

3



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE ESTUDIOS SUPERIORES
CUAUTITLAN

“BASES DE DATOS,
MODELO MULTIDIMENSIONAL PARA UNA EMPRESA
DE CONTRATACION DE ESPACIOS PUBLICITARIOS”

TRABAJO DE SEMINARIO
QUE PARA OBTENER EL TITULO DE:
LICENCIADO EN INFORMATICA
P R E S E N T A:
FRANCISCO FERMIN FLORES FLORES

ASESOR: MCC. ARACELI NIVON ZAGHI

257296



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

UNIVERSIDAD NACIONAL AUTONOMA DE MÉXICO

**FACULTAD DE ESTUDIOS SUPERIORES
CUAUTITLAN**

**“BASES DE DATOS
MODELO MULTIDIMENSIONAL PARA UNA EMPRESA
DE CONTRATACIÓN DE ESPACIOS PUBLICITARIOS”**

TRABAJO DE SEMINARIO

QUE PARA OBTNER EL TITULO DE:

LICENCIADO EN INFORMATICA

PRESENTA:

FRANCISCO FERMIN FLORES FLORES

ASESOR: MCC. ARACELI NIVON ZAGHI

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLAN
UNIDAD DE LA ADMINISTRACION ESCOLAR
DEPARTAMENTO DE EXAMENES PROFESIONALES



REPUBLICA NACIONAL
DE MEXICO

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLAN
UNIDAD DE LA ADMINISTRACION ESCOLAR
DEPARTAMENTO DE EXAMENES PROFESIONALES

DR. JUAN ANTONIO MONTARAZ CRESPO
DIRECTOR DE LA FES CUAUTITLAN
P R E S E N T E

ATN: Q. Ma. del Carmen García Mijares
Jefe del Departamento de Exámenes
Profesionales de la FES Cuautitlán

Con base en el art. 51 del Reglamento de Exámenes Profesionales de la FES-Cuautitlán, nos permitimos comunicar a usted que revisamos el Trabajo de Seminario: Bases de Datos

Modelo Multidimensional Para Una Empresa de Contratación de Espacios Publicitarios

que presenta el pasante: Francisco Fermín Flores Flores

con número de cuenta: 9006622-3 para obtener el título de
Licenciado en Informática

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el EXÁMEN PROFESIONAL correspondiente, otorgamos nuestro VISTO BUENO.

ATENTAMENTE

"POR MI RAZA HABLARA EL ESPIRITU"

Cuautitlán Izcalli, Méx. a 30 de Junio de 2000

MODULO

PROFESOR

II

Mcc. Araceli Nivón Zaghi

I

Ing. Víctor Hugo Amollo Hernández

IV

Lic. Carlos Pineda Muñoz

FIRMA

LE
AGRADEZCO
MUCHÍSIMO
A DIOS POR
PERMITIRME
VIVIR Y POR
CUIDARME

A MI MAMA POR
INCULCARMÉ
BUENOS PRINCIPIOS
Y SOBRE TODO POR
SER MI MAMA

A MI PAPA Y A
SAMUEL Y NERI
POR TODA SU
AYUDA Y
APOYO

A MI SUEGRA POR SU
CARIÑO, EL GRAN APOYO
QUE NOS HA BRINDADO
Y SOBRE TODO POR
HABER TRAIIDO A ESTE
MUNDO A ZULE... UNA
PERSONITA SIN IGUAL

HAYDE, ULISES,
JAIME, VICKY, LULU,
GERARDO, JACKY,
ELISEO, ANA KA Y
DARIO POR SU APOYO
Y POR SER MI
FAMILIA

Y A TI... CARIÑO
POR SER EL
AMOR, FUERZA Y
CARÁCTER QUE
NECESITO PARA
ESTE LARGO
CAMINO

A LA UNIVERSIDAD POR
PERMITIRME FORMAR
PARTE DE SUS ALUMNOS
Y POR LA OPORTUNIDAD
QUE CONTRIBUYE A
NUESTRO DESAROLLO
PROFESIONAL

A LA PROF.
ZAGUI POR SUS
CONOCIMIENTOS
Y SU PACIENCIA

AL PROF. CARLOS
POR SU
PACIENCIA Y
APOYO

A LA LIC ROSY
VALADEZ POR SU
APOYO

Y A TODOS MIS
MAESTROS DE ESTA
FACULTAD POR SU
APOYO Y
CONOCIMIENTOS

INDICE

INTRODUCCION	I
OBJETIVO GENERAL	IV
OBJETIVOS ESPECÍFICOS	IV
CAPITULO I. MODELOS DE BASES DE DATOS	1
I.1. CONCEPTO DE BASE DE DATOS	1
I.2. BASES DE DATOS DE RED Y JERÁRQUICAS	5
I.3. BASES DE DATOS RELACIONALES	7
I.4. BASES DE DATOS ORIENTADAS A OBJETOS	23
I.5. BASES DE DATOS MULTIDIMENSIONALES	24
CAPITULO II. BD MULTIDIMENSIONALES(BDM)	28
II.1. CONCEPTOS Y CARACTERÍSTICAS	28
II.2. JERARQUÍAS EN LAS DIMENSIONES	34
II.3. TIPOS DE ESQUEMAS	36
II.4. REGLAS DE E.F. CODD.	43
II.5. COMPARACIÓN DE UNA BDM CON UNA BD RELACIONAL	48
II.6. ACCESO A LAS BDM A TRAVÉS DE SQL MULTIDIMENSIONAL	49
CAPITULO III.METODOLOGÍA PARA EL DISEÑO DE SISTEMAS EN BASES DE DATOS MULTIDIMENSIONALES	62
III.1. ELEMENTOS DE ANÁLISIS EN LAS BD MULTIDIMENSIONALES	62
III.2. PASOS PARA EL DISEÑO DE UNA BDM	65
III.3. SOFTWARE DE APLICACIÓN SOBRE BDM.	77
CAPITULO IV. CASO PRÁCTICO	79
IV.1. DESCRIPCIÓN DEL NEGOCIO	79
IV.2. REQUERIMIENTOS DEL SISTEMA	80
IV.3. ESQUEMA DE TIPO SNOWFLAKE	86
CONCLUSIONES	94
INDICE DE ILUSTRACIONES	96
BIBLIOGRAFÍA	98

INTRODUCCION

La información en las organizaciones es de gran importancia ya que en ello se basa la toma de decisiones. La estructura de la información que presenta las operaciones de una empresa conforma lo que se conoce como los Sistemas de Información.

Un sistema de información debe ser oportuno y confiable, además de proveer la información necesaria que permita tener un margen mínimo de error al tomar las decisiones, puesto que una mala decisión puede tener un alto costo para la empresa, ya sea de tipo monetario o bien costar la propia vida de la organización.

Las Bases de Datos (BD) juegan un papel primordial en el desempeño de los sistemas de Información de la empresa junto con los sistemas de cómputo. Las BD fueron diseñadas para facilitar el manejo de grandes cantidades de información, las cuales a través del tiempo han ido evolucionando, según se han satisfecho las necesidades de información que se han ido presentando en las organizaciones, tales como el incremento en el volumen de información y el cambio en las formas de la información que se utiliza para la toma de decisiones.

Con el proceso de evolución, se han desarrollado diferentes tipos de Bases de Datos. En un principio se manejaron las Bases de Datos Jerárquicas y de Red, más tarde surgió el concepto de las Bases de Datos Relacionales y en la actualidad, se están introduciendo nuevos conceptos con los que se pretende construir sistemas de Bases de Datos Inteligentes, tal es el caso de las Bases de Datos Orientadas a Objetos y las

Bases de Datos sobre sistemas OLAP (Procesamiento Analítico en Línea); es importante mencionar que cada tipo de BD incluyen varias características de sus antecesores.

Los sistemas administrativos de Bases de Datos que actualmente operan son bastante eficientes para el almacenamiento y recuperación de la información, de tal manera que un gerente puede generar los reportes que necesita para hacer sus interpretaciones y finalmente tomar una decisión; esto significa que la semántica o significado de un reporte lo obtiene aquella persona que lo interpreta.

Las tecnologías actuales de BD pretenden que las requisiciones de información a la BD se hagan de forma parecida al lenguaje humano y que los resultados que se obtengan de estas requisiciones sean de una forma inteligente, es decir que la exploración de la información sea rápida, que la información sea suficiente y además, que la información contenga cierta semántica para que sea fácil la toma de decisiones. Este tipo de soluciones se están obteniendo con los sistemas OLAP y el diseño de datos multidimensionales.

Las BD multidimensionales se siguen mejorando con el fin de servir con eficacia en el procesamiento de información para la toma de decisiones y aún no están muy difundidos aquí en México.

Con esta tesis se pretende dar a conocer la tecnología que se utiliza para optimizar la explotación de la información, además de dar a conocer en forma concreta, algunas de las herramientas que utilizan Datos Multidimensionales.

Para ello se aporta información sobre los principales conceptos de las Bases de Datos diseñadas en forma multidimensional utilizados por los sistemas OLAP, así como la metodología empleada para el desarrollo de BD multidimensionales, aplicado a un caso práctico.

Objetivo General

Describir las características de las Bases de Datos Multidimensionales y sus ventajas sobre las Bases de Datos Relacionales en el proceso de la información para la toma de decisiones.

Objetivos Específicos

- Estudiar la evolución de las Bases de Datos en forma concreta.
- Indicar como interactúan las viejas tecnologías con las Bases de Datos Multidimensionales.
- Desarrollar el modelo de tipo snowflake (copo de nieve) de una aplicación, utilizando Bases de Datos Multidimensionales.

CAPITULO I. Modelos de Bases de Datos

1.1. Concepto de Base de Datos

Una base de datos es un almacén de grandes cantidades de información interrelacionada que se encuentra estructurada en uno o varios archivos en disco; esta es manejada por un Sistema Administrador de Bases de Datos (también son conocidos como SMDB) para que a través de este se pueda almacenar y recuperar información eficientemente.

La información que es almacenada en la BD debe reunir las siguientes características:

- **Información Única:** La información redundante, en vez de servirle a la empresa para el asesoramiento en la toma de decisiones, le produce un problema de elección de la información que debe ser aceptada.
- **Información Específica:** Debe de contener un nivel de detalle suficiente y solo la información necesaria.
- **Información Consistente:** Es decir debe ser verídica sin ambigüedades.
- **Información Rápida:** La información sirve para la toma de una decisión y debe tenerse antes de la decisión.

El SMDB forma parte del *Sistema de Información* de la empresa. Un Sistema de Información es un grupo de procedimientos ordenados que al ser ejecutados, proporcionan información para apoyar la toma de decisiones y el control de las

operaciones en la organización. La tecnología ha proporcionado los medios necesarios, tales como transmisión de datos y computadoras, para el procesamiento de la información con el fin de obtenerla con las características antes mencionadas. La Ilustración. I-1 muestra gráficamente un sistema de información.

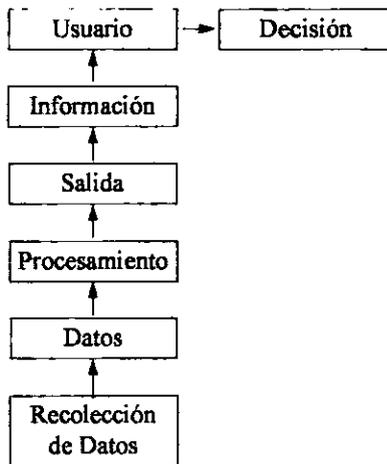


Ilustración I-1 Representación Esquemática de un Sistema de Información

Un sistema de información debe contener los siguientes atributos para que sea significativo para la empresa:

1. Debe proveer la suficiente información para la toma de decisiones en la organización.
2. Debe medir el entorno que rodea a cada actividad para poder pronosticar y controlar las circunstancias externas que lo rodean; detectando y proponiendo posibles soluciones.

3. Debe permitir conocer a tiempo, el desarrollo de cada una de las áreas de la organización para introducir medidas correctivas.

Un SMDB tiene cuatro componentes principales: Programas de Aplicación, la BD, un servidor de archivos y usuarios finales

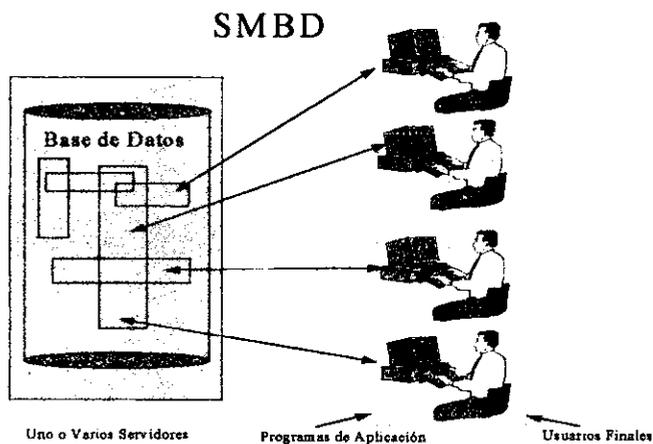


Ilustración I-2 SMDB

La administración de una BD implica la definición adecuada de las estructuras para el almacenamiento de la información, seguridad ante caídas del sistema o intentos de acceso de usuarios no permitidos; debe mantener la consistencia de la información y además permitir el uso de la Base de Datos a uno o varios usuarios que accesan al mismo tiempo la información, y además de controlar de los programas que accesan a esa información. (Ilustración. I-2)

La arquitectura de las BD se basan en un modelo de tres niveles ó capas: *Lógico, Conceptual, Físico.*

- El nivel Físico es el que se ocupa de como se almacenan físicamente los datos en los discos.
- El nivel Lógico es que se ocupa de la forma en que los usuarios finales entienden los datos.
- El nivel Conceptual es el nivel medio entre el físico y lógico. Es un panorama de los datos tal y como son, es decir que no describe los datos como lo perciben los usuarios y tampoco describe las limitaciones del equipo o lenguaje de programación.

Las Bases de Datos han sido un punto de investigación muy importante, buscando siempre una administración más eficiente de los datos, debido a que la cantidad de información es cada vez mayor en las empresas y por lo tanto su manejo es más complejo, pero además en la actualidad, las exigencias en la calidad de la información se van haciendo mayores, por que ya no solo se requiere obtener la información procesada y confiable en un tiempo corto sin importar la cantidad de información a procesar.

La tendencia en la actualidad, es la creación de modelos de BD que permitan obtener información procesada que contenga una semántica, en donde se muestre la información según como piensa la empresa, y que proponga soluciones, que pronostique ciertos problemas y además, que permita tener la información a la mano sin tener que hacer procesos complicados que tarden demasiado tiempo. Es por eso que a través de

la historia del procesamiento de datos, se han propuesto diferentes formas de organizar la información con el fin de satisfacer las diferentes necesidades de las empresas.

A continuación se explican brevemente las diferentes aportaciones de la tecnología sobre Bases Datos comprendida por:

1. Bases de Datos de Red
2. Bases de Datos Jerárquicas
3. Bases de Datos Relacionales
4. Bases de Datos Orientadas a Objetos
5. Bases de Datos Multidimensionales

Cabe mencionar que las BD jerárquicas y relacionales se explican más a detalle a través de el desarrollo del trabajo, ya que se son la base para el establecimiento de una BD multidimensional.

1.2. Bases de Datos de Red y Jerárquicas

Bases de Datos de Red

En este tipo de Base de Datos, la información está organizada en colecciones de registros relacionados entre sí a través de enlaces. Un enlace es una asociación de dos registros exclusivamente y su organización se describe por medio de grafos arbitrarios.

Un esquema para representar este de tipo de datos, es a través de el diagrama de estructuras de datos, el cual esta compuesto por cajas que representan a tipos de registros y por líneas que indican el enlace entre los registros. Cada registro contiene un campo puntero por cada enlace con el que esté asociado; solo se podrán tener relaciones de uno a uno y de uno a muchos; ya que las relaciones de muchos a muchos indica que cada registro debe tener una cantidad ilimitada de atributos puntero, por lo tanto esto no es viable.

