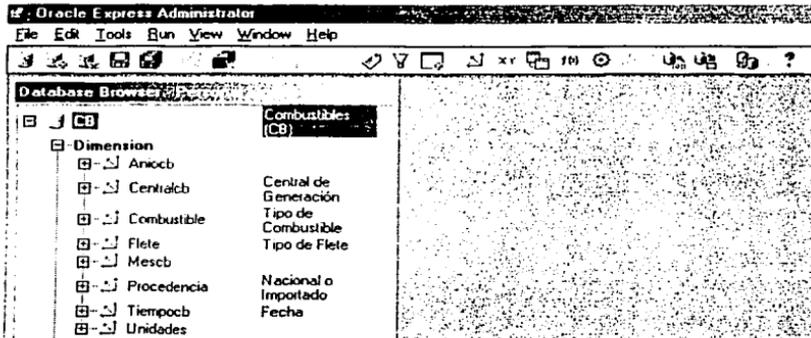
#### 4.5. Desarrollo de las estructuras multidimensionales.

Desarrollo del repositorio multidimensional en Oracle Express Server.



Implantación de estructuras multidimensionales en Oracle Express.

Creación de dimensiones.



TRABAJA CON  
PALLA DE ORIGEN

**Carga de datos en dimensiones.**

**Import Text Data** [X]

Data File: \_\_\_\_\_ Program Name: [Untitled]

File Type: \_\_\_\_\_ [Browse...]

Character delimited fields
  Fixed length fields
 Advanced...

Delimiters  
 Text Field Start: "" Text Field End: "" Close

General Field: [Comma (,)] Help

File Sample:

Recrd #: [1] Column

File Layout:

Field #: [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100] [101] [102] [103] [104] [105] [106] [107] [108] [109] [110] [111] [112] [113] [114] [115] [116] [117] [118] [119] [120] [121] [122] [123] [124] [125] [126] [127] [128] [129] [130] [131] [132] [133] [134] [135] [136] [137] [138] [139] [140] [141] [142] [143] [144] [145] [146] [147] [148] [149] [150] [151] [152] [153] [154] [155] [156] [157] [158] [159] [160] [161] [162] [163] [164] [165] [166] [167] [168] [169] [170] [171] [172] [173] [174] [175] [176] [177] [178] [179] [180] [181] [182] [183] [184] [185] [186] [187] [188] [189] [190] [191] [192] [193] [194] [195] [196] [197] [198] [199] [200] [201] [202] [203] [204] [205] [206] [207] [208] [209] [210] [211] [212] [213] [214] [215] [216] [217] [218] [219] [220] [221] [222] [223] [224] [225] [226] [227] [228] [229] [230] [231] [232] [233] [234] [235] [236] [237] [238] [239] [240] [241] [242] [243] [244] [245] [246] [247] [248] [249] [250] [251] [252] [253] [254] [255] [256] [257] [258] [259] [260] [261] [262] [263] [264] [265] [266] [267] [268] [269] [270] [271] [272] [273] [274] [275] [276] [277] [278] [279] [280] [281] [282] [283] [284] [285] [286] [287] [288] [289] [290] [291] [292] [293] [294] [295] [296] [297] [298] [299] [300] [301] [302] [303] [304] [305] [306] [307] [308] [309] [310] [311] [312] [313] [314] [315] [316] [317] [318] [319] [320] [321] [322] [323] [324] [325] [326] [327] [328] [329] [330] [331] [332] [333] [334] [335] [336] [337] [338] [339] [340] [341] [342] [343] [344] [345] [346] [347] [348] [349] [350] [351] [352] [353] [354] [355] [356] [357] [358] [359] [360] [361] [362] [363] [364] [365] [366] [367] [368] [369] [370] [371] [372] [373] [374] [375] [376] [377] [378] [379] [380] [381] [382] [383] [384] [385] [386] [387] [388] [389] [390] [391] [392] [393] [394] [395] [396] [397] [398] [399] [400] [401] [402] [403] [404] [405] [406] [407] [408] [409] [410] [411] [412] [413] [414] [415] [416] [417] [418] [419] [420] [421] [422] [423] [424] [425] [426] [427] [428] [429] [430] [431] [432] [433] [434] [435] [436] [437] [438] [439] [440] [441] [442] [443] [444] [445] [446] [447] [448] [449] [450] [451] [452] [453] [454] [455] [456] [457] [458] [459] [460] [461] [462] [463] [464] [465] [466] [467] [468] [469] [470] [471] [472] [473] [474] [475] [476] [477] [478] [479] [480] [481] [482] [483] [484] [485] [486] [487] [488] [489] [490] [491] [492] [493] [494] [495] [496] [497] [498] [499] [500] [501] [502] [503] [504] [505] [506] [507] [508] [509] [510] [511] [512] [513] [514] [515] [516] [517] [518] [519] [520] [521] [522] [523] [524] [525] [526] [527] [528] [529] [530] [531] [532] [533] [534] [535] [536] [537] [538] [539] [540] [541] [542] [543] [544] [545] [546] [547] [548] [549] [550] [551] [552] [553] [554] [555] [556] [557] [558] [559] [560] [561] [562] [563] [564] [565] [566] [567] [568] [569] [570] [571] [572] [573] [574] [575] [576] [577] [578] [579] [580] [581] [582] [583] [584] [585] [586] [587] [588] [589] [590] [591] [592] [593] [594] [595] [596] [597] [598] [599] [600] [601] [602] [603] [604] [605] [606] [607] [608] [609] [610] [611] [612] [613] [614] [615] [616] [617] [618] [619] [620] [621] [622] [623] [624] [625] [626] [627] [628] [629] [630] [631] [632] [633] [634] [635] [636] [637] [638] [639] [640] [641] [642] [643] [644] [645] [646] [647] [648] [649] [650] [651] [652] [653] [654] [655] [656] [657] [658] [659] [660] [661] [662] [663] [664] [665] [666] [667] [668] [669] [670] [671] [672] [673] [674] [675] [676] [677] [678] [679] [680] [681] [682] [683] [684] [685] [686] [687] [688] [689] [690] [691] [692] [693] [694] [695] [696] [697] [698] [699] [700] [701] [702] [703] [704] [705] [706] [707] [708] [709] [710] [711] [712] [713] [714] [715] [716] [717] [718] [719] [720] [721] [722] [723] [724] [725] [726] [727] [728] [729] [730] [731] [732] [733] [734] [735] [736] [737] [738] [739] [740] [741] [742] [743] [744] [745] [746] [747] [748] [749] [750] [751] [752] [753] [754] [755] [756] [757] [758] [759] [760] [761] [762] [763] [764] [765] [766] [767] [768] [769] [770] [771] [772] [773] [774] [775] [776] [777] [778] [779] [780] [781] [782] [783] [784] [785] [786] [787] [788] [789] [790] [791] [792] [793] [794] [795] [796] [797] [798] [799] [800] [801] [802] [803] [804] [805] [806] [807] [808] [809] [810] [811] [812] [813] [814] [815] [816] [817] [818] [819] [820] [821] [822] [823] [824] [825] [826] [827] [828] [829] [830] [831] [832] [833] [834] [835] [836] [837] [838] [839] [840] [841] [842] [843] [844] [845] [846] [847] [848] [849] [850] [851] [852] [853] [854] [855] [856] [857] [858] [859] [860] [861] [862] [863] [864] [865] [866] [867] [868] [869] [870] [871] [872] [873] [874] [875] [876] [877] [878] [879] [880] [881] [882] [883] [884] [885] [886] [887] [888] [889] [890] [891] [892] [893] [894] [895] [896] [897] [898] [899] [900] [901] [902] [903] [904] [905] [906] [907] [908] [909] [910] [911] [912] [913] [914] [915] [916] [917] [918] [919] [920] [921] [922] [923] [924] [925] [926] [927] [928] [929] [930] [931] [932] [933] [934] [935] [936] [937] [938] [939] [940] [941] [942] [943] [944] [945] [946] [947] [948] [949] [950] [951] [952] [953] [954] [955] [956] [957] [958] [959] [960] [961] [962] [963] [964] [965] [966] [967] [968] [969] [970] [971] [972] [973] [974] [975] [976] [977] [978] [979] [980] [981] [982] [983] [984] [985] [986] [987] [988] [989] [990] [991] [992] [993] [994] [995] [996] [997] [998] [999] [1000]

**Creación de variables.**

File Edit View Run Window Help

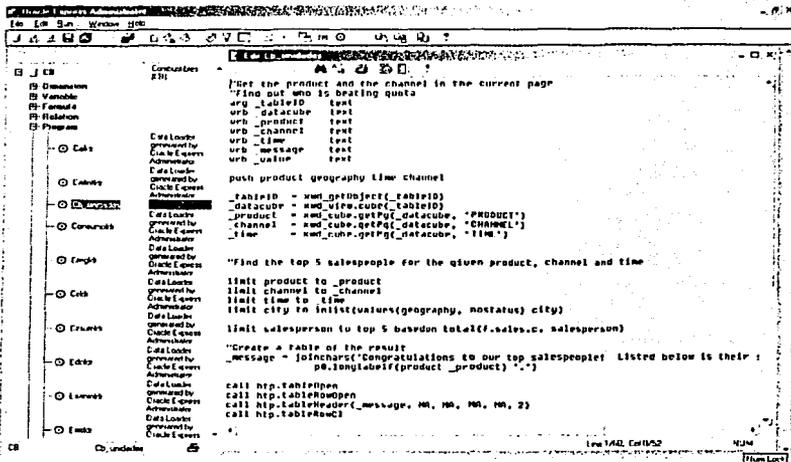
Dimensiones

- Dimensiones
- Variable
  - Calc P Ed + 10710
  - Calc Inc
  - Consumo CONSUMO
  - Es
  - Exit
  - Fact Factor de Planta
  - GENERACION GENERACION
  - Generador BWh
  - GRUPO GRUPO
  - Impulso
  - Instal INSTAL
  - Unidad
- Formulas
- Relaciones
- Programa
- Modelo
- Variables

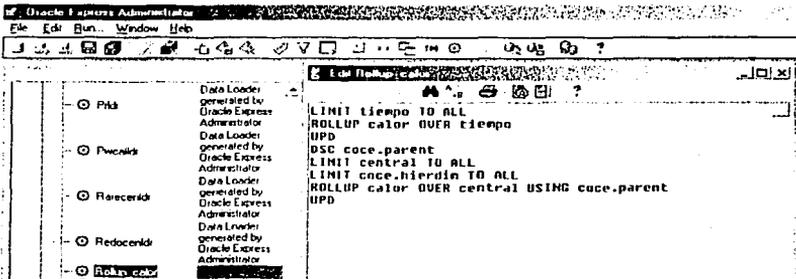
List of Express databases and database objects

TESIS CON  
 FALLA DE ORIGEN

Creación de programas de carga de variables.



Creación de programas de agregación.



TESIS CON  
FALLA DE ORIGEN

#### 4.6. Implantación de las estructuras relacionales orientadas a la transformación de datos del Data Mart de Energéticos.

A continuación se presentan los scripts para la creación de tablas espejo en el servidor destino sun6000, sistema operativo Solaris y con manejador de bases de datos Oracle 8.04.

Script de creación de la tabla de eb\_fletes.

create table eb\_fletes (

```
CL_CENTR CHAR(5) CONSTRAINT eb_fletes_leva1_pk PRIMARY KEY,
FE_VIGEN DATE eb_fletes_leva2_pk PRIMARY KEY,
CA_COMBU CHAR(1) eb_fletes_leva3_pk PRIMARY KEY,
CL_TITILE CHAR(1) eb_fletes_leva4_pk PRIMARY KEY,
NU_PRECKO DECIMAL(10,1))
```

Script de creación de la tabla de eb\_diario.

create table eb\_diario (

```
CL_CENTR CHAR(5) CONSTRAINT eb_diario_leva1_pk PRIMARY KEY,
FE_REGIS DATE eb_diario_leva2_pk PRIMARY KEY,
CA_POCA1 DECIMAL(10,1),
CA_POCA2 DECIMAL(10,1),
CA_POCA3 DECIMAL(10,1),
CA_POCA4 DECIMAL(10,1),
NU_COCOM NUMBER,
NU_COCOI NUMBER,
NU_COGAN NUMBER,
NU_COGAI NUMBER,
NU_CODIN NUMBER,
NU_CODOI NUMBER,
NU_COCAN NUMBER,
NU_COCOI NUMBER,
NU_EXCON NUMBER,
NU_EXCOI NUMBER,
NU_EXCAN NUMBER,
NU_EXCAI NUMBER,
NU_RECON NUMBER,
NU_RECOI NUMBER,
NU_RECON NUMBER,
NU_RECOI NUMBER,
NU_REDI NUMBER,
NU_REDAI NUMBER,
NU_RECAI NUMBER,
NU_RECAI NUMBER,
NU_TRACO NUMBER,
NU_TRADI NUMBER,
NU_TRACA NUMBER,
NU_DERCO NUMBER,
NU_DERDIA NUMBER,
NU_DERCA NUMBER,
NU_GENCE NUMBER)
```

Script de creación de la tabla de eb\_condia.

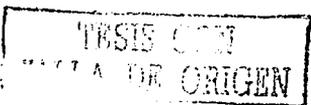
create table eb\_condia (

```
CL_CENTR CHAR(3) CONSTRAINT eb_condia_leva1_pk PRIMARY KEY,
FE_REGIS DATE,
NU_UNIDAD CHAR(1),
NU_COCOM INTEGER,
NU_COCOI INTEGER,
NU_COGAN INTEGER,
NU_COGAI INTEGER,
NU_CODIN INTEGER,
NU_CODOI INTEGER,
NU_COCAN INTEGER,
NU_COCOI INTEGER)
```

Script de creación de la tabla de eb\_mensua.

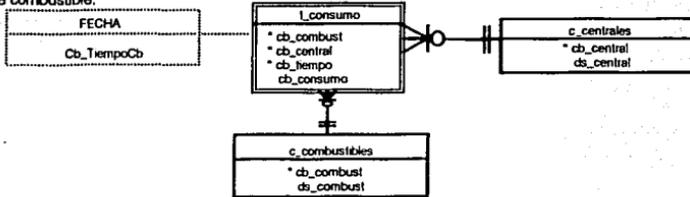
create table eb\_mensua (

```
CL_CENTR CHAR(3) CONSTRAINT eb_mensua_leva1_pk PRIMARY KEY,
FE_REGIS DATE,
FE_MOOF DATE,
NU_COCOM DECIMAL(10,1),
NU_COCOI DECIMAL(10,1),
NU_COCOM DECIMAL(10,1),
NU_COCAN DECIMAL(10,1),
NU_COCOI DECIMAL(10,1),
NU_COCAR DECIMAL(10,1),
NU_COCODI DECIMAL(10,1),
NU_COGAN DECIMAL(10,1),
NU_COGAI DECIMAL(10,1),
NU_COCAN DECIMAL(10,1),
NU_COCOI DECIMAL(10,1),
NU_EXCON DECIMAL(10,1),
NU_EXCOI DECIMAL(10,1),
NU_EXCAN DECIMAL(10,1),
NU_EXCAI DECIMAL(10,1),
NU_RECON DECIMAL(10,1),
NU_RECOI DECIMAL(10,1),
NU_RECON DECIMAL(10,1),
NU_RECOI DECIMAL(10,1),
NU_RECAN DECIMAL(10,1),
NU_RECAI DECIMAL(10,1),
NU_RECAR DECIMAL(10,1),
NU_RECOI DECIMAL(10,1),
NU_REDI DECIMAL(10,1),
NU_REDAI DECIMAL(10,1),
NU_GENER DECIMAL(10,1),
NU_CAGOM DECIMAL(10,1),
NU_CAGAS DECIMAL(10,1),
NU_CADEI DECIMAL(10,1),
NU_CAGAR DECIMAL(10,1),
NU_TOTCA DECIMAL(10,1),
NU_CONME DECIMAL(10,1),
NU_CONMA DECIMAL(10,1),
NU_CONMI DECIMAL(10,1),
NU_TOTCO DECIMAL(10,1),
NU_CARNA DECIMAL(10,1),
NU_CARMA DECIMAL(10,1),
NU_CENZA DECIMAL(10,1),
NU_TOTCA DECIMAL(10,1),
NU_DIENA DECIMAL(10,1),
NU_DIEM DECIMAL(10,1),
NU_TOTDI DECIMAL(10,1),
NU_GASMA DECIMAL(10,1),
NU_GASMI DECIMAL(10,1),
NU_TOTGA DECIMAL(10,1),
NU_TOCMB DECIMAL(10,1),
NU_TTCOM DECIMAL(10,1),
NU_TMCOM DECIMAL(10,1),
NU_TTDE DECIMAL(10,1),
NU_TMDE DECIMAL(10,1),
NU_TRANS DECIMAL(10,1),
NU_TOTAL DECIMAL(10,1),
NU_CPROC DECIMAL(10,1),
NU_CPROT DECIMAL(10,1),
NU_TCOSP DECIMAL(10,1),
CA_FAPLA DECIMAL(10,1))
```



#### 4.7. Diseño del repositorio relacional (ROLAP) en Oracle 8 en función de la estructura del MOLAP.

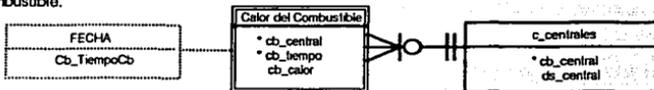
Consumo de combustible.



SISTEMA:	COMBUSTIBLES		
BASE DE DATOS:	Combustibles DB		
TABLA:	T_Consumo		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
Cb_Consumo	Consumo	decimal					
Cb_Combust	Tipo de Combustible	texto	20	si	no	Nombre del Combustible	C_Tipo de Combustible
Cb_TiempoCb	Fecha	texto	10	si	no	mm/aaaa	
Cb_CentralC	Central de generación	texto	40	si	no	Cve de la planta	c_generacion

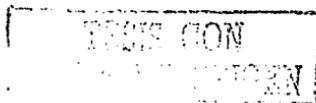
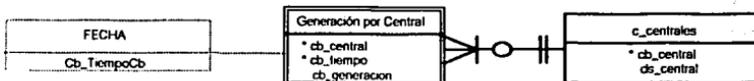
Calor de combustible.



SISTEMA:	COMBUSTIBLES		
BASE DE DATOS:	Combustibles DB		
TABLA:	T_CalorCb		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
Cb_Calor	Calor del Combustible	decimal					
Cb_TiempoCb	Fecha	texto	10	si	no	mm/aaaa	
Cb_CentralC	Central de generación	texto	40	si	no	Cve de la planta	c_generacion

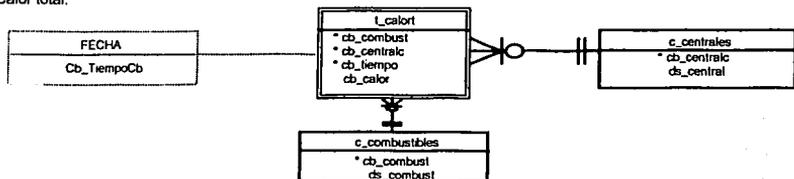
Generación por central.



SISTEMA:	COMBUSTIBLES		
BASE DE DATOS:	Combustibles DB		
TABLA:	T_Gen		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
Cb_Generar	Generacion por Central	decimal					
Cb_TiempoCB	Fecha	texto	10	si	no	mm/aaaa	
Cb_CentralC	Central de generacion	texto	40	si	no	Cve de la planta	c_generacion

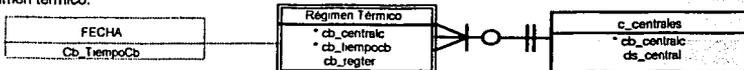
Calor total.



SISTEMA:	COMBUSTIBLES		
BASE DE DATOS:	Combustibles DB		
TABLA:	T_Cabtr		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
Cb_Cabtr	Cabtr Total (Kcal x 10 + 10)	decimal					
Cb_Combust	Tipo de Combustible	texto	20	si	no	Nombre del Combustible	C_Tipo de Combustible
Cb_TiempoCB	Fecha	texto	10	si	no	mm/aaaa	
Cb_CentralC	Central de generacion	texto	40	si	no	Cve de la planta	c_generacion

Régimen térmico.

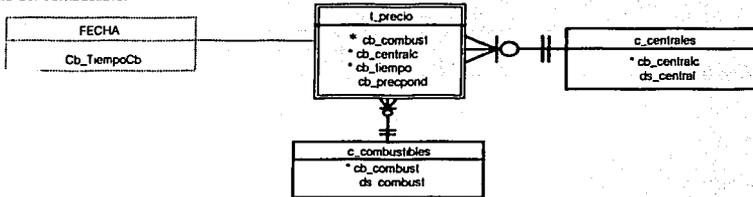


SISTEMA:	COMBUSTIBLES	COMBUSTIBLES	
BASE DE DATOS:	Combustibles DB		
TABLA:	T_Regimen		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
Cb_Registr	Regimen termico	decimal					
Cb_TiempoCB	Fecha	texto	10	si	no	mm/aaaa	
Cb_CentralC	Central de generacion	texto	40	si	no	Cve de la planta	c_generacion

TESIS CON  
FALLA DE ORIGEN

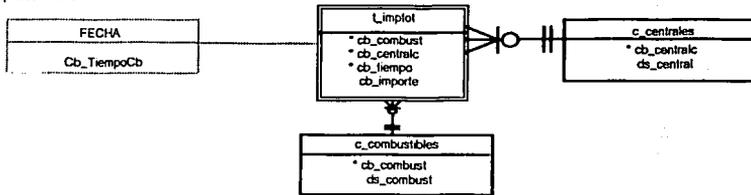
Precio del combustible.



SISTEMA:	COMBUSTIBLES	COMBUSTIBLES	
BASE DE DATOS:	Combustibles DB		
TABLA:	T_Precio		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
Cb_Precio	Precio Ponderado	decimal					
Cb_Combust	Tipo de Combustible	texto	20	si	no	Nombre del Combustible	C_Tipo de Combustible
Cb_TiempoCb	Fecha	texto	10	si	no	mm/aaaa	
Cb_CentralC	Central de generacion	texto	40	si	no	Ove de la planta	c_generacias

Importe total.



SISTEMA:	COMBUSTIBLES	COMBUSTIBLES	
BASE DE DATOS:	Combustibles DB		
TABLA:	T_Importe		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
Cb_Importe	Importe Total del Combustible	decimal					
Cb_Combust	Tipo de Combustible	Texto	20	si	no	Nombre del Combustible	C_Tipo de Combustible
Cb_TiempoCb	Fecha	Texto	10	si	no	mm/aaaa	
Cb_CentralC	Central de generacion	Texto	40	si	no	Ove de la planta	c_generacias

TESIS CON  
FALLA DE ORIGEN

#### 4.8. Diccionario de datos de índices calculados.

##### CONSUMO

DATOS DE SALIDA	OPERACIÓN	TABLA DE PROCEDENCIA
Total de consumo nacional=	gas nac+combustóleo nac+carbón nac+diesel nac	b_mensua
Total de consumo importado=	gas imp+combustóleo imp+carbón imp+diesel imp	b_mensua
Total de combustóleo=	total mensual combustóleo nacional+total consumo importado	b_mensua
Total de gas=	total mensual gas nacional+total mensual gas importado	b_mensua
Total de combustóleo=	total mensual combustóleo nacional+total mensual de combustóleo importado	b_mensua
Total de carbón=	total mensual carbón nacional+total mensual carbón importado	b_mensua
Total de diesel=	total mensual diesel nacional+total mensual diesel importado	b_mensua

BASE DE DATOS:	ener097		
TABLA:	sum_consumos		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
tot_consumnacional	Total nacional	decimal	15,5				
tot_consumoimp	Total importado	decimal	15,5				
tot_combustible	Total de combustibles	decimal	15,5				
tot_combustoleo	Total de combustóleo	decimal	15,5				
tot_carbon	Total de carbón	decimal	15,5				
tot_diesel	Total de diesel	decimal	15,5				
tot_gas	Total de gas	decimal	15,5				
año	Año	texto	4	si	no	aaaa	
mes	Mes	texto	2	si	no	mm	
cl_central	Cave de la Central	texto	5	si	no		