Por ejemplo las personas que conforman un departamento en una organización (Ilustración. I-3)

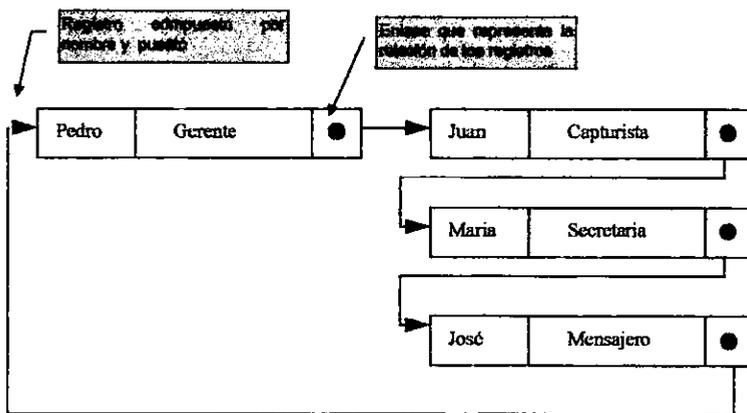


Ilustración I-3 Base de Datos de Tipo Red

Bases de Datos Jerárquicas

Al igual que el esquema de red, las BDJ utilizan registros relacionados entre sí a través de enlaces, la diferencia consiste en que en vez de utilizar grafos arbitrarios para

representar la información, se organiza en forma de registros que representan los nodos de un árbol, se clasifican en registros padre y registros hijos, un registro padre puede tener varios hijos, y un registro hijo solo puede tener un padre. Se tiene la misma estructura y representación que las BD de red, la única diferencia que existe es que en las BD jerárquicas se representan en estructuras de árboles en vez de grafos arbitrarios.

Un ejemplo clásico, es la información de las personas en cada departamento indicando su puesto (Ilustración. I-4)

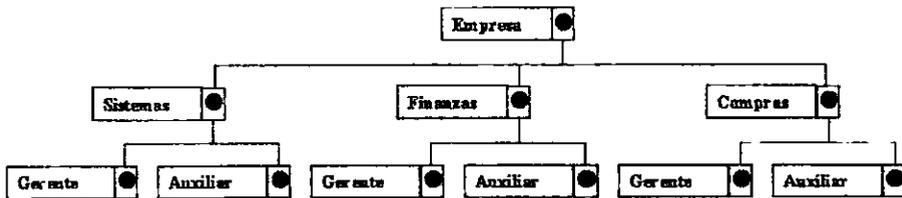


Ilustración I-4 Base de Datos Tipo Jerárquica

1.3. Bases de Datos Relacionales

Las BD relacionales constituyen una de las principales aportaciones en el modelado de BD ya que se encuentra fundamentado en el Álgebra Relacional, lo que permite asegurar modelos consistentes y bastante eficientes.

Las BD relacionales se han constituido como el principal modelo de datos y se constituyen al igual que su terminología como el punto de partida para el desarrollo de nuevas implementaciones y modelos, tales como el caso de las BD multidimensionales.

Estas bases de datos se organizan en "Entidades" contenedoras de tuplas, las cuales a su vez están divididos en "Atributos", cada tupla se relaciona con otras Entidades a través de los valores que contienen en sus atributos.

Por ejemplo la información de la entidad empleados y la entidad departamento, se encuentran representadas en forma relacional como lo muestra la Ilustración. I-5.

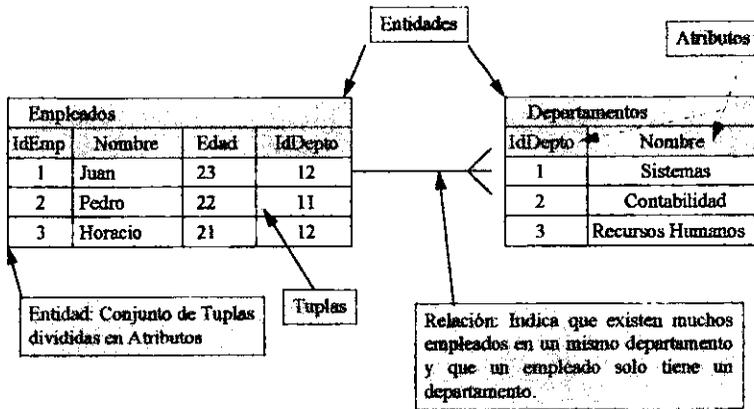


Ilustración I-5 Modelo Relacional y sus Elementos

Una entidad es cualquier objeto acerca del cual deseamos registrar información. Cada entidad tiene uno varios atributos, un atributo es una cualidad de la entidad.

Existe diferencia en la terminología empleada para este tipo de BD. Algunos libros utilizan el concepto de tablas, campos y registros, para describir entidades, atributos y tuplas respectivamente.

Terminología:

Términos Formales	Otros Términos
Tuplas	Tabla relacional o Tabla
Atributos	Columna o campo
Entidades o relaciones	Renglón o registro

En este trabajo en adelante se utilizarán los términos formales.

Las Propiedades que se deben considerar para la definición de Entidades son:

1. Cada atributo tiene un valor simple ó atómico, es decir, no son grupos repetitivos o arreglos.
2. Los valores en un mismo atributo son del mismo tipo y vienen del mismo dominio. Dominio es un conjunto de valores posibles que puede tener un atributo.
3. Cada tupla es única, existe por lo menos uno o varios atributos que diferencian una tupla de otra. Estos atributos se llaman *Llaves Primarias*.
4. Cada atributo tiene un solo nombre.
5. La secuencia de las columnas es indistinto. Las columnas pueden ser obtenidas en cualquier orden; el beneficio de esta propiedad es permitir a varios usuarios compartir la misma entidad sin conocer como está organizada, además de que si cambia la estructura física no afecta su relación con otras entidades.
6. La secuencia de las tuplas es indistinto. Es análogo a la propiedad anterior solo que aplica a tuplas. La ventaja de esta propiedad es que las tuplas se pueden obtener en diferente orden y secuencia.

Una **relación** se define como una asociación de dos o más Entidades. Las relaciones son expresadas mediante los valores de los atributos de una o varias llaves

primarias de una entidad, mismas que se encuentran definidas como foráneas en una o varias Entidades.

Una llave Primaria es uno o varios atributos que diferencian una tupla de otra; una llave foránea es uno o varios atributos de una entidad que hacen referencia a una llave primaria de otra entidad.

Tipos de Relaciones entre entidades:

1.Una a una: Es cuando una llave (primaria o foránea) en una entidad se relaciona con exactamente con una llave (primaria o foránea) en otra entidad.

2.Una a muchos: Es cuando una llave (primaria o foránea) de una entidad se relaciona con al menos una llave (primaria o foránea) de otra entidad.

3.Muchas a muchos: Es cuando una llave (primaria o foránea) de una entidad se relaciona con al menos una llave (primaria o foránea) de otra entidad y viceversa.

4.Cero a uno o a muchos :En este caso es cuando alguna de las relaciones anteriores no existe correspondencia entre una entidad y otra, es decir, puede o no existir una llave(primaria o foránea) en la otra entidad.

La manipulación y navegación en la BD depende de la integridad de los datos, esto significa que los datos deben ser consistentes y correctos. Existen dos reglas para asegurar la integridad de los datos: la integridad de la entidad y la integridad referencial.

La *Integridad de la Entidad* significa que la llave primaria no debe tener valores nulos (algo que no tiene valor), ya que esta se utiliza para identificar una única tupla.

La *Integridad Referencial* nos indica que una llave foránea puede tener valor nulo ó si tiene valor este debe existir en la llave primaria de la entidad a la que hace referencia.

La manipulación de las entidades y sus relaciones se basa en el *Álgebra Relacional*. El álgebra relacional consiste en un conjunto de operadores que trabajan sobre las relaciones. Cada uno de estos operadores toman una o más relaciones como entrada y producen una relación como salida. Existen ocho tipos de operadores y son las siguientes:

- **Unión:** Es formada por la agregación de tuplas de una entidad con las tuplas de otra entidad, y se eliminan las tuplas repetidas: por ejemplo la unión de la Entidad A con la B

Entidad A	Entidad B	A \cup B																												
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">x</th> <th style="width: 90%;">y</th> </tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>2</td><td>Azul</td></tr> <tr><td>3</td><td>Café</td></tr> </tbody> </table>	x	y	1	Amarillo	2	Azul	3	Café	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">x</th> <th style="width: 90%;">y</th> </tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>4</td><td>Verde</td></tr> <tr><td>5</td><td>Negro</td></tr> </tbody> </table>	x	y	1	Amarillo	4	Verde	5	Negro	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">x</th> <th style="width: 90%;">y</th> </tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>2</td><td>Azul</td></tr> <tr><td>3</td><td>Café</td></tr> <tr><td>4</td><td>Verde</td></tr> <tr><td>5</td><td>Negro</td></tr> </tbody> </table>	x	y	1	Amarillo	2	Azul	3	Café	4	Verde	5	Negro
x	y																													
1	Amarillo																													
2	Azul																													
3	Café																													
x	y																													
1	Amarillo																													
4	Verde																													
5	Negro																													
x	y																													
1	Amarillo																													
2	Azul																													
3	Café																													
4	Verde																													
5	Negro																													

- **Diferencia:** La diferencia de dos entidades es una relación que contiene aquellas tuplas que ocurren en la primera entidad pero no en la segunda. Basándonos en las dos entidades anteriores A - B queda de la siguiente forma:

A - B	B - A												
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">x</th> <th style="width: 90%;">y</th> </tr> </thead> <tbody> <tr><td>2</td><td>Azul</td></tr> <tr><td>3</td><td>Café</td></tr> </tbody> </table>	x	y	2	Azul	3	Café	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">x</th> <th style="width: 90%;">y</th> </tr> </thead> <tbody> <tr><td>4</td><td>Verde</td></tr> <tr><td>5</td><td>Negro</td></tr> </tbody> </table>	x	y	4	Verde	5	Negro
x	y												
2	Azul												
3	Café												
x	y												
4	Verde												
5	Negro												

- **Intersección:** La intersección de dos entidades es una relación que contiene tuplas comunes. Por ejemplo:

Entidad A	Entidad B	$A \cap B$																				
<table border="1" style="border-collapse: collapse;"> <thead> <tr><th>x</th><th>y</th></tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>2</td><td>Azul</td></tr> <tr><td>3</td><td>Café</td></tr> </tbody> </table>	x	y	1	Amarillo	2	Azul	3	Café	<table border="1" style="border-collapse: collapse;"> <thead> <tr><th>x</th><th>y</th></tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>4</td><td>Durazno</td></tr> <tr><td>5</td><td>Negro</td></tr> </tbody> </table>	x	y	1	Amarillo	4	Durazno	5	Negro	<table border="1" style="border-collapse: collapse;"> <thead> <tr><th>x</th><th>y</th></tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td></tr> </tbody> </table>	x	y	1	Amarillo
x	y																					
1	Amarillo																					
2	Azul																					
3	Café																					
x	y																					
1	Amarillo																					
4	Durazno																					
5	Negro																					
x	y																					
1	Amarillo																					

- **Producto:** Es también llamado producto cartesiano, es la combinación de cada tupla de una entidad con cada tupla de una segunda entidad. Por ejemplo:

Entidad A	Entidad B	$A \times B$																																																								
<table border="1" style="border-collapse: collapse;"> <thead> <tr><th>x</th><th>y</th></tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>2</td><td>Azul</td></tr> <tr><td>3</td><td>Café</td></tr> </tbody> </table>	x	y	1	Amarillo	2	Azul	3	Café	<table border="1" style="border-collapse: collapse;"> <thead> <tr><th>x</th><th>y</th></tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>4</td><td>Durazno</td></tr> <tr><td>5</td><td>Negro</td></tr> </tbody> </table>	x	y	1	Amarillo	4	Durazno	5	Negro	<table border="1" style="border-collapse: collapse;"> <thead> <tr><th>A_x</th><th>A_y</th><th>B_x</th><th>B_y</th></tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td><td>1</td><td>Amarillo</td></tr> <tr><td>1</td><td>Amarillo</td><td>4</td><td>Durazno</td></tr> <tr><td>1</td><td>Amarillo</td><td>5</td><td>Negro</td></tr> <tr><td>2</td><td>Azul</td><td>1</td><td>Amarillo</td></tr> <tr><td>2</td><td>Azul</td><td>4</td><td>Durazno</td></tr> <tr><td>2</td><td>Azul</td><td>5</td><td>Negro</td></tr> <tr><td>3</td><td>Café</td><td>1</td><td>Amarillo</td></tr> <tr><td>3</td><td>Café</td><td>4</td><td>Durazno</td></tr> <tr><td>3</td><td>Café</td><td>5</td><td>Negro</td></tr> </tbody> </table>	A _x	A _y	B _x	B _y	1	Amarillo	1	Amarillo	1	Amarillo	4	Durazno	1	Amarillo	5	Negro	2	Azul	1	Amarillo	2	Azul	4	Durazno	2	Azul	5	Negro	3	Café	1	Amarillo	3	Café	4	Durazno	3	Café	5	Negro
x	y																																																									
1	Amarillo																																																									
2	Azul																																																									
3	Café																																																									
x	y																																																									
1	Amarillo																																																									
4	Durazno																																																									
5	Negro																																																									
A _x	A _y	B _x	B _y																																																							
1	Amarillo	1	Amarillo																																																							
1	Amarillo	4	Durazno																																																							
1	Amarillo	5	Negro																																																							
2	Azul	1	Amarillo																																																							
2	Azul	4	Durazno																																																							
2	Azul	5	Negro																																																							
3	Café	1	Amarillo																																																							
3	Café	4	Durazno																																																							
3	Café	5	Negro																																																							

- **Proyección:** El operador de proyección obtiene un conjunto de valores de uno o varios atributos de una entidad, quitando los valores duplicados. Por ejemplo:

Entidad A	$\Pi(x)$																						
<table border="1" style="border-collapse: collapse;"> <thead> <tr><th>x</th><th>y</th></tr> </thead> <tbody> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>1</td><td>Amarillo</td></tr> <tr><td>2</td><td>Azul</td></tr> <tr><td>2</td><td>Azul</td></tr> <tr><td>2</td><td>Azul</td></tr> <tr><td>3</td><td>Café</td></tr> <tr><td>3</td><td>Café</td></tr> </tbody> </table>	x	y	1	Amarillo	1	Amarillo	1	Amarillo	2	Azul	2	Azul	2	Azul	3	Café	3	Café	<table border="1" style="border-collapse: collapse;"> <thead> <tr><th>x</th></tr> </thead> <tbody> <tr><td>1</td></tr> <tr><td>2</td></tr> <tr><td>3</td></tr> </tbody> </table>	x	1	2	3
x	y																						
1	Amarillo																						
1	Amarillo																						
1	Amarillo																						
2	Azul																						
2	Azul																						
2	Azul																						
3	Café																						
3	Café																						
x																							
1																							
2																							
3																							

- **Selección:** Es algunas veces llamado restricción, obtiene un juego de tuplas de una entidad basado en los valores de sus atributos. Por ejemplo:

Entidad A

x	y
1	Amarillo
2	Azul
3	Café

$\sigma(x \geq 2)$

x	y
2	Azul
3	Café

- **Acoplamiento (Join):** Una operación de acoplamiento combina el producto, selección y posiblemente una proyección. El operador join horizontalmente combina datos de una tupla de una entidad de otra o de la misma entidad con ciertos criterios. El criterio envuelve una relación entre atributos de una entidad.

Si el criterio es basado en la igualdad de un atributo es llamada equijoin. Un join natural es un equijoin sin atributos redundantes.

Entidad A

x	y
1	Amarillo
2	Azul
3	Café

Entidad B

x	y
1	Amarillo
4	Durazno
5	Negro

Equijoin(A.x=B.x)

A.x	A.y	B.x	B.y
1	Amarillo	1	Amarillo

Join Natural

x	y
1	Amarillo

- **División:** Esta operación resulta en valores de un atributo en una entidad que se encuentran en todos los valores de otro atributo de otra entidad. Por ejemplo:

Entidad A		Entidad B		A ÷ B	
1	Azul	Azul		1	
2	Café	Negro			
3	Azul				
3	Bianco				
2	Azul				
2	Negro				
3	Negro				
3	Café				

El objetivo principal del álgebra relacional no es solo la obtención de los datos, sino es ayudar a escribir expresiones. La siguiente lista nos indica sus principales aplicaciones.

1. Definir el alcance de una recuperación de datos, es decir, definir los datos que se van a extraer como resultado de una petición a la BD.
2. Definir vistas, es decir, definir los datos que se podrán ver en forma de relación virtual.
3. Definir los derechos de acceso.
4. Definir restricciones de integridad, es decir, definir alguna regla que debe satisfacer la BD.

Normalización

Para el diseño de las BD relacionales se lleva a cabo un proceso de "normalización", el cual consiste en remover los grupos repetidos de una entidad con el

fin de que estén libres de datos redundantes, para así poder manejar la BD en forma adecuada y además asegurar que el mantenimiento de las tuplas sea confiable. Este proceso de normalización está basado en el concepto de las *"Formas Normales"*.

Existen cinco formas normales, las tres primeras fueron definidas por E. F. Codd, para entender la teoría de la normalización es necesario definir las *"Dependencias Funcionales"*.

El concepto de dependencias funcionales es la base para las tres primeras formas normales. Una entidad es funcionalmente dependiente cuando por lo menos uno de sus atributos depende del valor de otro atributo de la misma entidad, es decir que un atributo X identifica los valores de un atributo Y. Cuando una entidad es totalmente funcional dependiente, existe un atributo Z cuyo valor es identificado por un atributo X y Y, es decir que el valor de Z es identificado por ambos atributos X y Y, no solo es identificado por uno solo; aquí X y Y forman una llave primaria compuesta.