##### EXISTENCIAS

DATOS DE SALIDA	OPERACIÓN	TABLA DE PROCEDENCIA
Total existencia nacional=	exist gas nac+exist combustóleo nac+exist carbón nac+exist diesel nac	b_mensua
Total existencia importado=	exist gas imp+exist combustóleo imp+exist carbón imp+exist diesel imp	b_mensua
Total de existencias=	total existencia nacional+total de existencia importado	b_mensua
Total de existencias de gas=	total mensual existencias gas nacional+total mensual existencias gas importado	b_mensua
Total de existencias de combustóleo=	total mensual existencia combustóleo nacional+total mensual existencia combustóleo importado	b_mensua
Total de existencias de carbón=	total mensual existencia de carbón nacional+total mensual existencia carbón importado	b_mensua
Total de existencias de diesel=	total mensual existencia diesel nacional+total mensual existencia diesel importado	b_mensua

TESIS CON  
VALOR DE ORIGEN

BASE DE DATOS	ener0997		
TABLA	sum_existencias		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
tot_existnacional	Total nacional	decimal	15,5				
tot_existimp	Total importado	decimal	15,5				
tot_existcombustible	Total de combustibles	decimal	15,5				
tot_existcombustibleo	Total de combustibleo	decimal	15,5				
tot_existcarbon	Total de carbon	decimal	15,5				
tot_existdiesel	Total de diesel	decimal	15,5				
tot_existgas	Total de gas	decimal	15,5				
anio	Año	texto	4	si	no	aaaa	
mes	Mes	texto	2	si	no	mm	
c1central	Clave de la Central	texto	5	si	no		

### IMPORTES

DATOS DE SALIDA	OPERACION	TABLA DE PROCEDENCIA
Total importes nacional=	importes de gas nacional+importes de combustibleo nacional+importes de carbon nacional+importes de diesel nacional+importes de ceniza nacional	b_mesnua
Total de importes importado=	importes de gas importado+importes de combustibleo importado+importes de carbon importado+importes de diesel importado+importes de ceniza importado	b_mesnua
Total de importes=	total de importes nacional+total de importes importado	b_mesnua
Total de importes de gas=	total de importe mensual de gas nacional+total de importes mensual gas importado	b_mesnua
Total de importes de combustibleo=	total de importe mensual de combustibleo nacional+total de importe mensual combustibleo importado	b_mesnua
Total de carbon=	total de importe mensual de carbon nacional+total de importe mensual carbon importado	b_mesnua
Total de importes de diesel=	total de importe mensual de diesel nacional+total de importe mensual diesel importado	b_mesnua

BASE DE DATOS:	ener0997		
TABLA	sum_importes		

NOMBRE	DESCRIPCION	TIPO	LONGITUD	LLAVE	NULOS	VALIDACION	RELACION
tot_importnacional	Total nacional	decimal	15,5				
tot_importimp	Total importado	decimal	15,5				
tot_impcombustible	Total de combustibles	decimal	15,5				
tot_impcombustibleo	Total de combustibleo	decimal	15,5				
tot_impcarbon	Total de carbon	decimal	15,5				
tot_impdiesel	Total de diesel	decimal	15,5				
tot_impceniza	Total de ceniza	decimal	15,5				
tot_impgas	Total de gas	decimal	15,5				
anio	Año	texto	4	si	no	aaaa	
mes	Mes	texto	2	si	no	mm	
c1central	Clave de la Central	texto	5	si	no		

TESIS CON  
 FALLA DE ORIGEN

#### 4.9. Desarrollo del repositorio relacional (ROLAP).

PROGRAMA 1 create table t_calorcb ( CB_CENTRAC CHAR(40) CONSTRAINT t_calorcb_leva1_pk PRIMARY KEY, CB_TIEMPOCB CHAR(10) CONSTRAINT t_calorcb_leva2_pk PRIMARY KEY, CB_CALOR DECIMAL(10,1))	PROGRAMA 6 create table t_gen ( CB_CENTRAC CHAR(40) CONSTRAINT t_gen_leva1_pk PRIMARY KEY, CB_TIEMPOCB CHAR(10) CONSTRAINT t_gen_leva2_pk PRIMARY KEY, CB_GENERAR DECIMAL(10,1))
PROGRAMA 2 create table t_calort ( CB_CENTRAC CHAR(40) CONSTRAINT t_calort_leva1_pk PRIMARY KEY, CB_COMBUST CHAR(20) CONSTRAINT t_calort_leva2_pk PRIMARY KEY, CB_TIEMPOCB CHAR(10) CONSTRAINT t_calort_leva3_pk PRIMARY KEY, CB_CALORT DECIMAL(10,1))	PROGRAMA 7 create table t_regimen ( CB_CENTRAC CHAR(40) CONSTRAINT t_regimen_leva1_pk PRIMARY KEY, CB_TIEMPOCB CHAR(10) CONSTRAINT t_regimen_leva2_pk PRIMARY KEY, CB_REGTER DECIMAL(10,1))
PROGRAMA 4 create table t_preco ( CB_CENTRAC CHAR(40) CONSTRAINT t_preco_leva1_pk PRIMARY KEY, CB_COMBUST CHAR(20) CONSTRAINT t_preco_leva2_pk PRIMARY KEY, CB_TIEMPOCB CHAR(10) CONSTRAINT t_preco_leva3_pk PRIMARY KEY, CB_PRECPOND DECIMAL(10,1))	PROGRAMA 8 create table t_impot ( CB_CENTRAC CHAR(40) CONSTRAINT t_impot_leva1_pk PRIMARY KEY, CB_COMBUST CHAR(20) CONSTRAINT t_impot_leva2_pk PRIMARY KEY, CB_TIEMPOCB CHAR(10) CONSTRAINT t_impot_leva3_pk PRIMARY KEY, CB_IMPORTE DECIMAL(10,1))
PROGRAMA 5 create table c_centrales ( CB_CENTRAC CHAR(40) CONSTRAINT c_centrales_pk PRIMARY KEY, DS_CENTRAS CHAR(80) CONSTRAINT t_gen_leva2_pk PRIMARY KEY)	PROGRAMA 9 create table c_combust ( CB_COMBUST CHAR(40) CONSTRAINT c_centrales_pk PRIMARY KEY, DS_COMBUST CHAR(80) CONSTRAINT t_gen_leva2_pk PRIMARY KEY)

#### 4.10. Diseño y desarrollo de programas express para la conexión, carga y agregación de datos de Oracle 8.

Programa de conexión a la base de datos relacional de Oracle en el tablespace de cb:

Nombre	Descripción
pcb_conect	Programa de conexión a la base de datos siopc, table space cb
ldr centrales	Programa de carga de la dimensión centrales
ldr tipocombust	Programa de carga de la dimensión tipo de combustibles
ldr consumo	Programa de carga de la variable consumo de combustibles
ldr calor	Programa de carga de la variable calor
ldr generacion	Programa de carga de la variable generación
ldr calorlot	Programa de carga de la variable calorlot
ldr importe	Programa de carga de la variable importes
rollup consumo	Programa de roll up de la variable consumo de combustibles
rollup calor	Programa de roll up de la variable calor
rollup generacion	Programa de roll up de la variable generación
rollup calorlot	Programa de roll up de la variable calorlot
rollup importe	Programa de roll up de la variable importes
Ejecuta	Programa que ejecuta todos los programas

TESIS CON  
FALLA DE ORIGEN

Programa de conexión a la base de datos (pcb\_connect):

```
DEFINE INIT_Oracle PROGRAM
LD Attach to Oracle

sql dbms = 'oracle'
sql messages = yes
sql connect cb identified by 'cb@scorp'
if sqtcode eq 0
then show 'Estas conectado a oracle'
else show 'Conexión fallida'
```

Programa de carga de la dimensión cb\_centrales (ldr.centrales):

```
SQLODBMS = 'ORACLE'

sql declare cu_centrales cursor for -
select cb_central, ds_central from c_centrales

if sqtcode ne 0
then signal derror 'Error en la Conexión'

sql open cu_centrales

if sqtcode ne 0
then signal operator 'Error abriendo cursor'

while sqtcode eq 0
do
sql fetch cu_centrales into :append Centralcb : All.shorttable
doend

sql close cu_centrales
SQL ROLLBACK WORK
```

Programa de carga de la dimensión cb\_combustibles

```
(ldr.combustibles):

SQLDBMS = 'ORACLE'

sql declare cu_combustibles cursor for -
select cb_combustible, cb_combustible from c_combustible

if sqtcode ne 0
then signal derror 'Error en la Conexión'

sql open cu_combustibles

if sqtcode ne 0
then signal operator 'Error abriendo cursor'

while sqtcode eq 0
do
sql fetch cu_combustibles into :append Combustiblecb : All.shorttable
doend

sql close cu_combustibles
SQL ROLLBACK WORK
```

Programa de carga de la variable consumo (ldr.consumo):

```
vb_cb_central: text
vb_cb_combust: text
vb_cb_tiempocb: text
vb_cb_consumo: decimal

SQLDBMS = 'ORACLE'

limit Centralcb to all
limit Combustible to all
limit Tiempocb to all

show joinchars ('Celdas Ocupadas: ' count(consumo ne NA))

sql declare cur1co cursor for -
select cb_central, cb_combust, cb_tiempocb, cb_consumo from
consumo
```

```
if sqtcode ne 0
then signal derror 'Error en la Conexión'
```

```
sql open cur1co
```

```
if sqtcode ne 0
then signal operator 'Error abriendo cursor'
```

```
while sqtcode eq 0
do
sql fetch cur1co into :_cb_central, :_cb_combust, :_cb_tiempocb,
:_cb_consumo
limit cb_central to :_cb_central
limit cb_combust to :_cb_combust
limit cb_tiempocb to :_cb_tiempocb

cb_consumo = :cb_consumo
show joinchars ( :_cb_central: ' ' :_cb_combust: ' ' :_cb_tiempocb: ' '
:_cb_consumo)
doend
```

```
sql close cur1co
SQL ROLLBACK WORK * Para que elimine la definición del Cursor al final
de la sesión
```

```
limit Centralcb to all
limit Combustible to all
limit Tiempocb to all
```

```
show joinchars ('Celdas Ocupadas: ' count(consumo ne NA))
```

Programa de carga de la variable calor (ldr.calor):

```
vb_cb_central: text
vb_cb_tiempocb: text
vb_cb_calor: decimal

SQLDBMS = 'ORACLE'

limit Centralcb to all
limit Tiempocb to all

show joinchars ('Celdas Ocupadas: ' count(calor ne NA))
```

```
sql declare cur1ca cursor for -
select cb_central, cb_tiempocb, cb_calor from l_calor
```

```
if sqtcode ne 0
then signal derror 'Error en la Conexión'
```

```
sql open cur1ca
```

```
if sqtcode ne 0
then signal operator 'Error abriendo cursor'
```

```
while sqtcode eq 0
do
sql fetch cur1co into :_cb_central, :_cb_tiempocb, :_cb_calor
limit cb_central to :_cb_central
limit cb_tiempocb to :_cb_tiempocb

cb_calor = :_cb_calor
show joinchars ( :_cb_central: ' ' :_cb_tiempocb: ' ' :_cb_calor)
doend
```

```
sql close cur1ca
SQL ROLLBACK WORK * Para que elimine la definición del Cursor al final
de la sesión
```

```
limit Centralcb to all
limit Tiempocb to all
```

```
show joinchars ('Celdas Ocupadas: ' count(calor ne NA))
```

Programa de carga de la variable generación (ldr.generacion):

```
vb_cb_central: text
vb_cb_tiempocb: text
vb_cb_generar: decimal
```

TESIS CON  
FALLA DE ORIGEN

```
SOLDBMS = 'ORACLE'
```

```
limit Central: to all  
limit Tiempo: to all
```

```
show joinchans('Celdas Ocupadas: ' count(generar ne NA))
```

```
sql declare cur1gc cursor for -  
select cb_central, cb_tiempo, cb_generar from L_gen
```

```
if sqcode ne 0  
then signal dclerr 'Error en la Conexión'
```

```
sql open cur1gc
```

```
if sqcode ne 0  
then signal openr 'Error abriendo cursor'
```

```
while sqcode eq 0  
do  
sql fetch cur1gc into :_cb_central, :_cb_tiempo, :_cb_generar  
limit cb_central to :_cb_central  
limit cb_tiempo to :_cb_tiempo
```

```
cb_generar = :_cb_generar  
show joinchans ( :_cb_central ' ' :_cb_tiempo ' ' :_cb_generar )  
doend
```

```
sql close cur1gc  
SQL ROLLBACK WORK * Para que elimine la definición del Cursor al final  
de la sesión
```

```
limit Central: to all  
limit Tiempo: to all
```

```
show joinchans('Celdas Ocupadas: ' count(generar ne NA))
```

Programa de carga de la variable calor.total (ldr.calor.tot):

```
vb _cb_central: text  
vb _cb_combust: text  
vb _cb_tiempo: text  
vb _cb_cabrir: decimal
```

```
SOLDBMS = 'ORACLE'
```

```
limit Central: to all  
limit Combustible: to all  
limit Tiempo: to all
```

```
show joinchans('Celdas Ocupadas: ' count(cabritotal ne NA))
```

```
sql declare cur1cat cursor for -  
select cb_central, cb_combust, cb_tiempo, cb_cabrir from L_cabrir
```

```
if sqcode ne 0  
then signal dclerr 'Error en la Conexión'
```

```
sql open cur1cat
```

```
if sqcode ne 0  
then signal openr 'Error abriendo cursor'
```

```
while sqcode eq 0  
do  
sql fetch cur1cat into :_cb_central, :_cb_combust, :_cb_tiempo,  
:_cb_cabrir
```

```
limit cb_central to :_cb_central  
limit cb_combust to :_cb_combust  
limit cb_tiempo to :_cb_tiempo
```

```
cb_cabrir = :_cb_cabrir  
show joinchans ( :_cb_central ' ' :_cb_combust ' ' :_cb_tiempo ' ' :  
_cb_cabrir )  
doend
```

```
sql close cur1cat  
SQL ROLLBACK WORK * Para que elimine la definición del Cursor al final  
de la sesión
```

```
limit Central: to all  
limit Combustible: to all  
limit Tiempo: to all
```

```
show joinchans('Celdas Ocupadas: ' count(cabrir ne NA))
```

Programa de carga de la variable Importa (ldr.importe):

```
vb _cb_central: text  
vb _cb_combust: text  
vb _cb_tiempo: text  
vb _cb_impot: decimal
```

```
SOLDBMS = 'ORACLE'
```

```
limit Central: to all  
limit Combustible: to all  
limit Tiempo: to all
```

```
show joinchans('Celdas Ocupadas: ' count(impot ne NA))
```

```
sql declare cur1ic cursor for -  
select cb_central, cb_tiempo, cb_impot from L_importe
```

```
if sqcode ne 0  
then signal dclerr 'Error en la Conexión'
```

```
sql open cur1ic
```

```
if sqcode ne 0  
then signal openr 'Error abriendo cursor'
```

```
while sqcode eq 0  
do  
sql fetch cur1ic into :_cb_central, :_cb_combust, :_cb_tiempo,  
:_cb_impot
```

```
limit cb_central to :_cb_central  
limit cb_combust to :_cb_combust  
limit cb_tiempo to :_cb_tiempo
```

```
cb_impot = :_cb_impot  
show joinchans ( :_cb_central ' ' :_cb_combust ' ' :_cb_tiempo ' ' :  
_cb_impot )  
doend
```

```
sql close cur1ic  
SQL ROLLBACK WORK * Para que elimine la definición del Cursor al final  
de la sesión
```

```
limit Central: to all  
limit Combustible: to all  
limit Tiempo: to all
```

```
show joinchans('Celdas Ocupadas: ' count(impot ne NA))
```

Programa de sumariación de la variable consumo (rollup.consumo):

```
show joinchans('Inicio de Sumariación')
```

```
chgdir Consumo Inplace * Esta variable la sube a RAM
```

```
alltbl  
upd  
limit Central: complement Ccoco parent  
limit Combustible: complement Ccoco parent  
limit tiempo: complement Ccbi parent
```

```
show joinchans('Inicio de Sumariación por Central...')  
limit Central: to all  
rollup consumo over Central using Ccoco parent  
upd
```

```
show joinchans('Inicio de Sumariación por Tipo de combustible...')  
limit Combustible: to all  
rollup Consumo over Combustible using Ccoco parent  
upd
```

```
show joinchans('Inicio de Sumariación por Tiempo...')  
limit Tiempo: to all
```

rollup Consumo over Tiempooc using Cbl.parent  
upd

chgdfn consumo permanent

show joinchars(tod 'Fin de Sumarización')

Programa de sumarización de la variable calor (rollup.calor):

show joinchars(tod 'Inicio de Sumarización')

chgdfn Cblor inpace \* Esta variable le sube a RAM

akstat

upd

limit Centralcb complement Ccoo.parent

limit tiempooc complement Cbl.parent

show joinchars(tod 'Sumarización por Central...')

limit Centralcb to all

rollup Cblor over Central using Ccoo.parent

upd

show joinchars(tod 'Sumarización por Tiempo...')

limit Tiempooc to all

rollup Cblor over Tiempooc using Cbl.parent

upd

chgdfn Cblor permanent

show joinchars(tod 'Fin de Sumarización')

Programa de sumarización de la variable generación  
(rollup.generación):

show joinchars(tod 'Inicio de Sumarización')

chgdfn Generacion inpace \* Esta variable le sube a RAM

akstat

upd

limit Centralcb complement Ccoo.parent

limit tiempooc complement Cbl.parent

show joinchars(tod 'Sumarización por Central...')

limit Centralcb to all

rollup Generacion over Central using Ccoo.parent

upd

show joinchars(tod 'Sumarización por Tiempo...')

limit Tiempooc to all

rollup Generacion over Tiempooc using Cbl.parent

upd

chgdfn Generacion permanent

show joinchars(tod 'Fin de Sumarización')

Programa de sumarización de la variable calor.tot (rollup.calortot):

show joinchars(tod 'Inicio de Sumarización')

chgdfn Cblor.tot inpace \* Esta variable le sube a RAM

akstat

upd

limit Centralcb complement Ccoo.parent

limit Combustible complement Ccoo.parent

limit tiempooc complement Cbl.parent

show joinchars(tod 'Sumarización por Central...')

limit Centralcb to all

rollup Cblor.tot over Central using Ccoo.parent

upd

show joinchars(tod 'Sumarización por Tipo de combustible...')

limit Combustible to all  
rollup Cblor.tot over Combustible using Ccoo.parent  
upd

show joinchars(tod 'Sumarización por Tiempo...')

limit Tiempooc to all

rollup Cblor.tot over Tiempooc using Cbl.parent

upd

chgdfn Cblor.tot permanent

show joinchars(tod 'Fin de Sumarización')

Programa de sumarización de la variable importe (rollup.importe):

show joinchars(tod 'Inicio de Sumarización')

chgdfn Importe inpace \* Esta variable le sube a RAM

akstat

upd

limit Centralcb complement Ccoo.parent

limit Combustible complement Ccoo.parent

limit tiempooc complement Cbl.parent

show joinchars(tod 'Sumarización por Central...')

limit Centralcb to all

rollup Importe over Central using Ccoo.parent

upd

show joinchars(tod 'Sumarización por Tipo de combustible...')

limit Combustible to all

rollup Importe over Combustible using Ccoo.parent

upd

show joinchars(tod 'Sumarización por Tiempo...')

limit Tiempooc to all

rollup Importe over Tiempooc using Cbl.parent

upd

chgdfn Importe permanent

show joinchars(tod 'Fin de Sumarización')

Programa que ejecuta secuencialmente todos los programas de conexión,  
carga y rollup (Ejecute):

Call pcb\_connect

Call tr.centralcb

Call tr.tipocombus

Call tr.consumo

Call tr.calor

Call tr.generacion

Call tr.calortot

Call tr.importe

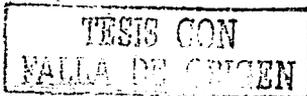
Call Rollup.consumo

Call Rollup.calor

Call Rollup.generacion

Call Rollup.calortot

Call Rollup.importe

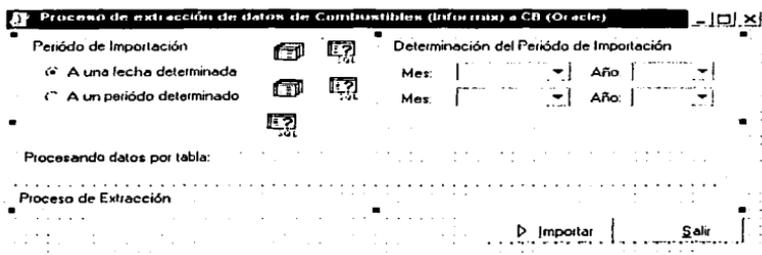


#### 4.11. Mecanismos para el acopio automático de información.

Se consideran:

- Cargas históricas de datos
- Cargas automáticas de datos mensuales
- Cargas específicas por mes y año.

Interface gráfica de la aplicación desarrollada en delphi 5.0 para la extracción, transporte, transformación e inserción de datos.



#### 4.12. Integración de las aplicaciones ETTI y desarrollo de un tablero de control que permita realizar la gestión y administración centralizada de estos procesos.

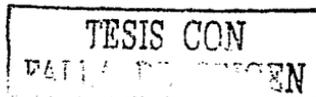
El control y la administración de las herramientas ETTI deberá registrar siguientes elementos:

Extracción:

- Nombre de la fuente
- Nombre del servidor fuente
- Tipo de manejador
- Ubicación física de la fuente de datos
- Responsable
- Nombre de la tabla
- IP del servidor fuente
- Número de registros de la tabla fuente
- Número de registros extraídos
- Número de registros rechazados

Transporte, Limpieza y filtrado:

- IP del servidor destino
- Nombre de la tabla
- Número de registros eliminados
- Número de cortes en la transmisión
- Número de registros antes de transporte
- Número de registros después de transporte
- Tiempo estimado de transmisión
- Tiempo real de transmisión



## Inserción y transformación

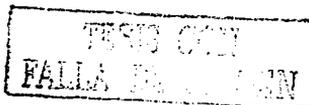
- Número de registros antes de la transformación
- Número de registros después de la transformación
- Número de registros antes de la inserción
- Número de registros después de la inserción

Adm. de Información		Principal		Data Mart		
<b>Tablero de Control de Mecanismos Automáticos de Acopio de Información</b>						
<b>Extracción</b>						
Nombre de la fuente	1531611165	Responsable	Gerencia de Generación	Num. de registros fuente	18.000.000	
Nombre del servidor	rs59-2	Nombre de la tabla	b_conda	Num. registros extraídos	18.000.000	
Horario	Informe 7 0 5 E	IP del servidor Fuente	153 16 11 136	Num. registros rechazados	0	
Utilización	GIT-CPD					
<b>Transporte Limpieza y Filtrado</b>						
IP del servidor destino	153 16 11 165	Num. registros antes de transporte	18.000.000	Transporte		
Nombre de la tabla	b_conda	Num. registros después de transporte	10.000.000	Limpieza		
Num. de registros eliminados	40	Tiempo estimado de transporte	35	(seg)	Filtrado	
Num. de cortes de transmisión	0	Tiempo real de transmisión	50	(seg)		
<b>Inserción y Transformación</b>						
Num. registros antes de transformación	18.000.000	Num. registros antes inserción	40	Inserción		
Num. registros después de transformación	10.000.000	Num. registros después inserción	40	Transformación		
Fecha de inicio de proceso	26/04/2001	Fecha de fin de proceso	28/04/2001	Hora de inicio de proceso	8:48:28	
				Hora de fin de proceso	9:19:09	
					Ejecutar todo	Salir

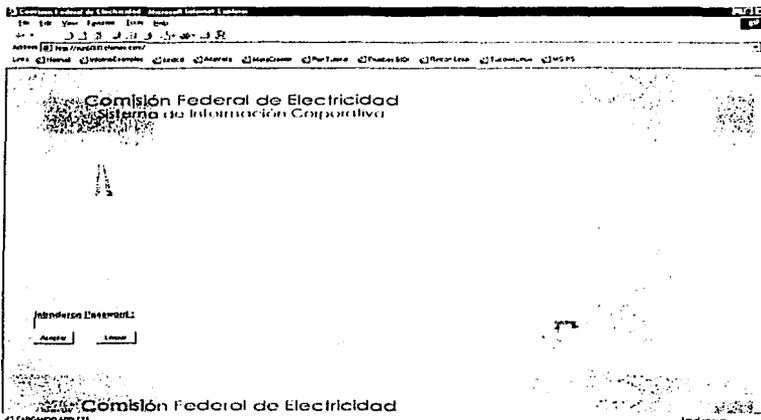
### 4.13. Diseño y desarrollo de la aplicación para la explotación de datos del Almacén de Datos de la CFE.