Para explicar las formas normales nos basaremos en el siguiente ejemplo referente a un registro de siniestros de una compañía ferrocarrilera. Cada tipo de siniestro en esta compañía está asociado a un tipo y un número de póliza, el cual puede ocurrir en un distrito cualquiera, mismo que pertenece a una división.

nro_siniestro	fecha_siniestro	descripcion_siniestro	division	distrito	muni_poliza	tipo_riesgo
1	02/02/1999	Robo de Auto	Chihuahua	Tolobampo	A320	Autos
2	02/02/1999	Descarrilamiento tren	Centro	Quebrada	A322	Todo riesgo
3	03/03/1999	Descarrilamiento tren	Centro	Tlalnepantla	A322	Todo riesgo
3	03/03/1999	Descarrilamiento tren	Centro	Tlalnepantla	A322	Todo riesgo
4	04/03/1999	Robo de dinero	Sur	La plaza	A323	Multipoliza

Ilustración I-6 Entidad de Siniestros No Normalizada

Primera Forma Normal: Se dice que una entidad está en primera forma normal si no tiene tuplas repetidas. Por lo tanto de la Entidad de siniestros quitamos la tupla repetida con *sild* =3.

nro_siniestro	fecha_siniestro	descripcion_siniestro	division	distrito	muni_poliza	tipo_riesgo
1	02/02/1999	Robo de Auto	Chihuahua	Tolobampo	A320	Autos
2	02/02/1999	Descarrilamiento tren	Centro	Quebrada	A322	Todo riesgo
3	03/03/1999	Descarrilamiento tren	Centro	Tlalnepantla	A322	Todo riesgo
4	04/03/1999	Robo de dinero	Sur	La plaza	A323	Multipoliza

Ilustración I-7 Ejemplo de Eliminación de Tuplas Repetidas

Como se puede observar todavía existen datos redundantes, por ejemplo se repite lo que es la división, distrito, número de póliza y tipo de póliza de las tuplas con *siId* 2 y 3. La redundancia en esta forma normal nos trae como consecuencia los siguientes problemas:

- No podemos insertar una nueva póliza por que los demás atributos quedan vacíos.
- Si quisiéramos borrar el siniestro 4 también borraríamos la división sur, el distrito La plaza, la póliza A323 y el tipo de póliza Multipóliza.
- Al modificar el siniestro 1 colocando otra división, desapareceríamos la división de Chihuahua.

Segunda Forma Normal: Se dice que una entidad esta en segunda forma normal si ya esta en primera forma normal y los atributos no llave son totalmente dependientes de una llave primaria.

Para pasar cualquier entidad a segunda forma normal se tienen que hacer los siguientes pasos:

1. Identificar los atributos llave y los atributos que dependen de ellos
2. Crear una nueva entidad para los atributos llave junto con los atributos que identifican.

En nuestro ejemplo los campos llave son :

- *siId*: Este identifica a los atributos *siFecha*, *siDesc*
- *piId*: Identifica a *piTipo*

Quedando las entidades de la siguiente forma:

SINIESTROS

id	fecha	descripcion	dvDivision	diDistrito	id
1	02/02/1999	Robo de Auto	Chihuahua	Tolobampo	A320
2	02/02/1999	Descarrilamiento tren	Centro	Quebrada	A322
3	03/03/1999	Descarrilamiento tren	Centro	Tlalnepantla	A322
4	04/03/1999	Robo de dinero	Sur	La plaza	A323

POLIZAS

id	descripcion
A320	Autos
A322	Todo riesgo
A3232	Multipóliza

Ilustración 1-8 Entidades en Segunda Forma Normal

En este caso existe todavía el problema de que si borramos o modificamos un siniestro corremos el riesgo de borrar una división o distrito.

Tercera Forma Normal: Una entidad está en tercera forma normal si ya esta en segunda forma normal y los atributos no llave son transitoriamente independientes de la llave primaria, es decir, que los atributos no llave dependen de otros atributos no llave.

En este ejemplo la entidad de pólizas ya está en tercera forma normal, pero la entidad de siniestros tiene el atributo *diDistrito* que depende de *dvDivision*, por lo tanto estos dos se llevan a una entidad llamada Divisiones y a otra de Distritos.

SINIESTROS

1	02/02/1999	Robo de Auto	1	1	A320
2	02/02/1999	Descarrilamiento tren	2	1	A322
3	03/03/1999	Descarrilamiento tren	2	2	A322
4	04/03/1999	Robo de dinero	3	3	A323

DIVISIONES

1	Chihuahua
2	Centro
3	Sur

DISTRITOS

1	1	Tolobampo
2	1	Quebrada
2	2	Tlalnepantla
3	3	La plaza

Ilustración I-9 Entidades en Tercera Forma Normal

La ventaja de estar en esta forma normal es de que eliminamos los datos redundantes y los datos son más independientes, en el ejemplo se eliminan los problemas que se mencionó para cada forma normal.

Con este nuevo diseño si se inserta un nuevo distrito no es necesario tener un siniestro en este distrito; un siniestro puede ser borrado o modificado sin afectar alguna división o distrito.

Cuarta y Quinta Formas Normales: Estas dos formas normales son avanzadas y nos ayudan a reorganizar las relaciones. Existen dos dependencias más a las antes descritas, que son: las *multivaluadas*, y las de *Reunión*.

La dependencia *multivaluada* existe cuando un atributo X multidetermina a un atributo Y. Esta dependencia se produce cuando hay una relación de muchos a muchos. Por ejemplo una entidad que contenga una relación de profesores, materia e institución.

Profesor	Materia	Institución
María Jiménez	BD de red	UNAM
Eusebio Cruz	BD jerarquicas	UNAM
Jimena Ortiz	BD de red	UNAM
Jimena Ortiz	BD relacionales	UNVM
Patricia Hernández	BD jerarquicas	UNVM

Ilustración I-10 Entidad en Tercera Forma Normal

La Entidad de la Ilustración I-10 como podemos ver se encuentra en tercera forma normal puesto que no existe una dependencia transitoria; la institución no depende de las materias ni las materias del profesor; aquí existe una relación de muchos a muchos: muchas materias existen para un profesor, muchas materias se dan en una institución y un profesor pertenece a muchas instituciones.

En esta entidad si existen datos redundantes, aunque ya este en tercera forma normal; la desventaja con esta dependencia de datos, es que si deja de existir la materia de BD de red, también dejan de existir los profesores María Jiménez y Jimena Ortiz. Si queremos Insertar un nuevo profesor a la entidad, quedarían nulos los atributos de materia e institución.

Una entidad se dice que esta en cuarta forma normal si ya esta en tercera forma normal y no tiene dependencias multivaluadas. La entidad anterior queda en cuarta forma normal como se muestra en la Ilustración I-11.

Profesores

prof	Nombre
1	María Jimenez
2	Eusebio Cruz
3	Jimena Ortiz
4	Patricia Hernández

Materias

matr	Materia
1	BD de red
2	BD jerarquicas
3	BD relacionales

Instituciones

inst	Institución
1	UNAM
2	UNVM

Materias impartidas

Profesor	Materia	Institución
1	1	1
2	2	1
3	1	1
3	3	2
4	2	2

Ilustración I-11 Entidades en Cuarta Forma Normal

La dependencia de Reunión significa que una entidad, después de haber sido descompuesta en tres o más entidades pequeñas, debe ser capaz de reacomplarse otra vez en la forma original de la tabla, cuando la entidad no esta en quinta forma normal el resultado de la operación anterior produce tuplas adicionales (espuria). Basándose en la Ilustración. I-11 se tratará de probar si esta relación se encuentra o no en quinta forma normal.

Se dividirá primero en tres diferentes proyecciones. Ilustración. I-12

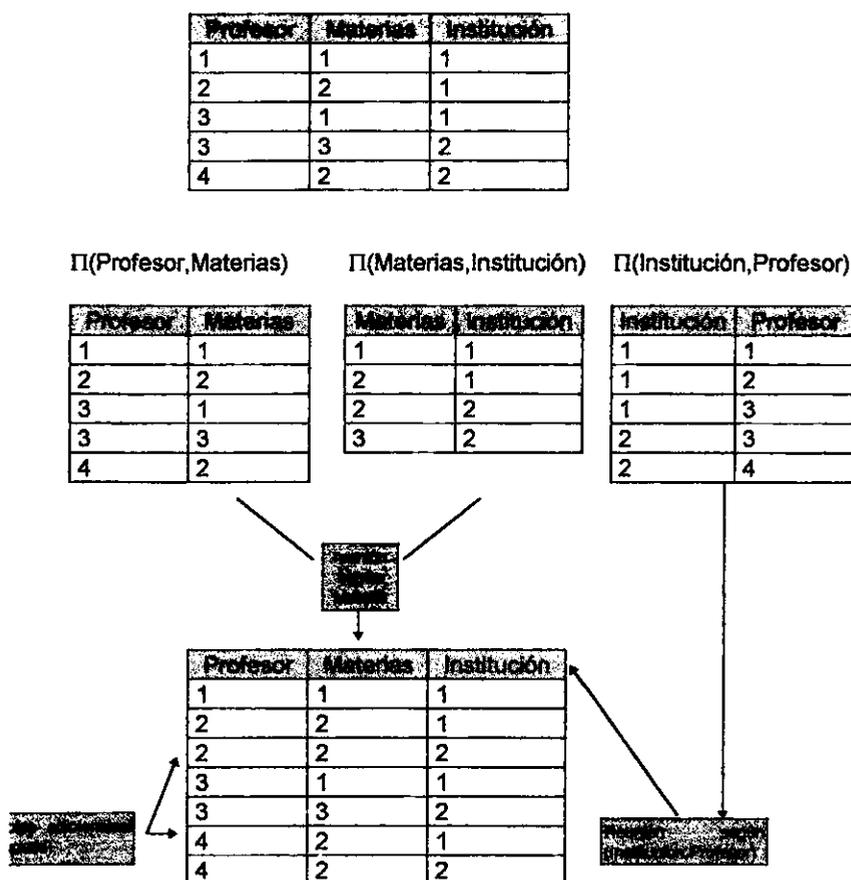


Ilustración I-12 Ejemplo de Dependencia de Reunión

Como se puede ver en la Ilustración I-12 existen tuplas adicionales y por lo tanto esta entidad no está en quinta forma normal. Una entidad de este tipo se divide en dos entidades más, para que las reuniones no produzcan tuplas adicionales. (Ilustración. I-13).

Profesor	Materias
1	1
2	2
3	1
3	3
4	2

Profesor	Institución
1	1
2	1
3	1
3	2
4	2

Ilustración I-13 Entidades en 5FN

1.4. Bases de Datos Orientadas a Objetos

Una forma innovadora en el diseño de BD la constituyen las BD orientadas a objetos (BDOO). Las BDOO no se basan en registros como en las BD relacionales, jerárquicas o de red, estas encapsulan datos y código en una misma unidad llamada objeto. La forma de comunicarse o relacionarse entre los objetos es por medio de mensajes.

Un objeto esta compuesto por:

- Atributos: Estos contienen valores que definen el estado del objeto.
- Métodos: Definen el comportamiento del objeto.
- Mensajes: Son las solicitudes para comunicarse con otros objetos.

Por ejemplo :

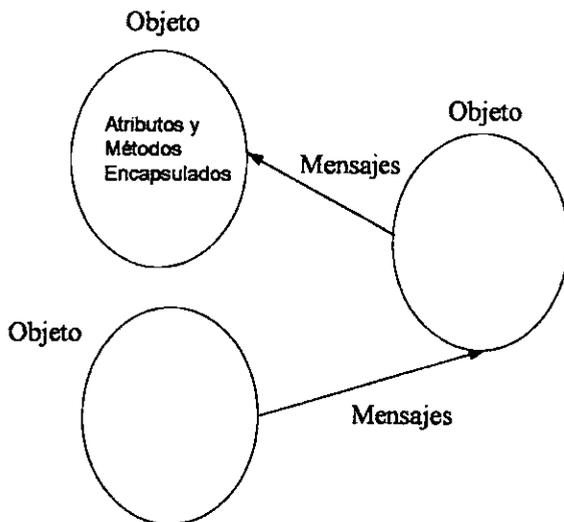


Ilustración I-14 Componentes de una BDOO

Este tipo de Bases de Datos fueron creados en base al paradigma de programación orientada a objeto como es el caso de C++, por lo que igualmente conservan las características de polimorfismo, herencia, encapsulamiento.

El modelado de estas BD es muy complejo y no cuenta con un fundamento robusto como es el caso de las BD relacionales, actualmente se encuentran en pleno desarrollo por lo que no están muy difundidas.

1.5. Bases de Datos Multidimensionales

Como ya se mencionó anteriormente las BD de relacionales en su forma conceptual están formadas por una matriz de dos dimensiones: una dimensión son las

tuplas y otra son los atributos; en las BD dimensionales se le agrega una dimensión más que puede ser el tiempo u otro dato. Sus datos tienen una estructura jerárquica siendo esta, la base para indicar el nivel de detalle con el cual se quiere ver la información. La forma de ver estos datos conceptualmente es de forma cúbica y cada dimensión tiene una estructura jerárquica. Su estructura se analizará a detalle en el siguiente capítulo.

La finalidad de este tipo de BD es que su modelo contempla datos jerárquicos y por cada nivel un concentrado de estos datos.

Este tipo de tecnología es una propuesta para mejorar la toma de decisiones, y es la utilizada por los sistemas OLAP.

Sistemas OLAP

Para una mejor explicación de esta tecnología, se darán algunas definiciones obtenidas de información en el WEB.

“OLAP es una categoría en la tecnología de software que permite a los analistas, administradores y ejecutivos echar un vistazo a los datos a través de un rápido, consistente e interactivo acceso a una amplia variedad de posibles vistas de información, que ha sido transformada de datos atómicos y que reflejan la dimensionalidad real de la empresa como la entiende el usuario.”¹

¹Datamation www.datamation.com/Plugin/issues/1996/april15/04bevalgis.html

"OLAP describe una clase de herramientas que pueden extraer y presentar datos multidimensionales desde diferentes puntos de vista. Diseñado para administradores que buscan darle sentido a su información, OLAP estructura sus datos jerárquicamente, la manera en que los administradores piensan en su empresa. Las funciones de OLAP incluyen análisis, distintos niveles de detalle, sumarización de datos y rotación de datos para vistas comparativas."²

De las definiciones anteriores se puede describir a OLAP como una tecnología que permite a los analistas de una organización, ver su información desde diferentes puntos de vista en forma interactiva y rápida. Esta tecnología presenta al usuario los datos en forma dimensional, tal y como el los entiende.

Como se describió anteriormente la forma de organizar los datos en una BDM es cúbica, sin embargo, la forma física de almacenamiento puede ser a través de un formato multidimensional, en una estructura relacional estándar, ó la combinación de ambos. Siendo esto posible su acceso a través de tres métodos: el MOLAP, el ROLAP, y el HOLAP, que a continuación se describen.

Procesamiento Analítico Multidimensional en Línea (MOLAP).

Este método tiene las siguientes principales características:

- Guarda la información en una estructura multidimensional.
- Ofrece el mejor funcionamiento en la obtención de la información (query), por que incluye información consolidada.

²OLAP www2.computerworld.com/home/features.nsf/all/981130qs

- Puede consumir gran cantidad de espacio, por que los cubos con datos consolidados pueden ser de gran variedad.

Procesamiento Analítico Relacional en Línea (ROLAP)

Este método tiene las siguientes principales características:

- Guarda las consolidaciones en entidades en un modelo Entidad Relación (E-R)
- Causa una baja respuesta en comparación con MOLAP ó HOLAP.
- Es la mejor opción de almacenamiento para medios que casi no son consultados.

Procesamiento Analítico Híbrido en Línea(HOLAP).

El procesamiento analítico híbrido en línea es un método híbrido de MOLAP y ROLAP y se distingue por :

- Las consolidaciones son guardadas en una estructura MOLAP.
- La información base se mantiene una estructura relacional.
- Las peticiones que son solo a información consolidada son tan rápidas como en MOLAP.
- Las peticiones que requieren información de un nivel consolidado a detallado son más lentas que en MOLAP, pero más rápidas que en ROLAP.
- Los cubos HOLAP son más pequeños que en MOLAP por que solo guardan información consolidada.

CAPITULO II. BD Multidimensionales(BDM)

II.1. Conceptos y Características

Aunque aún no se ha definido con precisión quién ó quienes fueron los iniciadores de este tipo de tecnología, se sabe con certeza que las BD Multidimensionales son principalmente usadas por los sistemas OLAP. El presidente de Kenan Systems; Dr. Kenan Sahin, menciona que esta tecnología ya se aplicaba en 1960 dentro de los productos Express y APL.