#### Especificaciones.

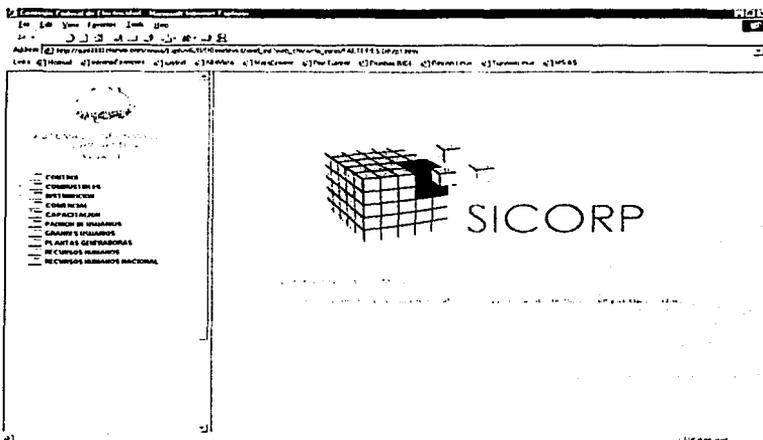
- El Sistema de Información Estratégica deberá contar con un mecanismo de acceso que permita:
  - Acceso a todas las áreas del sistema.
  - Acceso restringido a una o más áreas del sistema.
- El sistema deberá operar vía intranet, con el fin de tener la posibilidad de acceder a ella desde cualquier nodo de la red interna de la empresa.
- El sistema deberá presentar algunas pantallas informativas relativas a la empresa, procesos y áreas que se reportan.



## Desarrollo de la aplicación OLAP para la explotación de datos.



## Integración de la aplicación.



TESIS CON  
FALLA DE ORIGEN









## **CAPITULO V**

**Desarrollo de una aplicación para la minería de datos.**

TESIS CON  
FALLA DE ORIGEN

84-2

## 5.1. Planteamiento.

Los Almacenes de Datos son grandes bases de datos que contienen datos críticos relativos a la operación cotidiana de las organizaciones; sin embargo, un aspecto importante es la extracción de información significativa que revele patrones ocultos, tendencias y correlaciones que permita a sus directivos realizar predicciones que resuelvan problemas y que proporcionen ventajas competitivas.

Los Almacenes de Datos por sí solos no resuelven el problema de falta de información, tan solo garantiza la concentración de datos críticos. En muchas ocasiones se sufre por falta de información en medio de una inundación de datos.

Las herramientas de la minería de datos ayudan en la predicción de nuevas perspectivas, en el pronóstico de situaciones futuras de las organizaciones y en la realización de búsquedas inteligentes para localizar patrones significativos dentro de las grandes bases de datos.

En el caso del Almacén de Datos de la CFE, se pretende realizar un sistema de minería de datos que realice búsquedas inteligentes para encontrar patrones significativos relativos a la venta de energía eléctrica, los importes de cobranza, la generación y la demanda de energía eléctrica y, mediante el análisis de estos cuatro factores, determinar la existencia de suficientes condiciones para operar la CFE.

Evidentemente, el objetivo de este ejemplo es el de mostrar que a partir de los datos del Almacén es posible realizar descubrimientos relevantes haciendo uso de técnicas de minería de datos. Posiblemente para realizar conclusiones contundentes respecto a cualquier tema y área de la organización, es necesario considerar muchos más factores y hacer uso de una herramienta de minería de datos más sofisticada; sin embargo, este ejemplo solo pretende mostrar, de manera general, la potencialidad de la concentración de los datos y la obtención de información relevante mediante el uso de la minería de datos.

## 5.2. Estrategia de desarrollo.

Para realizar este ejemplo, se hace uso de la técnica de destilación de patrones, lo que supone la obtención de una muestra de datos cuyo fin es el de entrenar a la neurona; posteriormente, se preprocesan dichos datos a fin de obtener valores binarios que sirvan de entrada a la neurona y a continuación, se utiliza una neurona lineal adaptiva cuyo entrenamiento le permite generar las salidas esperadas sobre un conjunto finito de datos de entrada y, finalmente, se prueba la efectividad de su entrenamiento mediante la presentación de casos que no hayan sido mostrados previamente a la neurona. Véase la figura 5.1.

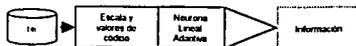
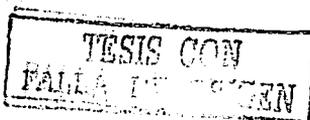


Figura 5.1. Minería de datos.

### Actividades.

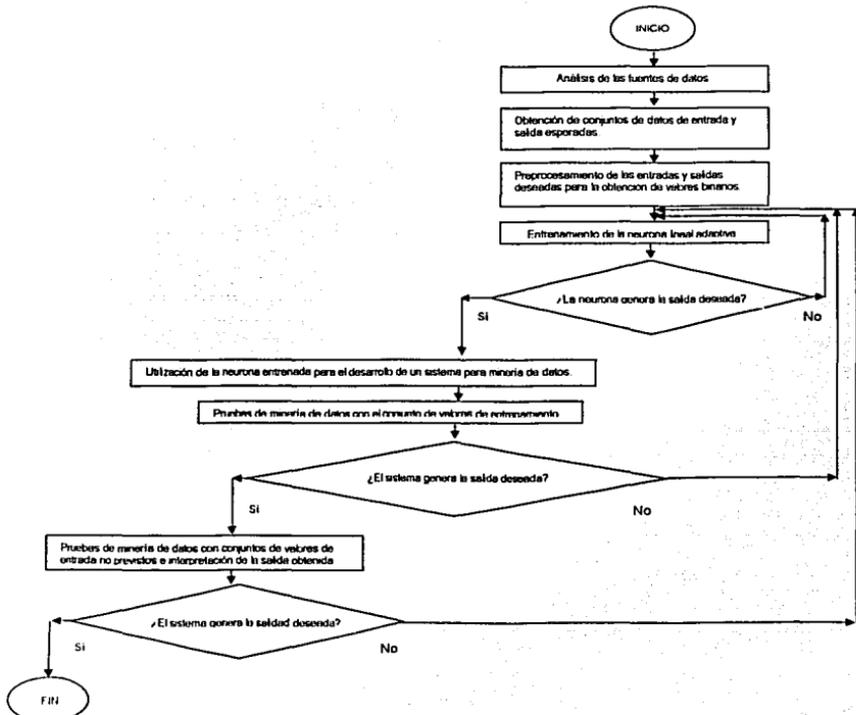
1. Se ubican dentro del Almacén las fuentes de datos que proporcionen los elementos relativos a los cuatro factores siguientes:

- Venta de energía eléctrica.
- Importes de cobranza.
- Generación.
- Demanda de energía eléctrica.



3. Mediante el uso del conjunto de valores de entrada y salida obtenidos en el punto 2, se realiza el entrenamiento de la neurona lineal adaptiva.
4. Se efectúan pruebas a la neurona entrenada.
5. Se desarrolla un programa que realice el procesamiento de los datos binarios de entrada con el uso de la neurona entrenada, a fin de que proporcione la salida correspondiente.

A continuación se muestra un diagrama que muestra la secuencia con la que se ejecutarán las acciones para la minería de datos. Véase figura 5.2.



En la figura 5.2 se muestran las etapas del proceso de minería de datos propuesto.

### 5.3. Descripción de elementos involucrados en la aplicación desarrollada.

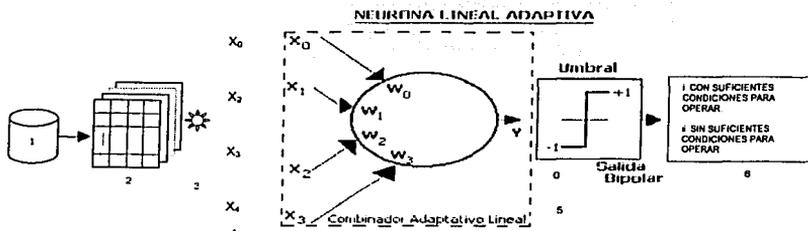


Figura 5.3. Elementos involucrados en la aplicación desarrollada.

1	Almacén de datos de la CPE.
2	Tabla: c_usuarios
	Tabla: c_agencias
	Tabla: c_tarifas
	Tabla: t_consumos
	Tabla: t_importes_ventas
	Tabla: t_importes_cobranza
	Tabla: c_plantash
	Tabla: c_plantast
3	Tabla: c_regiones
	Tabla: t_generación
	Tabla: t_region_agencia
	Tabla: t_demanda
3	Etapas de preprocesamiento para generar valores de entrada a la neurona lineal adaptativa.
4	Cuatro valores binarios de entrada a la neurona derivados de la etapa de preprocesamiento.
5	Cálculo de la salida binaria de la neurona lineal adaptativa.
6	Salida binaria.

Tabla 1. Descripción de Elementos de la Figura 5.2.

TESIS CON  
 FALLA DE ORIGEN

## 5.4. Análisis de la fuente de datos.

### Diccionario de datos.

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	c_usuarios

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
cl_agencia	Clave de la ubicación geográfica	varchar	5	si	no		c_agencias
cl_tarifa	Tipo de tarifa	varchar	1	si	no	D, A, C, I, S, O	c_tarifas
cl_usuario	Clave del usuario	varchar	13	si	no		
fe_alta	Fecha de alta del usuario	date		no	no	dd/mm/aaaa	
ds_usuario	Descripción del usuario	varchar	50	no	no		

### Muestra de datos.

cl_agencia	cl_tarifa	cl_usuario	fe_alta	ds_usuario
DA937	D	GAGI7R0426CD2	09/09/1970	Isabel Garrido
XD345	C	SOGE561208FR1	15/02/2000	Emmanuel Soto
TF967	I	MAZ1751011E1S8	24/12/1998	Iliana Magaña
GR368	S	SUIM60911RK9	09/10/1996	Maria Suárez

### II.

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	c_agencias

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
cl_agencia	Clave de la ubicación geográfica	varchar	5	si	no		
ds_agencia	Descripción de la ubicación geográfica	varchar	50	no	no		

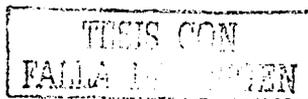
### Muestra de datos.

cl_agencia	ds_agencia
DA937	CORRAL
XD345	PROGRESO
TF967	SUR
GR368	OMAN

### IV.

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	c_tarifas

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
cl_tarifa	Tipo de tarifa	varchar	1	si	no	D, A, C, I, S, O	c_tarifas
ds_tarifa	Descripción de tarifa	varchar	20	no	no		
un_preciokwh	Precio unitario por KW hora de cada tarifa	real	17,2	no	no		



	KW hora de cada tarifa						
--	------------------------	--	--	--	--	--	--

**Muestra de datos**

cl_tarifa	ds_tarifa	nu_preciokwh
D	DOMÉSTICO	3 00
C	COMERCIAL	50 00
I	INDUSTRIAL	100 00
S	SERVICIOS	20 00
O	OTROS	10 00

**V.**

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	1 consumos

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
cl_agencia	Clave de la ubicación geográfica	varchar	5	si	no		c_agencias
cl_tarifa	Tipo de tarifa	varchar	1	si	no	D, A, C, I, S, O	c_tarifas
cl_usuario	Clave del usuario	varchar	5	si	no		c_usuarios
fe_corte	Fecha de corte	date		si	no	DD/MM/AAAA	
nu_consumo	KWH consumidos hasta la fecha de corte	number	17,2	no	no		

**Muestra de datos**

cl_agencia	cl_tarifa	cl_usuario	fe_corte	nu_consumo
DA937	D	GAGI780426CD2	15/04/1998	50
XD345	C	SOGE561208FR1	15/04/1998	800
TF587	I	MAZI751011EIS8	15/04/1998	1000
GR368	S	SUIM680911RK9	15/04/1998	700
Y1234	O	AMIK751011EH9	15/04/1998	500

**VI.**

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	1 importes ventas

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
cl_agencia	Clave de la ubicación geográfica	varchar	5	si	no		c_agencias
cl_tarifa	Tipo de tarifa	varchar	1	si	no	D, A, C, I, S, O	c_tarifas
cl_usuario	Clave del usuario	varchar	5	si	no		c_usuarios
fe_venta	Fecha de venta	date		si	no	DD/MM/AAAA	
un_ventas	Monio de cobranza	Number	17,2	no	no	nu_costokwh* nu_kwconsumo	c_tarifas, l_consumos

**Muestra de datos.**

cl_agencia	cl_tarifa	cl_usuario	fe_venta	nu_venta
DA937	D	GAGI780426CD2	30/04/1998	3*50=150 00
XD345	C	SOGE561208FR1	30/04/1998	50*800=4 000 00
TF587	I	MAZI751011EIS8	30/04/1998	100*1000=10 000 00
GR368	S	SUIM680911RK9	30/04/1998	20*700=14 000 00
Y1234	O	AMIK751011EH9	30/04/1998	10*500=5 000 00

IMPRESION CON  
FALLAS DE IMPRESION

## VII

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	I importes cobranza

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
c_l_agencia	Clave de la ubicación geográfica	Varchar	5	si	no		c_agencias
c_l_tarifa	Tipo de tarifa	Varchar	1	si	no	D A C 1 S O	c_tarifas
c_l_usuario	Clave del usuario	Varchar	5	si	no		c_usuarios
fe_cobranza	Fecha de cobro	date		si	no	DD/MM/AAAA	
nu_cobranza	Monto de cobranza	Number	17.2	no	no		

## Muestra de datos.

c_l_agencia	c_l_tarifa	c_l_usuario	fe_cobranza	nu_cobranza
DA937	O	CAGI780429CD2	30/04/1998	150.00
XD345	C	SCGE561268R1	30/04/1998	4.000.00
1F987	I	MAZ751011EIS8	30/04/1998	10.000.00
GR368	S	SUIM680911RK9	30/04/1998	14.000.00
Y1234	O	AMIK751011EH9	30/04/1998	5.000.00

## VIII

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	c_plantash

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
c_l_región	Clave de la región de generación	Varchar	5	si	no		c_regiones
c_l_planta	Clave de la planta generadora hidroeléctrica	Varchar	5	si	no		
nu_capacidad	Capacidad instalada MW	Number	17.2	no	no		
nu_autoabasto	Servicios propios autoabastecidos en MW	Number	17.2	no	no		

## Muestra de datos.

c_l_región	c_l_planta	nu_capacidad	nu_autoabasto
FE456	PH689	1230	120
HJ456	PH667	3456	345
TW345	PH799	1789	156

## IX

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	c_plantash

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
c_l_región	Clave de la región de generación	Varchar	5	si	no		c_regiones
c_l_planta	Clave de la planta generadora termoelectrica	Varchar	5	si	no		
nu_capacidad	Capacidad instalada MW	Number	17.2	no	no		
nu_autoabasto	Servicios propios autoabastecidos en MW	Number	17.2	no	no		

TEXIS CON  
FALLA DE ORIGEN

Muestra de datos.

cl_región	cl_planta	nu_capacidad	nu_autoabasto
DA234	PT123	1500	170
BC567	PT456	2000	240

X.

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	c_regiones

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
cl_región	Clave de la región de generación	Varchar	5	si	no		
ds_región	Descripción de la región	Varchar	40	no	no		

Muestra de datos.

cl_región	ds_región
DA234	Centro Oriente
BC567	Centro Occidente
DA234	Norte
BC567	Sur

XI.

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	t_generación

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
cl_región	Clave de la región de generación	Varchar	5	si	no		c_regiones
cl_planta	Clave de la planta generadora termoeléctrica	Varchar	5	si	no		c_plantash, c_plantast
fe_registro	Fecha de registro	Date		si	no	d3/mm/aaaa	
nu_generación	MW generados a la fecha de registro	number	17,2	no	no		

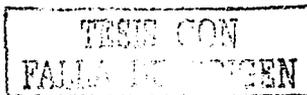
Muestra de datos.

cl_región	cl_planta	fe_registro	nu_generación
DA234	PT123	15/04/1998	1500
BC567	PT456	15/04/1998	2000
FE456	PH489	15/04/1998	1230
HJ456	PH567	15/04/1998	3456
TW345	PH799	15/04/1998	1789

XII

SISTEMA	SISTEMA DE INFORMACIÓN ESTRATÉGICA
BASE DE DATOS	ALMACÉN DE DATOS
TABLA	r_region_agencia

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
cl_región	Clave de la región de generación	Varchar	5	si	no		c_regiones
cl_agencia	Clave de la agencia	Varchar	5	si	no		c_agencias



Muestra de datos.

cl_region	cl_agencia
DA234	DA937
BC567	XD345
FE456	TF987
HJ456	CR368
TW345	YT234

XIII.

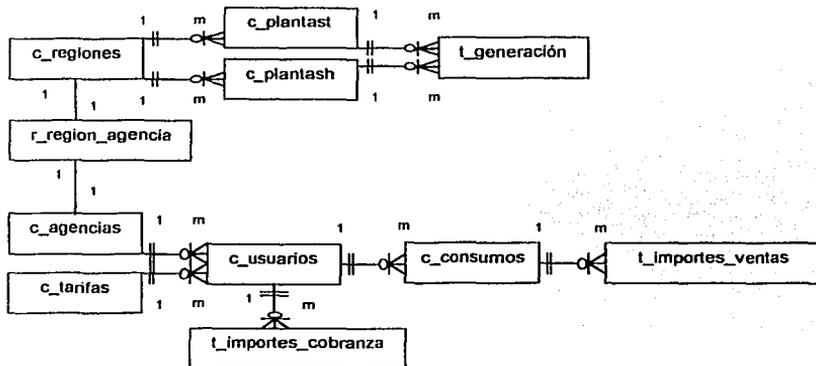
SISTEMA	SISTEMA DE INFORMACION ESTRATEGICA
BASE DE DATOS	ALMACEN DE DATOS
TABLA	t_demanda

Campo	Descripción	Tipo de dato	Longitud	Llave	Nulos	Validación	Relación
cl_agencia	Clave de la ubicación geográfica	varchar	5	si	no		c_agencias
cl_tarifa	Tipo de tarifa	varchar	1	si	no	D, A, C, I, S, O	c_tarifas
cl_usuario	Clave del usuario	varchar	5	si	no		c_usuarios
fe_corte	Fecha de corte	date		si	no	DD/MM/AAAA	
nu_demanda	MWII demandados hasta la fecha de corte	number	17,2	no	no		

Muestra de datos.

cl_agencia	cl_tarifa	cl_usuario	fe_corte	nu_demanda
DA937	D	GAGI/80426C1J2	15/04/1998	51
XD345	I	SOGE/561206R1	15/04/1998	830
TF987	I	MAZI/751011E158	15/04/1998	1233
CR368	S	SUM69091TRK9	15/04/1998	734
YT234	O	AMIK/751011EH9	15/04/1998	521

Diagrama Entidad-Relación.



TRIS CON FALLA DE ORIGEN

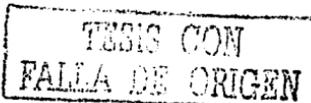
## 5.5. Etapa de preprocesamiento.

### Descripción de las operaciones con los datos.

	Variable	Descripción de operación	Operación	Tabla o índice calculado
1	Energía total generada MW	Sumatoria de la energía generada en todas las plantas de todas las regiones a una fecha determinada	$ETG = \sum (nu\_generación) \forall cl\_planta \& \forall cl\_región \& \forall fe\_registro$	L_generación
2	Total de servicios propios autoabastecidos	Sumatoria de todos los Servicios Propios Autoabastecidos de todas las plantas de todas las regiones a una fecha determinada	$SPA = \sum (nu\_autoabasto) \forall cl\_planta \& \forall cl\_región \& \forall fe\_registro$	c_plantash c_plantash
3	Energía total transmitida en MW	Energía total generada - Total de Servicios Propios Autoabastecidos	$ETT = ETG - SPA$	$ETT = (1-2)$ Índice tiempo
4	Energía total consumida en MW	Sumatoria de la energía consumida por todos los usuarios, de todas las tarifas, en todas las agencias a una fecha determinada	$ETC = \sum (nu\_consumo) \forall cl\_usuarios \& \forall cl\_tarifa \& \forall cl\_agencia \& \forall fe\_corte$	L_consumos
5	Diferencia de generación en MW	Energía total transmitida - Energía total consumida	$DG = ETT - ETC$	$DG = (3-4)$ Índice X2
6	Monto total de los importes de ventas \$	Sumatoria de las todas las ventas de todos los usuarios, de todas las tarifas, de todas las agencias a una fecha determinada	$MVI = \sum (nu\_ventas) \forall cl\_usuario \& \forall cl\_tarifa \& \forall cl\_agencia \& \forall fe\_corte$	L_importes_ventas
7	Importe total de cobranza en \$	Sumatoria de los montos de cobranza por agencia, tarifa, usuario a una fecha determinada	$ITC = \sum (nu\_cobranza) \forall cl\_usuario \& \forall cl\_tarifa \& \forall cl\_agencia \& \forall fe\_cobranza$	L_importes_cobranza
8	Adeudos \$	Monto total de los importes de ventas - Importe total de cobranza	$ADUUDOS = MVI - ITC$	$ADUUDOS = (6-7)$ Índice X3
8	Número de Usuarios	Sumatoria de usuarios de todas las tarifas, de todas las agencias a una fecha determinada	$USUARIO = \sum (cl\_usuario) \& \forall cl\_tarifa \& \forall cl\_agencia \& \forall fe\_corte$	c_usuarios
9	Total de nuevos usuarios	Cálculo de la diferencia entre el número de usuarios del periodo actual y el número de usuarios del periodo anterior	$INU = USUARIO_{fe\_nita} - USUARIO_{fe\_anterior}$	$DE = (8-8)$ fecha actual menos fecha anterior Índice X1
10	Demanda total MW	Sumatoria de la demanda por usuario, tarifa, agencia a una fecha determinada	$DEMANDA = \sum (nu\_demanda) \forall cl\_usuario \& \forall cl\_tarifa \& \forall cl\_agencia \& \forall fe\_corte$	L_demanda
11	Diferencia de demanda MW	Energía total transmitida - Demanda total	$DD = ETT - DEMANDA$	$DD = (3-10)$ Índice X4

### Estructuras de decisiones con salidas binarias.

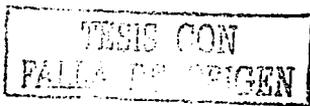
LOGICA PARA EL CÁLCULO DE PARÁMETROS*	PARÁMETROS SEGÚN LA TABLA ANTERIOR
if (Total de nuevos usuarios = 0) then x1 = 1 Else x1 = 0	if (INU = 0) then x1 = 1 Else x1 = 0
if (Diferencia de generación = 0) then x2 = 1 Else x2 = 0	if (DG = 0) then x2 = 1 Else x2 = 0
if (Diferencia de cobranza = 0) then x3 = 1 Else x3 = 0	if (ADEUDOS = 0) then x3 = 1 Else x3 = 0
if (Diferencia de demanda = 0) then x4 = 1 Else x4 = 0	if (DD = 0) then x4 = 1 Else x4 = 0



Aplicación de la estructuras de decisiones para la obtención de las salidas binarias.