Antes de continuar, quiero resaltar que las BDM no reemplazan a las BD relacionales, sino simplemente que una BDM es una técnica para conceptualizar modelos de negocio como un conjunto de características que describen sus facetas.

“Un buen modelo conceptual de modelo de datos multidimensional puede ser implementado en una BD relacional, en una BD multidimensional ó en una BD orientada a objetos”³. Esto quiere decir que el modelo multidimensional es una técnica de diseño de BD por lo tanto es independiente del motor de la base de datos.

Para entender claramente el concepto de las BDM se empezará por explicar los tipos de información que se maneja dentro de una empresa:

Primeramente están los *Datos Operacionales*, estos provienen de los sistemas OLTP (Procesamiento de Transacciones en Línea). Estos sistemas OLTP son

³MULTIDIMENSIONAL MODELING www.netmar.com/~nraden/tw0196_1.html

optimizados para el almacenamiento de grandes volúmenes de datos obtenidos de registro en registro. Los sistemas OLTP son comúnmente conocidos como sistemas de punto de venta ó los sistemas de registro de ordenes, entre otros, los sistemas operacionales están basados principalmente sobre los sistemas de BD relacionales.

La información operacional por lo general no se le presenta a un Gerente, a esta persona lo que le interesa es la información con cierto nivel de concentración, pues necesita manejar el concepto del entorno que está administrando, para ello requiere otro nivel de información que se genera al convertir un sistema operacional en un sistema informacional.

Los sistemas OLTP con grandes volúmenes de información no son muy eficientes en cuanto al tiempo de procesamiento para obtener los reportes de un sistema informacional a partir del sistema operacional, y es aquí donde entran los sistemas OLAP junto con las BDM.

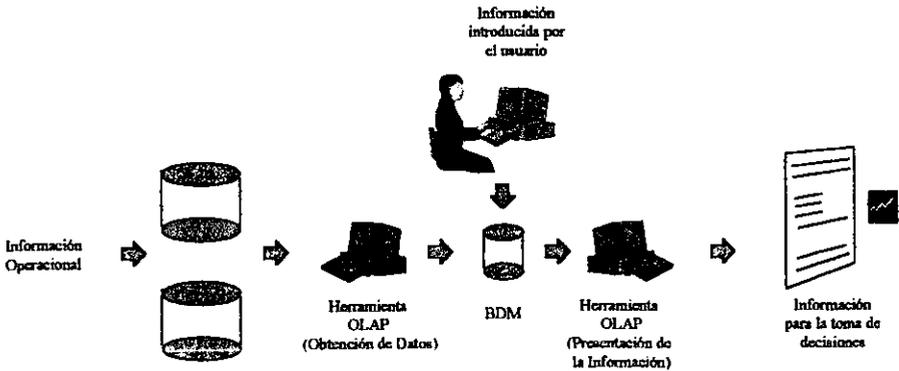


Ilustración II-1 Integración de la información en una BDM

La tecnología de las BDM fue diseñada para satisfacer las necesidades de los administradores, ejecutivos y analistas que quieren ver sus datos en un formato particular. Las BDM son también muy útiles para guardar datos en varios niveles de consolidación (por departamento, por división ó por compañía) con el fin de navegar rápidamente entre los niveles.

Una base de datos multidimensional se puede ver como un cubo y se diseña organizando los datos en términos de las actuales dimensiones de los negocios en una organización.

Basándonos en un típico ejemplo se entenderá mejor: en un negocio las ventas pueden ser clasificadas por tipos de productos, clientes, fecha de venta, vendedores, localización y cantidad vendida.

Un eje o dimensión puede estar representada por el tipo de producto, otra por la localización y otra dimensión por la fecha de venta. Si se quiere saber el total de las ventas hechas de un producto C, en la localización 2, en el mes de Enero este quedaría localizado de la siguiente forma:

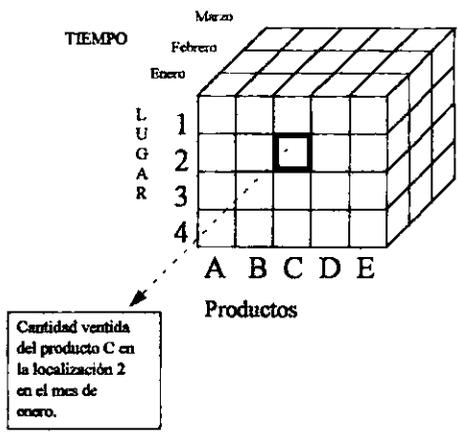


Ilustración II-2 Acceso a una BDM

La intersección en la localización 2, producto C y mes Enero es el dato que estamos buscando.

A diferencia de las BD comunes aquí es posible que en cada dimensión almacene información calculada, en el ejemplo anterior es posible tener guardada el total de ventas por cada dimensión, por ejemplo:

	A	B	C	D	E	Total
1	10	20	26	48	84	188
2	39	39	9	93	78	258
3	7	73	9	8	8	105
4	3	4	8	23	3	41
Total	59	136	52	172	173	592

— Elementos Consolidados

Ilustración II-3 Elementos Consolidados

Cada *celda* contiene datos fuente originales llamados *entradas* y los totales que son llamados *salidas*. Los números 1, 2, 3, 4 son *miembros de las entradas* en la *dimensión* de la Localización.

Los totales por producto son *miembros de las salidas* en la *dimensión* de Productos; y los totales por Localización son *miembros de las salidas* en la *dimensión* de Localización.

Aquí podemos ver que teniendo almacenada la información consolidada, es más rápido obtener información aún cuando el almacenamiento sea ineficiente.

Las *dimensiones* se pueden asemejar a los atributos en una BD relacional; y las *celdas* a tuplas.

Estos cubos se pueden rotar en cualquier dirección y ver la información de diferente manera. Por ejemplo de la ilustración il-2 tenemos la comparación de Productos vs Localización, si lo rotamos tenemos la comparación de Productos vs Tiempo.

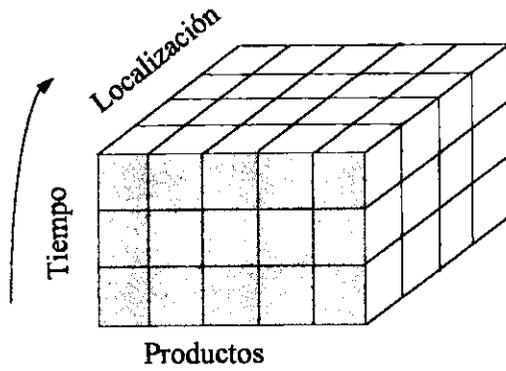


Ilustración II-4 Rotación de Cubo

Si lo volvemos a rotar podemos ver la comparación de Localización vs Tiempo.

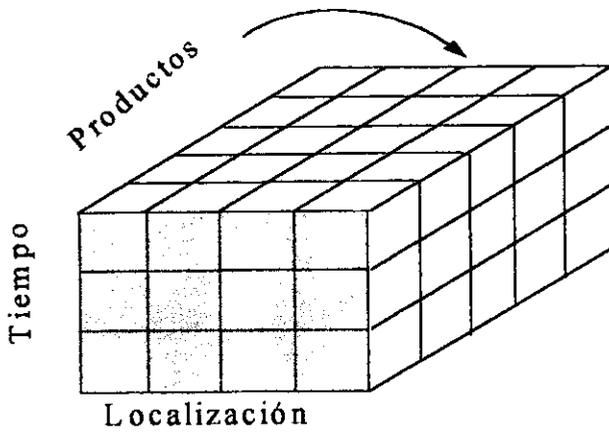


Ilustración II-5 Rotación de Cubo II

II.2. Jerarquías en las dimensiones

Las BDMs guardan la información de la misma manera en que es vista y la organizan de acuerdo con la visión empresarial. Esta forma de representar los datos nos da un claro entendimiento de la información. Las BDMs tienen un cierto grado de normalización en sus entidades y hay veces que solo se almacena información resumida o consolidada dando con esto un acceso mucho más rápido que en un BD relacional.

Las jerarquías se utilizan para simplificar el modelo de Dimensiones. En la Ilustración II-6 se muestra un modelo en donde las dimensiones de Producto, Localización y tiempo guardan a la vez una jerarquía que puede representarse gráficamente de la siguiente forma:

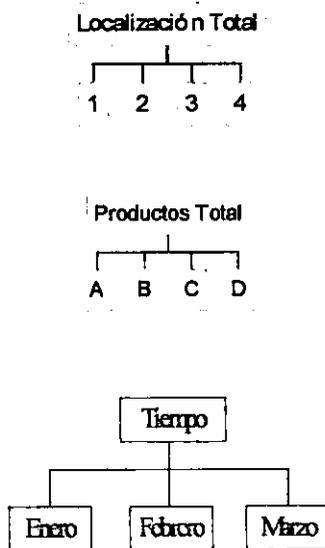


Ilustración II-6 Jerarquías en las dimensiones

En este caso los hijos no son padres y por lo tanto no tienen hijos, es decir, solo tienen dos niveles jerárquicos. Ahora supongamos que en la dimensión de localización cada hijo tuviera hijos, es decir, suponer que las localizaciones se subdividan en ciudades.

Una posible solución al ejemplo anterior es que se dimensione en Localizaciones, ciudades y productos

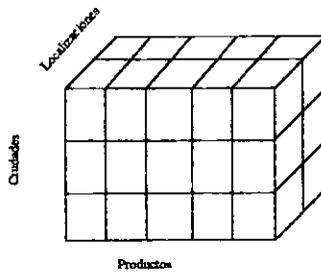


Ilustración II-7 Muestra de Cubo

La desventaja de optar por esta solución, es la dispersidad de los datos que provoca el hacer que una ciudad pertenezca a una localización en especial, implica que las demás celdas estén vacías. Otro problema es cuando el usuario no solo quiere ver sus datos por ciudad sino por sucursal o por color del producto, esto representarlo en más de tres dimensiones es inimaginable.

La mejor manera de resolver un problema como este es tener jerarquías en las dimensiones. Por ejemplo, las localizaciones divididas en ciudades y las ciudades en sucursales.

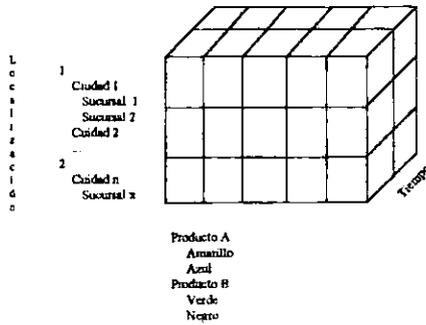


Ilustración II-8 Jerarquización de dimensiones

II.3. Tipos de esquemas

El esquema entidad-relación utilizado para los sistemas OLTP está optimizado para el almacenamiento de grandes cantidades de información, la organización de esta información es en base a las formas normales, es decir que no existe redundancia en los datos. Por otro lado los esquemas que se utilizan en los sistemas OLAP están optimizados para la explotación rápida de la información, esta información tiene un cierto grado de normalización, es decir, existen datos repetidos.

Los sistemas OLTP tratan con eventos individuales, mientras que sistemas OLAP tratan con múltiples eventos y situaciones que ocurren a través del tiempo.

Sistemas OLTP	Sistemas OLAP
<ol style="list-style-type: none"> 1. Detectan un evento 2. Responden al evento 3. Investiga acerca de estado del evento 	<ol style="list-style-type: none"> 1. Integran muchos eventos 2. Integran las respuestas del sistema a los eventos 3. Evalua las relaciones de los eventos, después del hecho, a gran escala.

Un diagrama multidimensional utiliza los mismos componentes que un esquema Entidad - Relación; los dos esquemas que se utilizan en los datos multidimensionales son los siguientes:

1. Tipo Estrella
2. Tipo Copo de nieve (Snow Flake)

Para el diseño de datos multidimensionales se necesitan identificar las entidades y los atributos de cada entidad para un problema determinado, pero además es necesario definir sus influencias:

- Identificar las entidades del negocio que se necesitan modelar.
- Determinar las actividades clave del negocio y sus influencias, por ejemplo las actividades clave de un negocio son sus ventas, inversiones y sus influencias son

el tiempo. Por influencias se quiere decir aquellas entidades que servirán como dimensiones en nuestro modelo.

Básicamente se tienen cuatro clases de entidades:

1. Una historia de las transacciones.
2. Personas, lugares o cosas que describen a las transacciones.
3. Tiempo, este es considerado por lo general en un análisis.
4. Consolidados, los cuales nos ayudan a obtener la información a un determinado nivel.

Una vez identificadas las entidades se procede a hacer el modelo multidimensional, para ello se utilizan los diagramas multidimensionales.

Esquema tipo Estrella: El nombre proviene de el patrón formado por las entidades y las relaciones cuando es representado en un diagrama de Entidad - Relación. Típicamente hay una actividad del negocio (Historia de Transacciones) en el centro de la estrella rodeado por gente, lugares, cosas, etc. que vienen junto con la actividad, es decir, el centro es una variable y lo que lo rodea son las dimensiones y sus jerarquías.

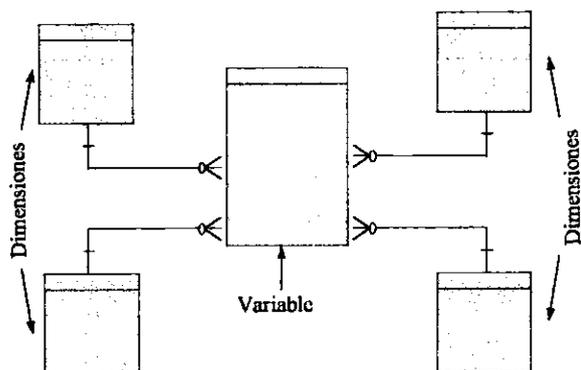


Ilustración II-9 Esquema de tipo Estrella

Los atributos dimensionales proveen de información descriptiva acerca de cada tupla en la entidad Variable. En el esquema las variables son llamadas como tabla de hechos.

Esta tabla de hechos tiene principalmente datos numéricos, los cuales pueden ser manejados para realizar diversas operaciones que nos sirven en el análisis, estas tablas deben contener llaves foráneas que le permitan relacionarse con las tablas de las dimensiones.

La relación que existe entre la tabla de hechos y las dimensiones es de uno a muchos; es decir, un registro de la dimensión se relaciona con muchos registros en la tabla de hechos. Estas dimensiones nos sirven también para guiarnos en la selección de los datos.

Principalmente las tablas de hechos contienen grandes volúmenes de información, mientras que las tablas de las dimensiones son pequeñas; esto nos da una ventaja en el rendimiento al relacionar tablas grandes con tablas pequeñas.

El tiempo es un componente crítico en el modelo. Los datos en los sistemas OLTP son orientados a los valores actuales. Los sistemas OLAP nos permiten analizar los datos y además de observarlos como cambian a través del tiempo.

Las actividades del negocio se analizan desde un tiempo específico a otro. Esto nos muestra los cambios y tendencias en periodos de tiempo. Estas tendencias son colocadas en donde se buscan las relaciones entre las actividades y sus dimensiones.

Teniendo una dimensión de tiempo podemos incluir unidades de tiempo que correspondan a algunos eventos significativos en la empresa. Estos eventos significativos serán utilizados para obtener reportes periódicos (temporalidad) cuando un Gerente lo requiera para analizar un concentrado de actividades del negocio. Basándonos en la ilustración II-9 el esquema de tipo estrella queda de la forma siguiente:

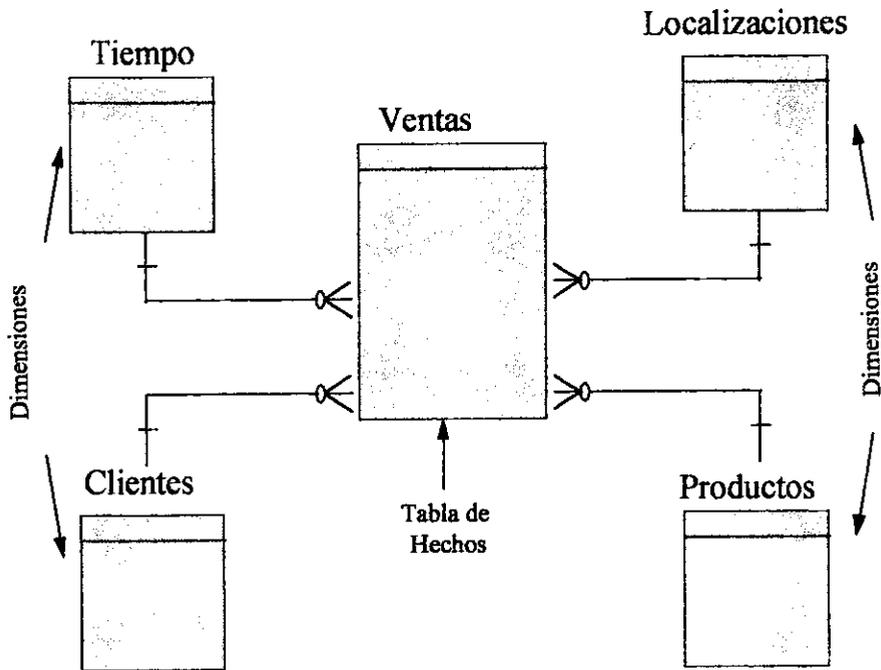


Ilustración II-10 Esquema de tipo Estrella de la Ilustración II-9

Esquema de Tipo Snowflake: Las tablas de dimensión pueden tener varios niveles o jerarquías, por ejemplo una unidad de tiempo escogida determinará el nivel de los datos que se quieren ver de la tabla de hechos.

Cuando en el modelo se tienen jerarquías en las dimensiones se llama esquema de tipo Snowflake.

En este tipo de esquema podemos incluir múltiples tablas de hechos relacionados entre si. Por ejemplo podemos tener una tabla de ventas, otra de inversiones, otra de inventarios.

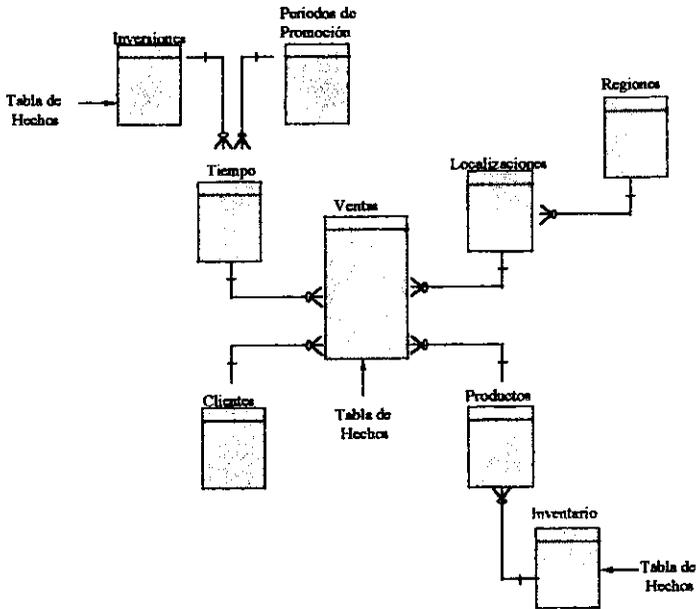


Ilustración II-11 Esquema de tipo Snowflake

Las tablas de hechos pueden contener consolidados de datos del más bajo nivel de detalle. Estos consolidados pueden ser definidos en varios niveles del mismo hecho.

Los consolidados nos ayudan en el rendimiento de la obtención de los datos. Los resultados de las peticiones de información a la BD es más rápido si se usan por ejemplo 100 registros precalculados a 1000000 registros al nivel más bajo de detalle.

En una tabla de ventas podemos consolidar datos por día y por región Ilustración II-12.

	A	B	C	D	E	Total
1	10	20	26	48	84	188
2	39	39	9	93	78	258
3	7	73	9	8	8	205
4	3	4	8	23	3	41
Total	59	136	52	172	173	592

Elementos
Consolidados
de la Región 1

Consolidado
por región y
producto

	A	B	C	D	E	Total
1	59	136	52	172	173	592
2	139	139	19	193	178	668
3	17	173	19	18	18	245
4	3	4	8	23	3	41
Total	218	452	98	406	372	1546

Ilustración II-12 Ejemplo de Tabla consolidada por región y producto

II.4. Reglas de E.F. Codd.

En 1993, E.F. Codd y Asociados publicaron un artículo, comisionado por Arbor Software (ahora Hyperion Solutions), titulado "Providing OLAP (Online Analytical Processing) to User-Analysts". El Dr. Codd es bien conocido por sus investigaciones desde 1960 a 1980 y es acreditado como el inventor del modelo de BD relacional y adepto de la tecnología OLAP, sin embargo sus reglas OLAP no son basadas en términos matemáticos como lo hizo en el modelo relacional.

No es muy clara su relación con esta actividad ya que este artículo se caracteriza como un documento de ventas más que como un documento académico.

El documento incluye catorce reglas divididas en cuatro grupos, doce de ellas se pueden extraer desde la mayoría de los sitios WEB de los vendedores de los productos

OLAP, las cuatro restantes no son muy difundidas. El objetivo de este documento es definir las características esenciales que deben estar incluidas en todos los sistemas OLAP.

El Doctor E.F. Codd divide estas reglas en relación a cuatro importantes características:

1. Características básicas.
2. Características especiales.
3. Características de los reportes.
4. Control de dimensiones.

CARACTERISTICAS BASICAS:

1. *Visión Conceptual Multidimensional:* Por naturaleza, un usuario empresarial tiene una visión multidimensional de la empresa. Por lo tanto, los modelos OLAP deben ser de naturaleza multidimensional. La parte central del funcionamiento OLAP debe incluir "slice and dice" (rebanado y picado) como parte de este requerimiento. La característica de "slice and dice" se refiere a la forma en que el usuario pueda consultar la información en la computadora, por ejemplo de toda información que se cuenta en la BDM se escoge una en especial a analizar(slice) y después esta información se pueda dividir en un nivel más detallado(dice).
2. *Manipulación de Datos Intuitivo :* La manipulación de los datos debe ser directamente a través de celdas en la vista, sin recurrir a menús o múltiples acciones. Uno asume que se está usando el mouse (o equivalente).

3. *Accesibilidad* : OLAP como un mediador. El Dr. Codd describe esencialmente los motores OLAP como la parte central de una arquitectura cliente/servidor o middleware, situado entre fuentes de datos heterogéneos y un cliente (front-end) OLAP.
4. *Extracción por lotes vs Interpretativo*: Este requerimiento requiere que los productos ofrezcan su propia BD para los estadíos de los datos OLAP y también ofrezcan un libre acceso a datos externos. El Dr. Codd esta uniendo las fases de los datos multidimensionales, además del precalculado de grandes cantidades de datos multidimensionales, junto con un transparente proceso de comportamiento del detalle.
5. *Modelos de análisis OLAP*: Los productos OLAP deben soportar los cuatro modelos de análisis (Categorico, Interpretativo, Contemplativo y Fomulaico), podemos describir esta característica como reportes parametrizados, picar y rebanar con "drill down"(de lo genérico a lo particular), análisis de tipo "que pasa si..." y modelos con fórmulas que definan las relaciones entre miembros y permitan personalizar los datos para una mejor visualización; respectivamente.
6. *Arquitectura Cliente-Servidor*. El componente servidor de herramientas OLAP debe ser lo suficientemente inteligente para que los diversos clientes se incorporen con un esfuerzo y programación de integración mínimos.
7. *Transparencia* : El lugar donde reside la herramienta analítica debe ser transparente para el usuario empresarial. OLAP debe existir dentro de la arquitectura de sistemas abiertos, permitiendo que la herramienta analítica se incorpore en cualquier parte que desee el usuario sin un efecto adverso sobre la funcionalidad de la herramienta anfitrión.

8. *Soporte multiusuario*: Las herramientas OLAP deben suministrar acceso concurrente (recuperación y actualización), integridad y seguridad para dar soporte a usuarios que necesitan trabajar de forma concurrente con el mismo modelo o crear diferentes modelos para los mismos datos.

CARACTERISTICAS ESPECIALES

9. *Tratamiento de datos no normalizados*: Este se refiere a la integración entre el motor OLAP y la fuente de datos no normalizados. Se apunta a que no se permita hacer cualquier modificación en el ambiente OLAP.
10. *El almacenaje de los resultados OLAP*: Deben separarse de los datos fuente.
11. *Extracción de valores faltantes*: Todos los valores faltantes (nulos) deben ser distinguidos de los valores cero.
12. *Tratamiento de los valores nulos*: Todos los valores nulos deben ser ignorados por el analizador OLAP.

CARACTERÍSTICAS DE LOS REPORTES:

13. *Reportes flexibles*: Usando un servidor OLAP y sus herramientas, un usuario puede manipular, analizar, sintetizar y observar los datos en cualquier forma deseada, incluyendo la creación de agrupamientos lógicos o la colocación de hileras, columnas y celdas juntas unas tras otras.
14. *Rendimiento consistente en los reportes*: No debe degradarse el rendimiento al aumentar la cantidad de dimensiones.

15. *Ajuste automático del nivel físico*: Se requiere que los sistemas OLAP ajusten sus sistemas físicos automáticamente para adaptar el tipo de modelo, volúmenes de datos y esparcidad.

CONTROL DE DIMENSIONES:

16. *Dimensionalidad genérica*: Cada dimensión de datos debe ser equivalente tanto en su estructura como en su capacidad operacional. Debe existir solo una estructura lógica para todas las dimensiones. Debe ser posible usar una en otra, cualquier función que se aplique a una dimensión.

17. *Dimensiones y niveles de adición ilimitados*: Un servidor OLAP debe alojar por lo menos quince dimensiones dentro de un modelo analítico común. Cada una de estas dimensiones genéricas debe permitir una cantidad ilimitada de adiciones definidas por el usuario y niveles de resumen dentro de una trayectoria de consolidación determinada.

18. *Operaciones cruzadas entre dimensiones sin restricción*: Todas las formas de cálculo deben ser permitidas en todas las dimensiones. Si los cálculos requieren definir varias fórmulas de acuerdo con el lenguaje, este último debe permitir que se calculen y manipulen datos a través de diversas dimensiones de datos, sin restringir las relaciones entre las celdas de datos y sin importar la cantidad de atributos comunes de datos que contenga cada celda.

Estas reglas fueron propuestas para tener un punto de partida sobre los datos que no es un punto de partida formal, y sirven para evaluar y comprender las necesidades y el uso de herramientas OLAP. Estas reglas no se llevan a cabo por completo como se

podrá ver en el capítulo tres, por ejemplo en las reglas se dice que no debe de afectar el rendimiento al aumentar la cantidad de dimensiones, más sin en cambio en los diferentes productos del mercado en el modelo de datos multidimensionales se recomienda lo contrario.

II.5. Comparación de una BDM con una BD relacional

" Una BD relacional típica puede sumar unos pocos cientos de registros por segundo; una típica BDM sumar números en tuplas y atributos en un promedio de 10000 por segundo o más."⁴

Lo anterior quiere decir que la BDM son particularmente útiles para sumarizar y agrupar datos para facilitar el análisis. Y solo trabajan con datos numéricos tales como valores, conteos, pesos y ocurrencias; dado que un dato alfabético no se puede sumarizar ni mucho menos tener un nivel de consolidación.

El modelo relacional es diseñado para soportar procesos operacionales, en contraste con el multidimensional este es diseñado para soportar los reportes y las necesidades analíticas de la gente que toma las decisiones.

Como ya se mencionó las BDM no reemplazan a las BD relacionales, sino más bien las BDM se complementan con las BD relacionales. Esta comparación de los dos tipos de BD es similar a diferenciar un auto con un trailer, cada uno de ellos fueron creados para fines distintos, el primero es para transportar gente y el otro para

⁴OLAP www.pilotsw.com/olap/olap.html

transportar mercancía pero al final los dos son medios de transportes; por otro lado las BD relacionales fueron creadas para soportar transacciones mientras que las BD multidimensionales fueron creadas para consultas complejas y rápidas para el análisis.

La diferencia entre un esquema de tipo estrella y un esquema relacional totalmente normalizada es la siguiente:

ESQUEMA DE TIPO ESTRELLA Y DENORMALIZADO	ESQUEMA RELACIONAL TOTALMENTE NORMALIZADO
Utilizan una denormalización para optimizar la velocidad en la obtención de la información. Sacrifica espacio por tiempo.	Esta totalmente normalizado, no tiene datos repetidos y sacrifica tiempo por espacio.
Las tablas de hechos solo guardan características numéricas	En todas las tablas se pueden guardar datos de transacción y de referencia
Son de consulta para el análisis	Son para mantener la consistencia de la información

II.6. Acceso a las BDM a través de SQL Multidimensional

En esta sección se hablará de cómo hacer una sentencia *Select* multidimensional y una comparación con una sentencia semejante en una BD relacional; además de una explicación breve de cómo construir y explotar la información de un cubo a través de los servicios de OLAP que vienen con la BD relacional de Microsoft SQL Server versión 7.0.

Dentro de los sistemas OLTP existe un lenguaje estructurado de consulta llamado SQL (Structured Query Language) el cual funciona con las operaciones del álgebra

relacional descrita en el capítulo 1.; este fue construido para manipular estructura de datos tabulares y solo regresa resultados de dos dimensiones.

Dentro de las BD multidimensionales también existe un lenguaje estructurado multidimensional (MSQL) el cuál como el SQL también maneja álgebra relacional, pero a diferencia de este regresa resultados de dos o más dimensiones.

En el SQL existe una sentencia para consolidar información (GROUP BY); en el MSQL ya no es necesaria esta instrucción ya que existen datos precalculados que hacen que la consulta sea rápida.

Estructura y resultados de una consulta de la sentencia SELECT en el SQL de los sistemas OLTP

Básicamente la estructura es la siguiente:

SELECT [Atributos de una o varias entidades]

FROM [Entidades a consultar]

WHERE [filtros o condiciones de la consulta]

GROUP BY [Atributos que determinan el agrupamiento de los resultados de los cálculos (suma,promedio,conteo) de las operaciones indicadas en el select]

ORDER BY [Atributos que determinan la ordenación de la consulta]

En la parte del *SELECT* se colocan todos los atributos de unas o varias entidades a consultar, en el *FROM* se colocan las entidades que contienen los atributos de la consulta, en la parte del *WHERE* contiene las restricciones de la consulta, en el *GROUP BY* indica que se agrupan las tuplas dependiendo de los valores de ciertos atributos y en la parte final se indica la ordenación

Por ejemplo: Basándose en los valores de la tabla de empleados de la ilustración. I-5, se hará la sentencia *SELECT* para obtener el nombre y edad de los empleados del departamento 12

Empleados			
IdEmp	Nombre	Edad	IdDepto
1	Juan	23	12
2	Pedro	22	11
3	Horacio	21	12

La sentencia *SELECT* quedaría de la siguiente forma:

```
SELECT Nombre, Edad
FROM Empleados
WHERE IdDepto = 12
```

El resultado de esta sentencia son los siguientes datos tabulares:

Nombre	Edad
Juan	23
Horacio	21

Estructura y resultados de una consulta de la sentencia *SELECT* en el SQL

Multidimensional

En la actualidad este tipo de lenguaje SQL no hay un estándar para que en todos los sistemas se utilice lo mismo; para explicar este punto se utilizará en el *SELECT* de

las expresiones MDX de los servicios OLAP de Microsoft. Estas expresiones MDX son las expresiones del lenguaje multidimensional de dicho software.

La sintaxis general de un SELECT multidimensional es la siguiente:

```
SELECT  [Ejes del cubo]
        CROSSJOIN [Nivel de consolidación de cada miembro
                  especificado en la parte del select ]
        [ON ROWS],
        [Ejes del cubo]
        CROSSJOIN [Nivel de consolidación de cada miembro
                  especificado en la parte del select ]
        [ON COLUMNS]
FROM    [Nombre del cubo]
WHERE   [Indicador de la parte a consultar del cubo (slicer)]
```

En la parte del SELECT se indica las dimensiones y/o miembros que se quieren visualizar a nivel de columna (ON COLUMNS) al igual que a nivel de renglón [ON ROWS]. En la sección del FROM se indica el cubo que se desea analizar, y en la parte final, en el WHERE se especifica la sección del cubo que se requiere analizar, también es conocido como el *slicer* que en español significa "rebanador".

Para este tipo de sentencias existen seis elementos que se deben entender para poder construir las sentencias las cuales son:

Cubos: Tal y como se ha estado definiendo en esta tesis, un cubo es un juego de dimensiones relacionadas que forman un "cuadrulado" de n-dimensiones. Cada punto en este es identificado únicamente por el juego de coordenadas los cuales son los valores que componen las dimensiones

Dimensiones: Específicamente una dimensión identifica el eje que se esta tratando de representar. Por ejemplo los productos y el tiempo son dimensiones.

Miembros: Es el valor en específico de cada dimensión Por ejemplo en la dimensión del tiempo los miembros "1997", "Semestre", etc.

Celdas: Una celda es el punto de intersección entre dos o más dimensiones. Las celdas contienen información de lo que se esta tratando de evaluar.

Jerarquías en las Dimensiones: Son los niveles de consolidación de cada dimensión, cada una de ellas contiene una combinación de múltiples miembros. Por ejemplo la dimensión del tiempo se divide en días, meses, años, semestres, etc.

Propiedades: Las propiedades representan características de cada miembro, por ejemplo dentro de los miembros de la dimensión productos se puede colocar el precio de cada uno de ellos.