Se plantean once casos para el entrenamiento de la neurona:

Disminuye el número de usuarios	X1+0
No se vende todo lo que se genera	X2+0
No se cobra todo lo que se vende	X3+0
No se generan todos los werts demandados	X4+0
No hay suficientes condiciones para operar	d+0
/	
Disminuye el número de usuarios	X1+0
No se vende todo lo que se genera	X2+0
Se cobra todo lo que se vende	X3+1
No se generan todos los werts demandados	X4+0
No hay suficientes condiciones para operar	d+0
/	
Disminuye el número de usuarios	X1+0
Se vende todo lo que se genera	X2+1
No se cobra todo lo que se vende	X3+0
Se generan todos los werts demandados	X4+1
Hay suficientes condiciones para operar	d+1
/	
Disminuye el número de usuarios	X1+0
Se vende todo lo que se genera	X2+1
Se cobra todo lo que se vende	X3+1
Se generan todos los werts demandados	X4+1
Hay suficientes condiciones para operar	d+1
/	
Aumenta o se mantiene el número de usuarios	X1+1
No se vende todo lo que se genera	X2+0
No se cobra todo lo que se vende	X3+0
No se generan todos los werts demandados	X4+0
No hay suficientes condiciones para operar	d+0
/	
Aumenta o se mantiene el número de usuarios	X1+1
No se vende todo lo que se genera	X2+0
No se cobra todo lo que se vende	X3+0
Se generan todos los werts demandados	X4+1
No hay suficientes condiciones para operar	d+1
/	
Aumenta o se mantiene el número de usuarios	X1+1
Se vende todo lo que se genera	X2+0
No se generan todos los werts demandados	X3+1
No se cobra todo lo que se vende	X4+0
No hay suficientes condiciones para operar	d+0
/	
Aumenta o se mantiene el número de usuarios	X1+1
Se vende todo lo que se genera	X2+1
No se cobra todo lo que se vende	X3+0
No se generan todos los werts demandados	X4+0
No hay suficientes condiciones para operar	d+0
/	
Aumenta o se mantiene el número de usuarios	X1+1
Se vende todo lo que se genera	X2+1
No se cobra todo lo que se vende	X3+0
Se generan todos los werts demandados	X4+1
Hay suficientes condiciones para operar	d+1
/	
Aumenta o se mantiene el número de usuarios	X1+1
Se vende todo lo que se genera	X2+1
Se cobra todo lo que se vende	X3+1
Se generan todos los werts demandados	X4+1
Hay suficientes condiciones para operar	d+1
/	
Aumenta o se mantiene el número de usuarios	X1+1
Se vende todo lo que se genera	X2+1
Se cobra todo lo que se vende	X3+1
Se generan todos los werts demandados	X4+1
Hay suficientes condiciones para operar	d+1



## 5.6. Entrenamiento de la neurona lineal adaptiva.

La regla de aprendizaje Least Mean Square (LMS) para el entrenamiento de la neurona.

Dado un vector de entrada  $x$ , resulta sencillo determinar un conjunto de pesos  $w$  que dé lugar a un valor de salida concreto  $y$ . Supongamos que se dispone de un conjunto de vectores de entrada  $\{x_1, x_2, \dots, x_L\}$ , cada uno de los cuales posee su propio valor correcto  $d_k$ ,  $k=1, L$ , que quizá sea único. El problema de hallar un único vector de pesos que pueda asociar con éxito cada vector de entrada con el valor de salida deseado ya no es sencillo. A continuación se desarrollará un método llamado regla de aprendizaje de mínimos cuadrados (Least Mean Square LMS), que es un método para hallar el vector de pesos deseado. Aludiremos a este proceso para hallar el vector de pesos deseado. Aludiremos este proceso para hallar el vector de pesos diciendo que estamos entrenando al ALC. La regla de aprendizaje se puede incorporar al propio dispositivo, que entonces se puede autoadaptar a medida que se le vayan presentando las entradas y salidas deseadas. Se hacen pequeños ajustes en los valores de los pesos cada vez que se procesa una combinación entrada-salida, hasta que el ALC da unas salidas correctas. En cierto sentido, este proceso es un verdadero proceso de entrenamiento, porque no necesitamos calcular explícitamente el valor de pesos. Antes de describir con detalle el proceso de entrenamiento, vamos a llevar a cabo el cálculo manualmente.

*Cálculo de  $w^*$ .*

Para empezar, vamos a formular el problema de un modo ligeramente distinto: dados los ejemplos  $(X_1, d_1)$ ,  $(X_2, d_2)$ , ...,  $(X_L, d_L)$ , de alguna función de procesamiento que asocia a los vectores de entrada  $X_k$  a (o los proyecta sobre) los valores de salida deseados,  $d_k$ , ¿cuál es el mejor de los vectores de pesos,  $w^*$ , para un ALC que lleve a cabo esta proyección?

Para responder a esta pregunta, primero hay que definir lo que constituye el mejor vector de pesos. Esta claro que, una vez que se haya encontrado el mejor, deseáramos que al aplicar todos los vectores de entradas obtuviésemos como resultado un valor de salida que fuese, con precisión, el valor correcto. Por tanto, es necesario eliminar, o por lo menos minimizar, la diferencia entre la salida deseada y la salida real para todos los vectores de entrada. La aproximación que se emplea aquí consiste en minimizar el error cuadrático medio para el conjunto de valores de entrada.

Si el valor de la salida es  $y_k$  para el  $k$ -ésimo vector de entrada, entonces el término de error correspondiente es:

$\epsilon_k = d_k - y_k$ . El error cuadrático medio, o valor esperado del error, se define en la forma:

$$\langle \epsilon^2 \rangle = \frac{1}{L} \sum_{k=1}^L \epsilon^2 k \quad \dots 2.2$$

en donde  $L$  es el número de vectores de entrada que haya en el conjunto de entrenamiento.

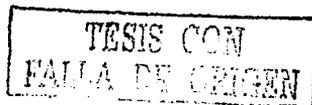
*Empleando la ecuación:*

$$y = w' x$$

Se puede desarrollar el error cuadrático medio como sigue:

$$\langle \epsilon^2 \rangle = \langle (d_k - w' x_k)^2 \rangle \quad \dots 2.3$$

$$\langle \epsilon^2 \rangle = \langle d^2 \rangle + w' \langle x_i x_i' \rangle w - 2 \langle d_k x_k' \rangle w \quad \dots 2.4$$



Al pasar de la Ecuación (2.3) a la (2.4), se ha hecho la suposición consistente en que el conjunto de entrenamiento permanece estacionario estadísticamente, lo cual quiere decir que los valores esperados sólo van a cambiar lentamente con el transcurso del tiempo. Esta suposición nos permite eliminar los vectores de pesos que aparecen en los términos de valores esperados dentro de la ecuación (2.4).

Defínase una matriz  $R = \langle x_k x_k^t \rangle$ , llamada matriz de correlación de entradas, y un vector  $p = \langle d_k x_k \rangle$ . Adicionalmente, hágase la identificación de  $\xi = \langle \xi_i^2 \rangle$ . Utilizando estas definiciones, se puede reescribir la ecuación 2.4 en la forma:

$$\xi = \langle d^2 \rangle + w^t R w - 2 p^t w$$

Esta ecuación muestra a  $\xi$  como función explícita del vector de pesos,  $w$ :  $\xi = \xi(w)$ .

Para hallar el vector de pesos correspondiente al error cuadrático medio mínimo se deriva de la ecuación (2.5), se evalúa el resultado en  $w^*$  y se hace el resultado igual a cero:

$$\begin{aligned} \frac{\partial \xi(w)}{\partial w} &= 2Rw - 2p \\ 2Rw^* - 2p &= 0 \\ R w^* &= p \\ w^* &= R^{-1} p \end{aligned}$$

Obsérvese que, aunque  $\xi$  es un escalar  $\frac{\partial \xi(w)}{\partial w}$  es un vector. La ecuación (2.6) es una expresión del gradiente de  $\xi$ ;  $\nabla \xi$ , que es un vector.

$$\nabla \xi = \left[ \frac{\partial \xi}{\partial w_1}, \frac{\partial \xi}{\partial w_2}, \dots, \frac{\partial \xi}{\partial w_n} \right]$$

Lo único que hemos hecho es demostrar que es posible hallar un punto en el cual la pendiente de la función  $\xi(w)$  es cero. En general, ese punto puede ser un máximo o un mínimo. En nuestro ejemplo, se tienen cuatro pesos. En esta situación, la gráfica de  $\xi(w)$  se conoce con el nombre de hiperparaboloides, la cual tiene que poseer una concavidad dirigida hacia arriba, puesto que todas las combinaciones de pesos deben dar lugar a un valor no negativo para el error cuadrático medio,  $\xi$ . Este resultado es general, y se obtiene independientemente de las dimensiones del vector de pesos.

*Especificación de los vectores de entrada para entrenamiento y salida esperada.*

CASEO	X1	X2	X3	X4	d
1	0	0	0	0	0
2	0	0	1	0	0
3	0	1	0	1	1
4	0	1	1	1	1
5	1	0	0	0	0
6	1	0	0	1	0
7	1	0	1	0	0
8	1	1	0	0	0
9	1	1	0	1	1
10	1	1	1	0	0
11	1	1	1	1	1

TESIS CON  
FALLA DE ORIGEN

Por lo que el vector de entradas es:

$$X = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

El vector de salidas deseadas es:

$$d := [0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 1]$$

Cálculo de las matrices: R y P.

TX:=transpose(X):

$$X^T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Td:=transpose(d):

$$d^T = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

P:=multiply(d,TX):

$$P := [2 \ 4 \ 2 \ 4]$$

TP:=transpose(P):

$$P^T = \begin{bmatrix} 2 \\ 4 \\ 2 \\ 4 \end{bmatrix}$$

R:=multiply(X,TX):

$$R = \begin{bmatrix} 7 & 4 & 3 & 3 \\ 4 & 6 & 3 & 4 \\ 3 & 3 & 5 & 2 \\ 3 & 4 & 2 & 5 \end{bmatrix}$$

Cálculo del vector de pesos.

$W := \text{multiply}(\text{INVR}, \text{TP})$ :

$$W = \begin{bmatrix} -3 \\ 14 \\ 37 \\ 98 \\ 3 \\ 49 \\ 59 \\ 98 \end{bmatrix}$$

Establecimiento de la neurona entrenada.

$$X^T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad W = \begin{bmatrix} -3 \\ 14 \\ 37 \\ 98 \\ 3 \\ 49 \\ 59 \\ 98 \end{bmatrix} \quad Y = X^T * W \quad \text{if } \begin{cases} y < 0.76, & \text{salida} = 0 \\ y \geq 0.76, & \text{salida} = 1 \end{cases}$$

### 5.7. Pruebas a la neurona entrenada.

Para probar que la neurona entrenada funciona correctamente, se debe de probar que la matriz de valores de entrada por el vector de pesos sea igual a la salida esperada:

$$X^T * W = d^T$$

$$X^T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} * W = \begin{bmatrix} -3 \\ 14 \\ 37 \\ 98 \\ 3 \\ 49 \\ 59 \\ 98 \end{bmatrix} = d^T = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

TESIS CON  
VALIA DE ORIGEN

Realizando el producto de la matriz de entradas por los pesos:

$$X^T W = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 14 \\ 37 \\ 98 \\ 3 \\ 49 \\ 59 \\ 98 \end{bmatrix} = \begin{bmatrix} 0 \\ 3/49 \\ 48/49 \\ 51/49 \\ -3/14 \\ 19/49 \\ -15/98 \\ 8/49 \\ 75/98 \\ 11/49 \\ 81/98 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.0612289796 \\ 0.9795918367 \\ 1.0408163265 \\ -0.2142857143 \\ 0.3877551024 \\ -0.1530612245 \\ 0.1632653061 \\ 0.7653061224 \\ 0.2244897959 \\ 0.8265061224 \end{bmatrix}$$

Aplicando el valor de umbral, se determina el valor de salida:

Y= 0 ⇒ Y<0.76 ∴ SAL=0
Y= 0.06 ⇒ Y<0.76 ∴ SAL=0
Y= 0.98 ⇒ Y≥0.76 ∴ SAL=1
Y= 1.04 ⇒ Y≥0.76 ∴ SAL=1
Y=-0.21 ⇒ Y<0.76 ∴ SAL=0
Y= 0.39 ⇒ Y<0.76 ∴ SAL=0
Y=-0.15 ⇒ Y<0.76 ∴ SAL=0
Y= 0.16 ⇒ Y<0.76 ∴ SAL=0
Y= 0.76 ⇒ Y≥0.76 ∴ SAL=1
Y= 0.22 ⇒ Y<0.76 ∴ SAL=0
Y= 0.83 ⇒ Y≥0.76 ∴ SAL=1

$$\Rightarrow \text{SALIDA} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Como: SALIDA = d<sup>1</sup> ∴ Se prueba el correcto entrenamiento de la neurona

## 5.8. Presentación de una aplicación para la minería de datos.

Introducción al sistema de minería de datos.