Por ejemplo, basándose en el siguiente cubo y consulta queda la sentencia *SELECT* armada con se describe a continuación:

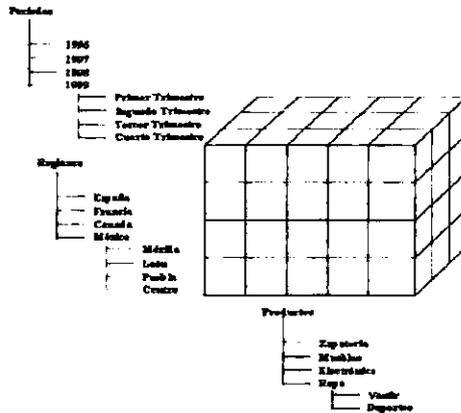


Ilustración II-13 Datos para Armar Una Sentencia SELECT Multidimensional

En base al cubo de la ilustración II-13, se necesita saber las ventas de la ropa de vestir y deportes en todos los años en España y en cada uno de los estados de la República Mexicana, para el año de 1999, la información se debe desglosar por periodo.

En este tipo de consulta en los sistemas OLTP son complicados en codificar y en su tiempo de respuesta suelen ser lentos cuando se requiere que la información se vea en un mismo cuadrículado , ya que en este caso se tendría que hacer lo siguiente:

1. Se tiene que saber específicamente en cuantos estados de la república se vende la ropa de vestir para poder desglosar las ventas por cada estado de la República
2. Se debe saber cuantos años existen en la historia de las ventas en España y por cada estado de la República, además se busca cada tipo de ropa por cada periodo del año de 1999
3. Una vez conocido la cantidad de valores de los dos puntos anteriores se crea una tabla temporal con ese mismo número de atributos.

4. Una vez creada la tabla temporal se inserta un registro a la vez por cada año, periodo, estado, país y tipo de producto

Por otro lado, en los sistemas OLAP para resolver este tipo de consultas solo se tiene que realizar el siguiente *Select*.

```
SELECT CROSSJOIN({Espana,México.CHILDREN}) ON COLUMNS,
          {Periodos.MEMBERS, 1999.CHILDREN} ON ROWS
FROM     Cubo_Ventas
WHERE    (Vestir, Deportes)
```

Este tipo de sentencias nos da un resultado en un cuadrículado como el siguiente:

		Ropa de Vestir					Ropa de Deportes				
		México				Espana	México				Espana
		Mérida	León	Puebla	Centro		Mérida	León	Puebla	Centro	
1999	1er Trim	80	80	75	110	325	105	125	125	147	510
	2do Trim	75	85	80	120	345	92	100	138	125	547
	3er Trim	80	74	85	115	310	198	105	145	123	512
	4to Trim	85	80	80	95	315	100	115	128	198	489
1998		215	270	265	382	1450	398	415	475	465	1541
1997		185	245	250	325	1358	356	395	448	432	1478
1996		150	200	190	296	1265	305	325	412	423	1365

En la actualidad existen herramientas que ayudan al usuario armar consultas multidimensionales en forma gráfica, es decir, él no necesita aprender a construir expresiones MDX, si no que a través de opciones se realiza la consulta en una forma clara, rápida y fácil; tal es el caso de Excel 2000 el cual es una herramienta de hoja de

cálculo que incluye una utilidad que ayuda al usuario a construir la consulta multidimensional de diversos motores de BD.

A continuación se describe como se construye una consulta multidimensional en Excel 2000 , accedendo a una BD de Access.

En la utilidad de obtención de datos externos existe una opción para amarrar cubos OLAP, esta funciona de la siguiente manera:

- 1.- Se indica el tipo de motor de BD a consultar, en este caso es Access.

Ilustración. II-14

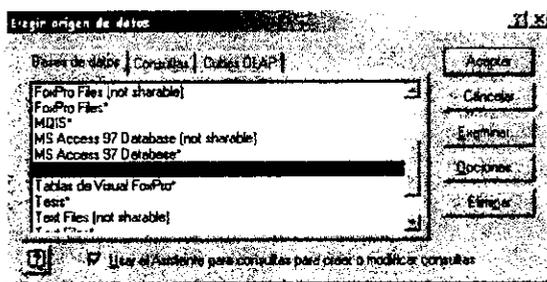


Ilustración II-14 Selecciona Tipo de Motor de BD

- 2.- Se da clic en el botón de "Aceptar" para indicar la BD. Ilustración II-15

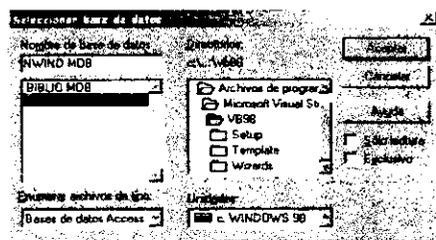


Ilustración II-15 Archivo de BD

3.- Se seleccionan los atributos que participarán como dimensiones y hechos en la consulta. Ilustración II-16

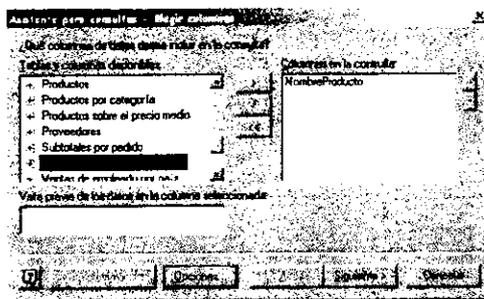


Ilustración II-16 Dimensiones y Hechos

4.- Se escoge la opción para crear un cubo para poder ver la información de forma multidimensional. Ilustración II-17

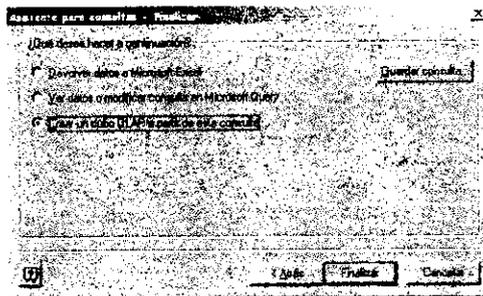


Ilustración II-17 Opciones de Consulta

5.- Se seleccionan los atributos que son de hechos, en este caso son las ventas.

Ilustración II-18

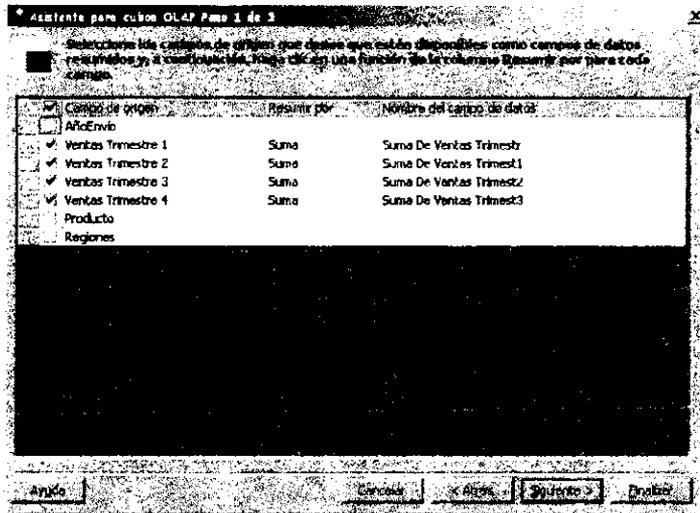


Ilustración II-18 Atributos de Hechos

6.- Se seleccionan los atributos que son dimensiones y sus jerarquías. Ilustración II-

19

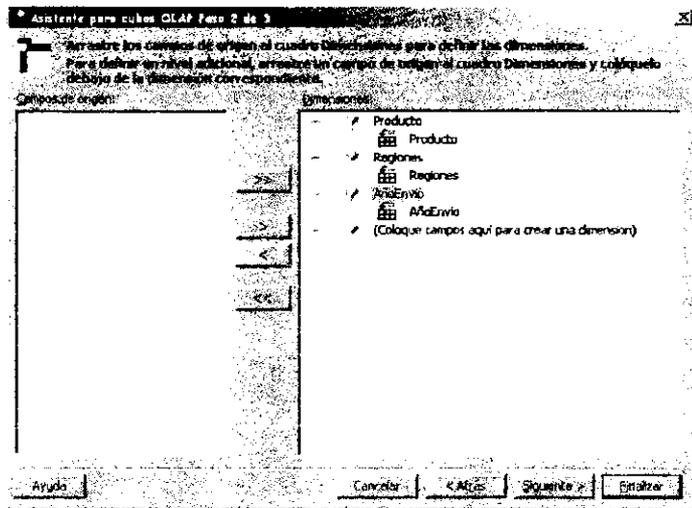


Ilustración II-19 Jerarquías y Dimensiones

7.- A través de esta utilería se indica como van a ir organizados los datos.

Ilustración. II-20

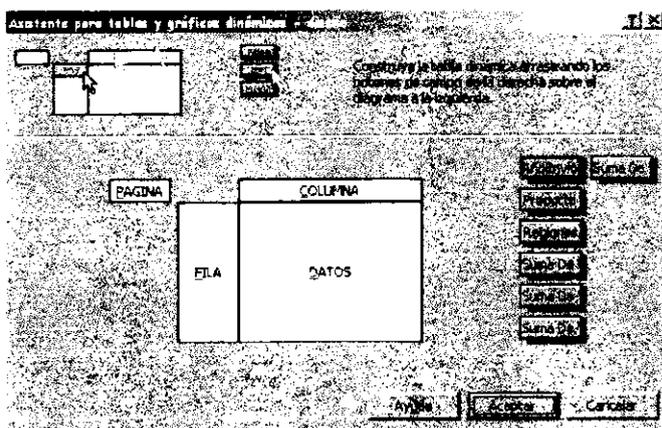


Ilustración II-20 Diseño de la Consulta

8.- Una vez diseñada la consulta los datos se presentan de las siguiente forma:

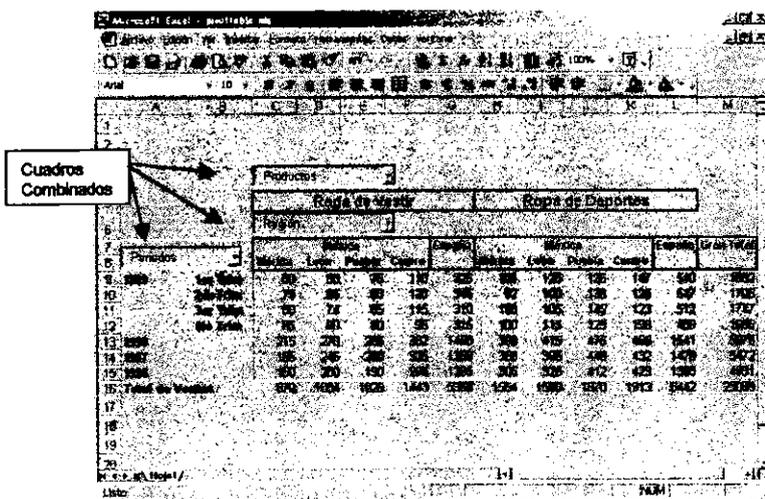


Ilustración II-21 Resultado de Consulta en Excel 2000

A este tipo de aplicaciones resulta dinámica ya que, en las dimensiones definidas existe un cuadro combinado para cada una de ellas, estos sirven para indicar que datos se requiere visualizar y cuales no; además permite ver todo el detalle de un nivel o todo el concentrado (drill-down), por ejemplo se requiere que no se visualice los trimestres del año de 1999, la opción que se desplegará en Excel es la siguiente:

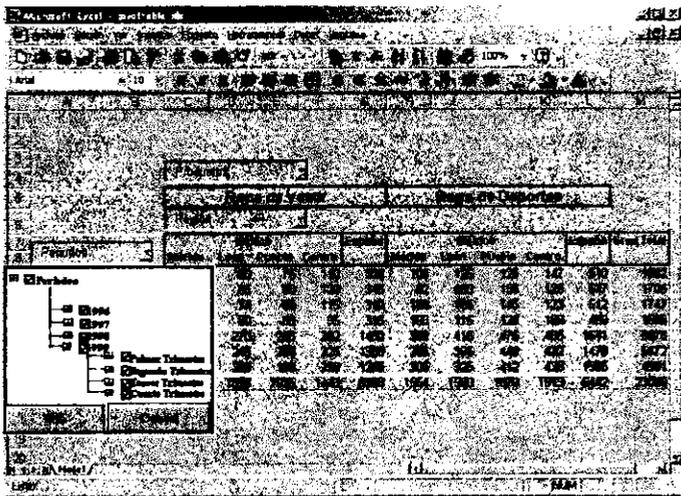


Ilustración II-22 Drill-Down en una Dimensión

Esta opción de la ilustración anterior permite desplegar los miembros de una dimensión y marcar o desmarcar aquellos que no se requiere ver en el momento del análisis.

En este tipo de aplicaciones existe la ventaja de dar el formato que requiera el usuario a la presentación además de manipular la información a su gusto Así como en Excel se pueden manejar libremente el formato de presentación y la información existen otras aplicaciones que funcionan semejante a este, tal es el caso de Holos o Crystal info

ambos de la compañía Seagate; entre otras; solo que estas últimas ayudan a crear consultas con condiciones sofisticadas: en Excel es un poco limitada la consulta ya que, por ejemplo, no se pueden poner condiciones como "Ventas mayores a 2000".

CAPITULO III. Metodología Para el Diseño de Sistemas en Bases de Datos Multidimensionales

En el diseño se deben tener en cuenta las técnicas de BD multidimensionales y los requerimientos de la empresa, para obtener un trabajo bien hecho que en un futuro no provoque más problemas que aciertos.

III.1. Elementos de análisis en las BD Multidimensionales

Para poder diseñar un BDM primero se debe comprender los siete componentes principales, estos son explicados a continuación:

1. **Dimensiones:** Es un lado particular del cubo el cual contiene una lista de miembros del mismo tipo en el que el usuario tiene la percepción de los datos. En la ilustración III-1 los números 1, 2, 3, 4 pertenecen a la dimensión de Localización.

Ofrecen una concisa e intuitiva manera de organizar y seleccionar datos para la obtención, exploración y análisis.

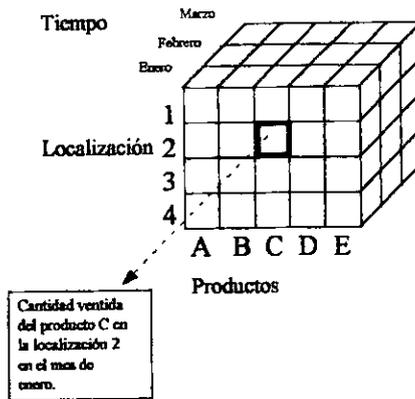


Ilustración III-1 Cubo de Ventas

2. **Atributos:** Estos proveen la descripción y jerarquías de las dimensiones, estas son analizadas en siguiente punto.

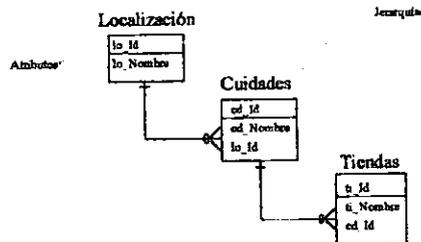


Ilustración III-2 Atributos y Jerarquías en las dimensiones

3. **Relaciones:** Las relaciones en una BDM son implícitas por la existencia de valores en la intersección de las dimensiones. Por ejemplo en la ilustración III-1, la relación que existe entre las ventas y localización es implícita por que las ventas de los productos obviamente se venden en cualquier localización.

4. **Jerarquías:** Cualquier miembro de la dimensión puede estar compuesto de relaciones padre-hijo; un miembro padre representa la consolidación de miembros que son sus hijos. Ilustración.(III-2)
5. **Esparcidad:** La esparcidad de los datos es calculada por el número de celdas que no contienen valores. Esta esparcidad es eliminada por el modelo de estrella simplemente no guardando aquellas combinaciones que no son válidas.

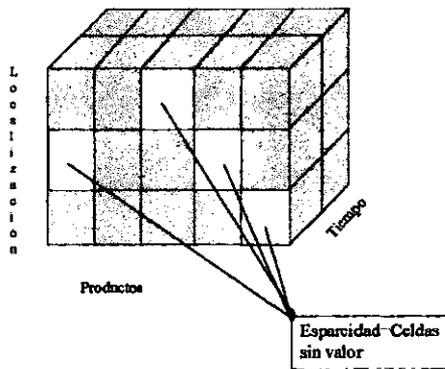


Ilustración III-3 Esparcidad de Datos

6. **Servicios:** Básicamente la utilidad que se le da los datos multidimensionales es para la fácil navegación en la información, cálculos, detalle a un nivel determinado de jerarquía para la obtención, exploración y análisis.
7. **Valores de las Celdas:** Los valores que manejan las BDM en sus celdas son numéricos tales como ventas, costos, precios. Estos valores en las celdas son conocidas en las BDM como variables, las cuales representan una actividad del negocio.
- Estas variables a veces pueden ser dimensionados, por ejemplo las ventas pueden ser dimensionadas por región, producto ó cliente; sin embargo los

precios de los productos no pueden ser dimensionados ya que son muy diversos y por lo tanto produciría celdas vacías (alta dispersidad).

Este concepto es llamado como Variables Dimensionadas Independientemente, la cual nos sirve para mejorar el rendimiento de la BDM y reducir la esparcidad de los datos

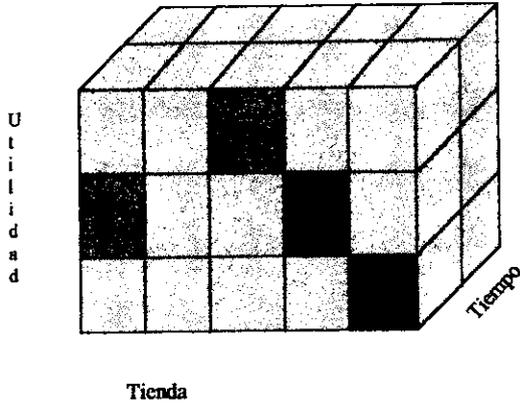
Las variables pueden ser definidas como complejas relaciones matemáticas con otras variables. En el ejemplo de la ilustración II-2 la relación entre los miembros de la dimensión es de una simple operación de adición.

Las relaciones complejas incluyen operaciones aritméticas, promedios ó ecuaciones.

III.2. Pasos para el diseño de una BDM

Primero identificar los requerimientos para hacer uso del sistema OLAP. Esto es, empezando por el modelo de BD actual de la organización. Obtener el juego de componentes clave o llave para establecer las dimensiones de los cubos y las variables. Por último conocer el entorno y las reglas del negocio.

La aplicación que va soportar la BD debe permitir a los usuarios hacer reportes, análisis de datos, identificar problemas y oportunidades del negocio e iluminar áreas clave.



-  Meses productivos
-  Utilidades normales
-  Tiendas que se necesitarán préstamos

Ilustración III-4 Identificación de problemas

De los requerimientos del negocio se identifican las salidas del nuevo sistema, estas pueden ser las ventas, costos, presupuestos, etc. En estas salidas se identifican los cálculos que necesitan los usuarios para el análisis de la información.

Los reportes existentes que manejan la empresa pueden ser un buen comienzo. Aunque estos reportes puedan tener algunas ineficiencias, la nueva aplicación debe de proveer una mejora a estos.

REPORTE DE VENTAS DEL SEMESTRE

Division Norte							
Productos	Ene	Feb	Mar	Abr	May	Jun.	Total
CD's	\$ XXXX						
Pc's	\$ XXXX						
Impresoras	\$ XXXX						
Division Sur							
CD's	\$ XXXX						
Pc's	\$ XXXX						

Ilustración III-5 Reporte Dimensional

El tiempo es importante en el análisis y por lo tanto se debe incluir en el diseño. Se debe tener cuidado de no producir información redundante, aunque el reporte tenga información histórica no quiere decir que ésta sea importante para el análisis.

Una vez identificados los componentes claves de la BD multidimensional se procede a construir los esquemas de tipo estrella ó Snowflake teniendo en cuenta las siguientes técnicas y consideraciones físicas.

1. Técnica de Denormalización.
2. Consultas (queries) usando múltiples tablas de hecho.
3. Volúmenes de datos en las tablas dimensionales.
4. Tablas consolidadas.
5. Número de niveles en las jerarquías (snowflakiness).
6. Datos compartidos.
7. Tipo de cubos.
8. Tipo de motor de base de datos.

Técnica de Denormalización: La denormalización en una BD se refiere a la duplicación de datos en una entidad. El diseño de datos multidimensionales requiere cierta denormalización en las entidades para un rápido acceso a la información. Aunque con esta técnica bien sabemos que se sacrificará espacio por tiempo; se realizarán más procesos cuando los datos duplicados cambien. La rapidez es más notoria cuando se manejan grandes cantidades de información.

Esta denormalización es utilizada en los sistemas OLTP, cuando los datos pueden ser planeados o predecibles.

La denormalización no ofrece una flexibilidad cuando surgen requerimientos inesperados. Existe mucha dependencia entre los datos y en cada entidad debe existir toda la información necesaria para la relación. Un cambio en la jerarquía no puede ser implementada en una sola parte del modelo, este afecta toda la jerarquía de la entidad.

Cuando los datos duplicados cambian, existen cientos o miles de registros que deben ser corregidos, por otro lado en un modelo totalmente normalizado se actualiza un registro en una sola entidad.

Las entidades que nos interesan por lo general denormalizar son las de hecho y aquellas que tienen varias jerarquías como se verá en los siguientes puntos. Por ejemplo veamos las entidades VENTAS Y DETALLE en una BD relacional (totalmente normalizada) Ilustración III-6.

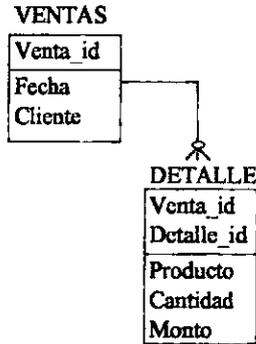


Ilustración III-6 Entidades normalizadas (Ventas)

La denormalización que nos interesa es de la siguiente forma (ilustración III-7):

VENTAS	
Venta_id	
Fecha	
Cliente	
Producto	
Cantidad	
Monto	

Ilustración III-7 Entidad denormalizada

La ventaja de la denormalización de la entidad de VENTAS es que se elimina una operación de join y por lo tanto la consulta a VENTAS se ejecutará en menos tiempo.

Consultas (Querías) usando múltiples tablas de hecho: En la misma forma en crece el modelo multidimensional existe una tendencia en incluir más tablas de hecho, por lo tanto en un query se querrá utilizar múltiples tablas de hechos.

Por ejemplo tal vez se quiera comparar las ventas en cierto periodo de promoción y compararlos con las inversiones hechas con estas ventas. Este query requiere de que el modelo incluya dos o más snowflakes.

El problema con esta técnica es que va en contra de una de las características principales para el rendimiento: "Principalmente las tablas de hechos contienen grandes volúmenes de información, mientras que las tablas de las dimensiones son pequeñas; esto nos da una ventaja en el rendimiento al relacionar tablas grandes con tablas pequeñas.", es decir al relacionar entidades que contienen grandes volúmenes de información con entidades similares en tamaño disminuye el rendimiento en la rapidez de las consultas.

Volúmenes de datos en las tablas dimensionales: Esta consideración nos sirve para recordar que la velocidad de nuestro modelo multidimensional depende de que solo las tablas de hecho deben tener grandes volúmenes de información y las tablas de dimensiones deben tener pocos datos.

Tablas consolidadas: Consolidados son datos precalculados que resuelven Queries ya conocidos. Los Queries no planeados no son soportados por este tipo de tablas.

También no es bueno tener todas las posibles combinaciones de consolidados por ejemplo si tenemos 5 dimensiones existen $7! = (5040)$ posibles tablas de consolidados.

Entre más tablas de consolidados tengamos más tiempo nos llevará mantenerlas actualizadas y por supuesto utilizan demasiado espacio.

Para definir consolidados se debe considerar la regla de 80:20, 80% del rendimiento implementado será utilizado por el 20% de los consolidados más utilizados.

El 80% restante de los consolidados principales deben caer en las siguientes categorías:

- Aquellas que toman en cuenta la optimización de los recursos computacionales para obtener los Queries. Parte de estos Queries se realizarán en las horas no pico.
- Aquellas tablas que mantienen consolidados que pocas veces se usan.

Número de niveles en las Jerarquías (snowflakiness): El tener más jerarquías en las dimensiones nos implica más operaciones de join , lo cual nos perjudica el rendimiento de la obtención de los datos.

Para evitar los joins con las tablas podemos utilizar la técnica de la denormalización, por ejemplo podemos tener la siguiente jerarquía.

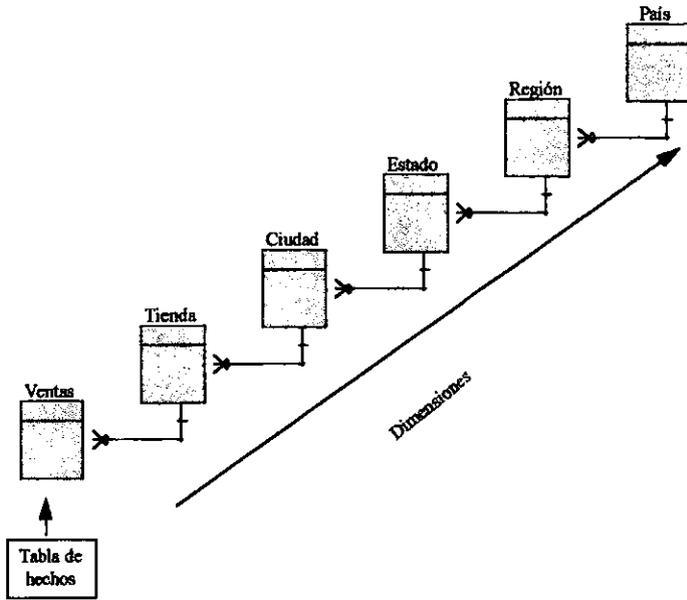


Ilustración III-8 Entidades Normalizadas

Lo recomendable para tener un rendimiento en la obtención de los datos es denormalizar las jerarquías, como por ejemplo:

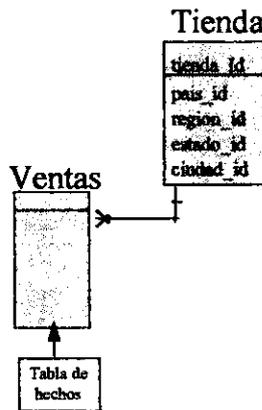


Ilustración III-9 Entidades Denormalizadas

Datos compartidos: Otra consideración es la de recordar entre más compartidos estén los datos más baja es nuestra velocidad en la obtención de la información. Se recomienda que cada esquema este una sola BD. Las tablas de hechos y las tablas de dimensiones estarán duplicadas en cada BD guardando su sincronía entre ellas mismas.

Existe un fenómeno que es llamado como "Explosión de la BD" y el cual es muy común tener los siguientes errores:

- Muy poca supresión de la esparcidad de los datos
- Almacenamiento de la BD multidimensional. Este puede causar problema por la generación de valores matemáticos muy grandes.

Tipos de cubos: Existen dos formas de determinar como los datos pueden ser procesados, estos son los Hiper cubos y los Multicubos.

Hiper cubos: Los objetos con tres o más dimensiones se describen como lados planos y cada dimensión con ángulos rectos en relación con los demás; se presenta como una simple estructura cúbica lógica singular. Utilizan métodos de compresión de datos dispersos.

Se permite usualmente la introducción de valores para cada combinación de miembros de dimensión y todas las partes de espacio de datos tienen idéntica dimensionalidad.

Los productos que utilizan este tipo de cubos es por ejemplo ESSBASE, CFO Vision, PowerPlay, entre otros.

Multicubos: En los productos de multicubos, la aplicación de diseño segmenta la base de datos en varias estructuras multidimensionales, cada una estas es compuesta por una subdivisión de varias dimensiones llamadas subcubos.

Se tienen dos tipos de multicubo:

El bloque multicubo el cual no utiliza dimensiones especiales en el nivel de datos. Un cubo puede consistir en cualquier número de dimensiones definidas y todas las dimensiones, incluyendo el tiempo, son tratados como dimensiones ordinarias.

Series Multicubo: Trata cada tabla de hechos como un cubo separado (a menudo en series de tiempo) con sus propias dimensiones.

Los bloques multicubo son más flexibles, por que no asumen ninguna dimensión y permiten múltiples para ser manejadas juntas, pero no son muy convenientes para los reportes que quieren ver un bloque a la vez.

Las series multicubos se pueden ver un cubo a la vez, además de que los cubos pueden ser unidos en uno solo aunque físicamente estén guardados en diferentes cubos.

Los proveedores que ofrecen esta tecnología son Hoios, Oracle Express, Microsoft OLAP services, Gentia, Informix, entre otros. ⁵

Tipos de motor de BD: Dada la necesidad de guardar los datos eficientemente denormalizados, existen tres opciones de hacerlo:

- **BD relacional:** Este tipo de BD son manejadas por los sistemas RDBMS
- **BD multidimensional:** Aquí los datos son guardados en un servidor aparte. Puede incluir datos extraídos y sumariados de sistemas legales, BD relacionales ó usuarios finales. La mayoría de los casos son guardados en disco algunos otros se manejan en la memoria RAM. En pocos casos se permite múltiples usuarios con concurrencia para leer y escribir, pero como hemos visto no es bueno. Algunos otros productos permiten que una sola persona actualice y que muchas puedan acceder, y otras solo permiten un acceso de solo lectura.
- **Archivos en máquinas clientes:** En este caso se tienen una parte de los datos y se mantienen en una maquina cliente. Pueden ser distribuidos o creados en demanda, o por medio de el WEB. También se pueden manejar en disco o en RAM y el acceso por lo general es de solo lectura.

De la misma forma que existen tres formas de guardar los datos, también existen tres formas de accederlos y procesarlos. Los cálculos para los datos multidimensionales no necesitan procesarse en el mismo lugar donde están guardados los datos.

⁵ Pendse Niegel, The OLAP Report, www.olapreport.com , USA 1999

- **SQL:** Es una opción muy inusual para hacer procesos complejos multidimensionales, aún estando en la plataforma de BD relacional. No se tiene la posibilidad de hacer estos cálculos en una simple sentencia.
- **Motor multidimensional:** En comparación con el SQL, aquí si se pueden hacer procesos y cálculos multidimensionales en una misma sentencia además de que esta hecho para mejorar el rendimiento en la obtención de los datos. También funciona muy bien en las aplicaciones basadas en un estructura cliente-servidor
- **Motor multidimensional en máquina cliente:** Hay veces que los usuarios tienen para ellos mismos máquinas poderosas y los vendedores aprovechan estas características para vender productos que manejan datos multidimensionales y que son procesados en la misma PC. Pero con la popularidad del WEB muchos vendedores están moviendo sus productos basados en esta arquitectura a aplicaciones en servidores WEB.

III.3. Software de aplicación sobre BDM.

A continuación se enumeran algunos de los principales productos encontrados, así como la plataforma sobre la cual pueden operar: ⁶

OPCIONES DE PROCESAMIENTO DE DATOS MULTIDIMENSIONALES	OPCIONES DE ALMACENAJE DE DATOS MULTIDIMENSIONALES		
	RDBMS	SERVIDOR DE BD MULTIDIMENSIONAL	ARCHIVOS EN MAQUINAS CLIENTE
SQL	MicroStrategy DSS suite		
SERVIDOR MULTIDIMENSIONAL	<ul style="list-style-type: none"> -IBM BD2 OLAP SERVER -IA DecisionSuite -Longview Khalix -Informix Metacube -Speed Ware Meda/MR -Microsoft OLAP Services -Oracle Express (modo ROLAP) -Pilot Analysis Server -Sagent -Seagate Holos (modo ROLAP) -Applix TM1 -WhiteLight 	<ul style="list-style-type: none"> -SAS CFO Vision -Comshare Decision -Hyperion Essbase -Oracle Express -Seagate Holos -Gentia -Speedware Media/M -Microsoft OLAP services -Pilot Analysis SERVER -Applix TM1 	
Motor multidimensional en cliente	<ul style="list-style-type: none"> -Oracle Discover -Informix Metacube -PowerPlay(opción del servidor) 	<ul style="list-style-type: none"> -Comshare FDC -Dimensional Insight -Hyperion Enterprise -Hyperion Pillar 	<ul style="list-style-type: none"> -Brio -BusinessObjects -Cognos PowerPlay -Personal Express -TM1 Perspectives

En la mayoría de estos productos se puede escoger la manera de presentar la multidimensionalidad al usuario, esta no es solo una opción de vista; también determinan como los datos son procesados y en particular cuantos cálculos son permitidos.

Cada una de estas opciones tienen sus ventajas y sus desventajas, no es fácil decidir una opción óptima. La decisión de escoger alguna opción de almacenaje o de tipo de proceso depende de las necesidades y capacidades de la empresa.