Delphi 5 de Borland es el editor y compilador empleado para desarrollar la aplicación para la minería de datos que se reporta en este trabajo de tesis.

La aplicación tiene como objetivo determinar si hay suficientes condiciones para operar a partir del análisis de cuatro factores: venta de energía eléctrica, los importes de cobranza, la generación y la demanda de energía eléctrica.

Con esta aplicación no se pretende realizar conclusiones contundentes, ya que para ello será necesario considerar más factores y hacer uso de una herramienta de minería más compleja.

En la figura 5.4. se muestra la interfase gráfica de la aplicación desarrollada

TESIS CON  
FALLA DE ORIGEN

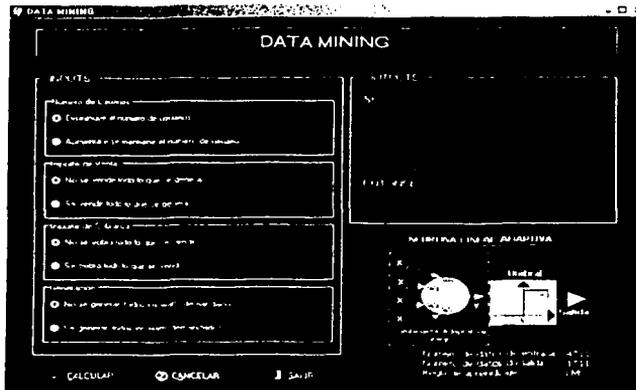


Figura 5.4. Pantalla principal de la herramienta de minería de datos

La aplicación realiza minería de datos a partir de cuatro factores de entrada mediante el uso de una neurona lineal adaptativa previamente entrenada. Véase la figura 5.5.



Figura 5.5 Gráfica de la neurona lineal adaptativa.

#### Estructura.

#### Valores de entrada.

En el caso real de una base de datos de producción o de un Almacén de Datos en constante actualización, los cuatro parámetros deberán ser recopilados e inmediatamente preprocesados periódicamente de manera automática a fin de obtener los valores binarios para el procesamiento de la neurona lineal adaptativa. En este ejemplo se presupone la etapa de preprocesamiento y la entrada de los datos será manual a través de selección de radio botones. Véase figura 5.6.

TRABAJO CON  
FALLAS DE ORIGEN

La zona de entradas "INPUTS" de la aplicación, se encuentra dividida en cuatro bloques, correspondientes a cuatro puntos estratégicos de la operatividad de la organización:

- Y Número de Usuarios
- Y Y Importe de Ventas
- Y Y Importe de Cobranza
- Y Y Generación

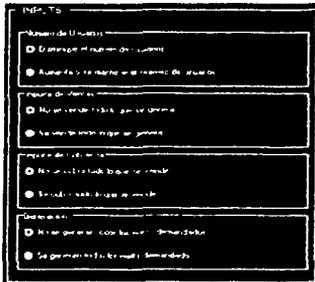


Figura 5.6. Zona de valores de entradas. INPUTS.

En caso de que se intente ejecutar la aplicación sin seleccionar previamente ningún valor, entonces los valores predeterminados por bloque son los siguientes:

Número de Usuarios	Disminuye el número de usuarios
Importe de Ventas	No se vende todo lo que se genera
Importe de la Cobranza	No se cobra todo lo que se vende
Generación	No se generan todos los watts demandados

Una vez seleccionados los valores de cada uno de los bloques, se deberá presionar el botón de calcular  para pasar los valores de entrada a la neurona lineal adaptiva previamente entrenada y de esa manera proceder a su procesamiento.

#### Procesamiento de los valores de entrada con la neurona entrenada y determinación de la salida.

En la etapa de procesamiento, se tomarán los cuatro valores binarios seleccionados previamente:

$$X=[X_1, X_2, X_3, X_4]$$

y se procederá a efectuar el producto cartesiano de  $x^T W$ . Recuérdese que el vector de pesos fue calculado en la etapa de entrenamiento de la neurona y que en el caso de la aplicación es un vector de valores constantes.

Una vez efectuado el producto cartesiano, se procederá a evaluar el escalar obtenido para verificar si se cumple que:

$$if \begin{cases} x^T W < 0.76 & salida = 0 \\ x^T W \geq 0.76 & salida = 1 \end{cases}$$

Una vez determinado el valor de salida, su interpretación deberá ser desplegada en la zona de salidas.

### Valores de salida.

La zona de salidas "OUTPUTS" se divide en dos áreas:

- En la parte superior se despliegan los valores seleccionados relativos a cada uno de los cuatro puntos estratégicos referentes a la operatividad de la organización.
- En la parte inferior se despliega el resultado del cálculo de dichos valores a través de la neurona lineal adaptiva, indicando la situación en la que se encuentra la organización.

Las salidas pueden tomar dos posibles valores:

- Y Hay suficientes condiciones para operar
- Y No hay suficientes condiciones para operar

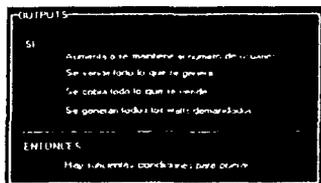


Figura 5.7. Zona de valores de salidas. OUTPUTS

Se espera que a partir del dato de salida, es decir, que existan o no condiciones suficientes para operar la organización considerando los factores de entrada mencionados, el directivo podrá tomar las acciones y decisiones convenientes a partir de esta información. Se reitera que es necesario considerar que esta operación deberá ser periódica y la herramienta generará una especie de alarma que le indique al directivo la presencia de una situación crítica.

Finalmente, respecto al manejo de herramienta podemos mencionar:

- Para ingresar nuevos valores de entrada a la neurona, se deberá presionar el botón de cancelar **CANCELAR**.
- Para salir de la aplicación se deberá presionar el botón de **SALIR**.

TESIS CON  
PALA DE ORIGEN

## **CONCLUSIONES**

TESIS CON  
FALLA DE ORIGEN

## CONCLUSIONES

Se comprueba la efectividad de la arquitectura propuesta mediante su aplicación en el diseño del Sistema de Información Estratégica de la CFE.

Se comprueba la efectividad de la metodología y herramientas de análisis y diseño propuestas en este trabajo de tesis, mediante su exitosa aplicación en el análisis, diseño y desarrollo del Data Mart de Energéticos de la CFE.

La carga histórica de datos en el Data Mart del área de Energéticos se efectuó exitosamente, lo que permite a los directivos hacer consultas en distintos periodos de tiempo, mediante las cuales es posible apreciar el comportamiento de esta área.

El Sistema de Información Estratégica reveló elementos significativos para diversas áreas de la CFE; por ejemplo, al conocerse el consumo de energéticos por planta en distintos periodos de tiempo, el departamento de Contabilidad puede anticipar con mayor certeza el presupuesto para esta área; por otro lado, el Centro Nacional de Control de Energía (CENACE) puede hacer un control más efectivo de las plantas generadoras, en vista de que es posible comparar la cantidad de energéticos consumidos contra la generación total; el departamento de Auditoría, puede supervisar que los pagos a proveedores como PEMEX, corresponda realmente con la cantidad de energéticos surtidos por planta generadora en distintos periodos de tiempo, etc.,

Se comprueba la efectividad del Data Mart de Energéticos como una fuente que aporta elementos clave para la toma de decisiones efectiva y el control de gestión de la CFE.

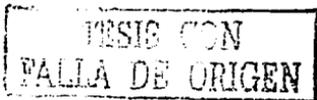
El uso de la tecnología OLAP permitió a los directivos y analistas de la CFE contestar preguntas estratégicas, e incluso se visualizaron nuevas perspectivas de la información y la consideración de opciones distintas; también brindó a los directivos la información base para realizar análisis comparativos, análisis de tendencias, modelado y proyecciones, lo que les permite anticiparse a ciertos eventos.

El desarrollo del Data Mart del área de Energéticos respondió a las expectativas de los directivos de la CFE, aunque se indicaron ciertos elementos de mejora que deben ser abordados en versiones posteriores, tales como el cálculo de índices y relaciones con otros data marts de la empresa, y la inclusión de nuevas variables.

Paralelamente al desarrollo del Data Mart del área de Energéticos se realizó el Data Mart relativo al área Comercial, lo que permitió probar la flexibilidad de la arquitectura propuesta en este trabajo de tesis.

La aplicación desarrollada para la minería de datos resultó exitosa en vista de que pudo inferir las respuestas correctas ante situaciones no presentadas previamente.

Tal y como se especificó en la introducción de este trabajo de tesis, el desarrollo de Sistemas de Información Estratégica no tiene fin, por lo que sin duda, en versiones posteriores, será posible mejorar lo desarrollado en este trabajo de tesis.



## REFERENCIAS

TESIS CON  
FALLA DE ORIGEN

## REFERENCIAS

### **Sistemas de información para la dirección**

El recurso estratégico crítico  
James C. Emery  
Ediciones Diaz Santos, S.A.

### **Oracle. Manual del administrador**

Kevin Loney  
McGraw Hill, 1996

### **Informix 4GL reference manual**

Database tool  
Volume 1 y 2  
Version. 4.00

### **Oracle. The complete reference**

Electronic edition  
George Koch and Kevin Loney  
McGraw-Hill

### **Introduction to Oracle: SQL and PL/SQL**

Student Guide  
Volume 1 & 2

### **Building the Operational Data Store**

Inmon, W.H.  
John Wiley & Sons, 1996

### **Fundamentos y modelos de bases de datos**

Adoración de Miguel Castaño, Mario G. Piattini Velhuis  
CompuTec / RAMA, 1998

### **The Data Warehouse toolkit.**

Ralph Kimball, Foreword by W.H. Inmon  
John Wiley and Sons, Inc. 1996

### **"Planning for a Data Warehouse"**

White Paper by Haistein, M.,  
<http://www.netlynx.com/cfbar/y9nla4.html>

### **"Architecting the Data Warehouse",**

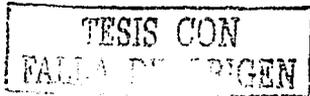
An Information Builders White Paper,  
Information Builders, Inc., 1995

### **"I want a Data Warehouse So, What is It Again?"**

White paper by Simon, A.,  
Supplement to Database Programming & Design. Diciembre, 1995.

### **"Building the Data Mart",**

White paper by Demarest, M.,  
DBMS, Julio, 1994



**Oracle 8 Data Warehousing**  
Michael J. Corey, Michael Abbey, Ian Abramson, Ben Taub  
Editorial McGraw-Hill

**Oracle express foundation**  
Student Guide  
Volume 1 & 2

**Oracle Express Objects for Application Developers**  
Student Guide  
Volume 1 & 2

**Develop Applications with Oracle Express Objects**  
Student Guide  
Volume 1 & 2

**Oracle Discoverer for Administrators**  
Student Guide

**Creating OLAP Briefings and Applications**  
Student Guide  
Volume 1 & 2

**Redes neuronales artificiales.**  
José R. Hijera, Víctor J. Martínez  
Alfaomega/RAMA, 2000

**Sian Ka'an**  
Notes from the International Conference  
The first joint Mexico-US  
International Workshop on Neural Networks and Neurocontrol.

TESIS CON  
FALLA DE ORIGEN