⁶ OLAP, www.busintel.com/syn5.htm

Características Principales de algunos de los productos que manejan

BDM

Productos	Principales Características
Commander OLAP de Comshare	Usa el motor Essbase más otros componentes de Comshare. Es promocionado por la capacidad de detectar excepciones en grandes BD "Detecta y Avisar"
Essbase de Arbor Software	Sofisticado producto de cliente/servidor para grandes y medianas aplicaciones. Utiliza mucho espacio en disco
Gentium de Planning Sciences	Son herramientas orientadas a objetos, tienen una amplia gama de funciones que no se habían usado hasta ahora, pero la mayoría de siguen madurando
Holos de Holistic Systems	Esta basado en la arquitectura de cliente/servidor. Es muy útil para la introducción de la información a través de los usuarios
Hyperion Enterprise de Hyperion Software	Especializado en reportes financieros
LightShip Suite de Pilot software	Orientado a reportes en series de tiempo y herramientas de análisis dirigido a aplicaciones de ventas y mercadeo.
Power Play de Cognos	Es una popular herramienta de OLAP. Trabaja aislado o en su propio servidor ESSBASE. Diseñado para fácil implementación.
Express de Oracle	Es el más vendido en la línea de productos de OLAP, incluye una amplia variedad de funciones y aplicaciones prescritas
Holos de Seagate	Trabaja en una arquitectura abierta cliente/servidor. Puede integrar diferentes BD dispersas en una sola aplicación. Tiene las siguientes capacidades de análisis: <ol style="list-style-type: none"> 1. Pronósticos: Ejecuta afirmaciones y simula escenarios de negocio, 2. Análisis: Proporciona análisis de riesgo y objetivos, estadísticas. 3. Datamining: Usa un avanzado datamining con redes neuronales y reconocimiento de patrones.

CAPITULO IV. Caso Práctico

IV.1. Descripción del negocio

Se trata de un sistema de contratación de medios publicitarios basado en la descripción del negocio de la tesis: *Modelo del Sistema de Contratación de Medios Para Arquitectura Cliente – Servidor. Visual Basic y SQL Server en una Empresa De Publicidad, México 1997, 48 p. Tesis (lic. Informática), UNAM, FES-C*, dicho negocio consiste en una empresa dedicada a contratar espacios en los diferentes medios de comunicación (radio, televisión, prensa) para que diversos clientes hagan publicidad de los productos que ofrecen en el mercado; en el transcurso de este tema se hará referencia a esta empresa como "Intermediaria de Medios". El proceso que se lleva es el siguiente:

1. Se tiene un plan general con una cierta cantidad de dinero para la publicidad de un determinado producto, básicamente este plan tiene una vigencia de un año.
2. Durante este periodo se emiten varios presupuestos mensuales, cada uno de ellos contempla un espacio publicitario y una o varias transmisiones o emisiones, estos están sujetos a la cantidad de dinero especificada en el plan general.
3. Una vez que se tiene un presupuesto se emiten una o varias órdenes las cuales son enviadas al medio correspondiente para su ejecución; este proceso tiene una duración promedio de una semana.
4. Cuando la publicidad se ha llevado a cabo, el medio emite la factura correspondiente al servicio. La "Intermediaria de Medios" monitorea las solicitudes de publicidad y una vez que corrobora que se llevaron a cabo de acuerdo a lo establecido en la orden, acepta la factura y a partir de ese

momento cuenta con cuarenta y cinco días para pagar al medio y a su vez expide al cliente la factura que contiene el monto a pagar al medio así como la comisión por su servicio. El cliente tiene un plazo de treinta días para cubrir la factura.

IV.2. Requerimientos del sistema

Dadas las condiciones de pago señaladas en el punto anterior, para la empresa es muy importante estar monitoreando el flujo de facturación vs montos presupuestados, puesto que no debe incurrir en el atraso de pago a los medios ya que se arriesgaría a pagar intereses moratorios.

Actualmente el monitoreo se lleva a cabo a través de reportes que despliegan los montos presupuestados y facturados bajo dos tipos de clasificaciones básicamente: por medio o por cliente. Ambas en un período determinado.

Los reportes antes mencionados son de gran utilidad, sin embargo se considera que hace falta una herramienta que facilite al usuario la identificación oportuna de pagos que estén a punto de vencer y que tengan algún problema en cuanto al cobro al cliente de tal forma que sean candidatos para el pago de intereses.

La BD y los reportes existentes, servirán como inicio para diseñar la BD multidimensional.

Reportes Actuales: Los siguientes reportes servirán como base para la identificación de los componentes clave en la construcción del modelo multidimensional.

INTERMEDIARIA DE MEDIOS

Resúmen anual de Presupuestado Vs. Actual

	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
TV												
Presupuestado												
Facturado												
Pendientes												
Vencidos												
RADIO												
Presupuestado												
Facturado												
Pendientes												
Vencidos												
TV												
Presupuestado												
Facturado												
Pendientes												
Vencidos												
RESUMEN												
Presupuestado												
Facturado												
Pendientes												
Vencidos												

Ilustración IV-1 Resumen Anual de Presupuestado Vs. Actual

INTERMEDIARIA DE MEDIOS

Detalle anual de Presupuestado Vs. Actual

CLIENTES	ENERO			
	Presupuestado	Facturado	Pendiente	Vencidas

Ilustración IV-2 Reporte detallado anual Presupuestado Vs. Actual

INTERMEDIARIA DE MEDIOS

Detalle anual Pago Pendientes (Representantes)

REPRESENTANTES DE MEDIOS PUB.	Presupuestado	ENERO Facturado	Pendiente
----------------------------------	---------------	--------------------	-----------

Ilustración IV-3 Detalle de Pagos Pendientes

A simple vista se puede determinar que existen en estos reportes los siguientes componentes clave:

1. Clientes
2. Presupuestados
3. Facturados
4. Vencidos
5. Pendientes
6. Medio publicitario
7. Representantes de los Medios Publicitarios
8. Tiempo

Siendo el esquema Entidad-Relación el que se muestra en la ilustración IV-4.

BD Actual en la empresa: Esta BD servirá para escoger las entidades que sirvan para la construcción del modelo multidimensional, además de indicar que entidades se le aplicará la técnica de la denormalización.

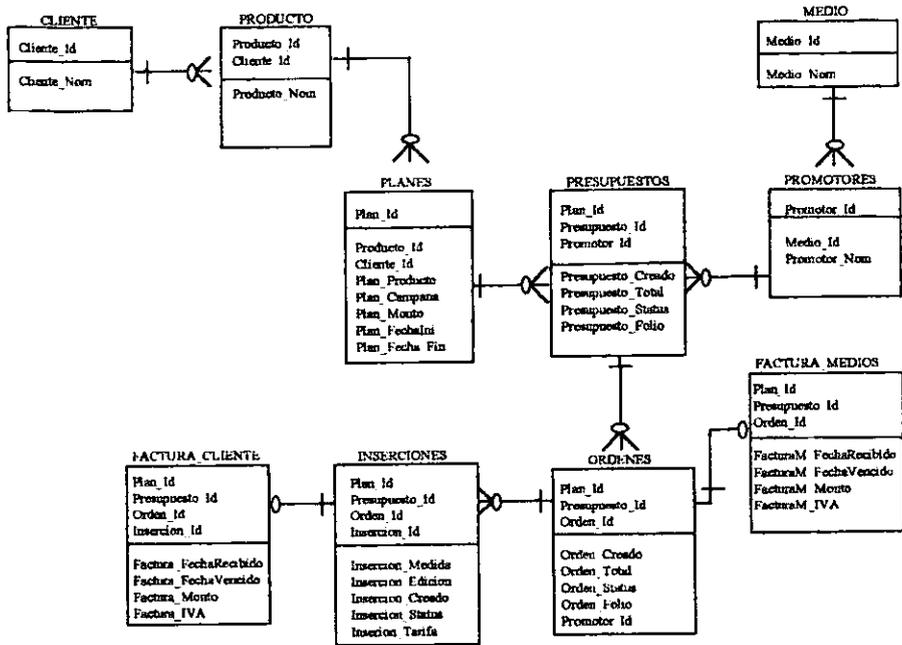


Ilustración IV-4 Diagrama Entidad-Relación actual de "Intermediaria de Medios"

Descripción de las Entidades y Atributos de la BD actual

MEDIO: Es un catálogo de los medios publicitarios en la actualidad, por ejemplo radio, televisión o prensa

ATRIBUTO	DESCRIPCION
Medio_Id	Identificador del medio
Medio_Nom	Nombre del medio

PRODUCTO: Catálogo de productos de un cliente en especial.

ATRIBUTO	DESCRIPCION
Producto Id	Identificador del producto
Cliente Id	Nombre del cliente propietario del producto
Producto_Nom	Nombre del producto

CLIENTE: Catálogo de clientes

ATRIBUTO	DESCRIPCION
Cliente Id	Identificador del cliente
Cliente_Nom	Nombre del cliente

PROMOTORES: Catálogo de promotores

ATRIBUTO	DESCRIPCION
Promotor Id	Identificador del Promotor
Medio Id	Identificador del medio
Promotor_Nom	Nombre del promotor

PRESUPUESTOS: Son los presupuestos de cada mes derivados de la Entidad de PLANES.

ATRIBUTO	DESCRIPCION
Plan Id	Identificador del Plan
Presupuesto Id	Identificador del presupuesto
Promotor Id	Identificador del promotor
Presupuesto_Creado	Fecha de creación del presupuesto
Presupuesto_Total	Monto del presupuesto
Presupuesto_Status	Status del presupuesto, puede estar pendiente o pagado
Presupuesto_Folio	Folio asignado al presupuesto

PLANES: Son los planes que se emiten cada año.

ATRIBUTO	DESCRIPCION
Plan Id	Identificador del Plan
Producto Id	Identificador del producto
Cliente Id	Identificador del cliente
Plan_Producto	Nombre del plan para el producto a publicar
Plan_Monto	Monto total, presupuestado para todo el año
Plan_Campaña	Nombre de la campaña del producto
Plan_Fecha_Ini	Fecha de Inicio de vigencia del plan
Plan_Fecha_Fin	Fecha de vencimiento del plan

ORDENES: Son las ordenes emitidas a un determinado medio para promocionar un producto.

ATRIBUTO	DESCRIPCION
Plan_Id	Identificador del Plan
Presupuesto_Id	Identificador del presupuesto
Orden_Id	Identificador de la orden
Orden_Creado	Fecha de creada la orden
Orden_Total	Monto de la orden
Orden_Status	Status de la orden, puede estar pendiente ó cubierta
Orden_Folio	Número de folio asignado a la orden.
Promotor_Id	Identificador del promotor de la orden

INSERCIONES: Es la descripción de cada publicidad hecha en una orden.

ATRIBUTO	DESCRIPCION
Plan_Id	Identificador del Plan
Presupuesto_Id	Identificador del presupuesto
Orden_Id	Identificador de la orden
Insercion_Id	Identificador de la inserción
Insercion_Medida	Duración de la publicidad
Insercion_Edición	Fecha de edición
Insercion_Creado	Fecha de creación de la inserción
Insercion_Status	Puede estar pendiente o pagada
Insercion_Tarifa	Monto de la Tarifa que aplica a la inserción.

FACTURA CLIENTE: Es la descripción de cada factura emitida a los clientes.

ATRIBUTO	DESCRIPCION
Plan_Id	Identificador del Plan
Presupuesto_Id	Identificador del Presupuesto
Orden_Id	Identificador de la Orden
Insercion_Id	Identificador de la Inserción
Factura_FechaRecibido	Fecha cuando recibió la factura el cliente
Factura_FechaVencido	Fecha de cuando vence la factura
Factura_Monto	Monto de la factura
Factura_Iva	Iva de la factura

FACTURA MEDIO: Es la descripción de cada factura emitida por los medios publicitarios a "Intermediaria de Medios".

ATRIBUTO	DESCRIPCION
Plan_Id	Identificador del Plan
Presupuesto_Id	Identificador del Presupuesto
Orden_Id	Identificador de la Orden
FacturaM_FechaRecibido	Fecha cuando se recibió la factura del medio
FacturaM_FechaVencido	Fecha de cuando vence la factura
FacturaM_Monto	Monto de la factura
FacturaM_Iva	Iva de la factura

IV.3. Esquema de Tipo Snowflake

Tablas de Hechos y Dimensiones: De los componentes clave determinados anteriormente, se establecen las dimensiones y las tablas de hechos (Ilustración IV-5). Las relaciones y atributos ya están explícitos en el diagrama Entidad-Relación (Ilustración IV-4); para este ejemplo los hechos y dimensiones detectados son:

TABLAS DE HECHOS	DIMENSIONES
Inserciones Factura_Cliente Factura_Medios Ordenes	Representantes Clientes Productos Medios Tiempo Planes Presupuestos

Ilustración IV-5 Identificación de Tabla de Hechos y Dimensiones

Establecimiento de Jerarquías : Las jerarquías se determinan por medio de las relaciones que existen entre las tablas de dimensiones (Ilustración IV-6).

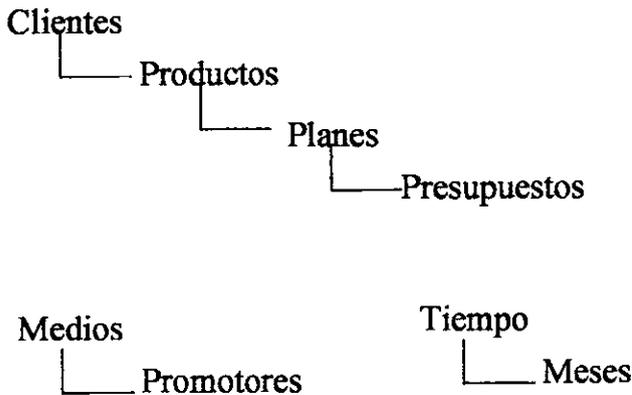


Ilustración IV-6 Jerarquías en las Dimensiones

Las entidades de Factura_Cliente, Factura_Medios no se jerarquizan ya que son le nivel más bajo de información detallada.

Una vez identificados los componentes del diagrama Multidimensional, se procederá a escoger el tipo de esquema (estrella ó snowflake) a implementar, en este caso se escoge un esquema de tipo snowflake ya que existen jerarquías en las dimensiones

Esquema de Tipo Snowflake: Las tablas que se denormalizarán son las de hechos(Inserciones, Factura_Cliente ,Factura_Medios, Ordenes), quedando el esquema de la forma siguiente: (Ilustración IV-7).

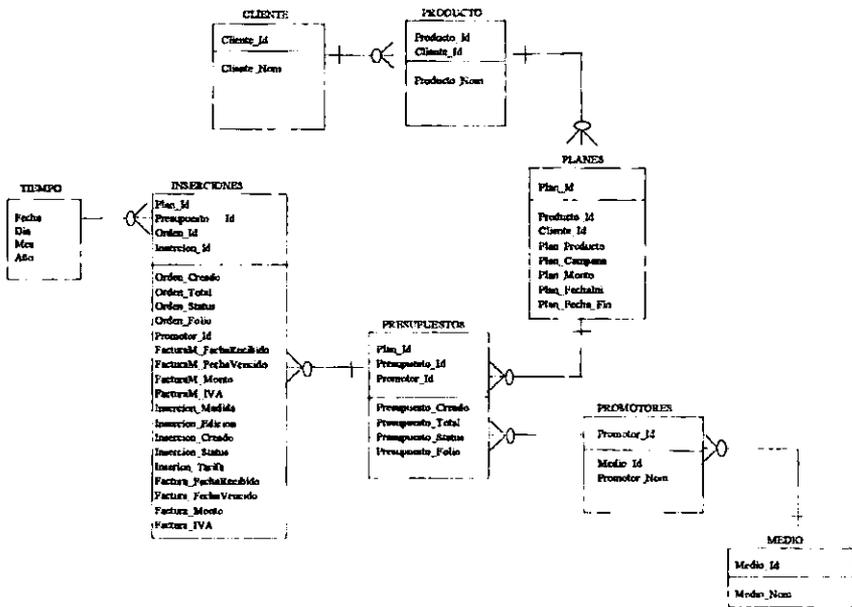


Ilustración IV-7 Diagrama Snowflake

Creación de Reportes en Excel

Los datos multidimensionales pueden ser explotados fácilmente por aplicaciones como Excel 2000, ya que a través de este el usuario define sus consultas a las necesidades del día, es decir él las define en el momento justo cuando se requiera, además de que este tipo de aplicaciones están optimizadas para manipular este tipo de información, por otro lado en datos relacionales estas consultas por lo general se hacen por un experto en desarrollo de software el cual conoce y maneja las diversas sentencias de SQL para la explotación de la información, en este caso el usuario tiene que esperar a que el experto acabe su reporte, por lo general este tipo de reportes no son dinámicos, es decir siempre muestran la información de la misma forma. Es por esto último que los reportes que se necesitan en la empresa, además de otros, pueden ser creados con facilidad por el usuario.

En base al esquema de la Ilustración IV-7, se muestra a continuación como construir un reporte de Detalle Anual de Pagos Pendientes, utilizando la utilería para construir consultas OLAP del programa Excel 2000 de la siguiente forma:

1.- Se selecciona el tipo de Motor de BD, en este caso es Access. Ilustración. IV-8

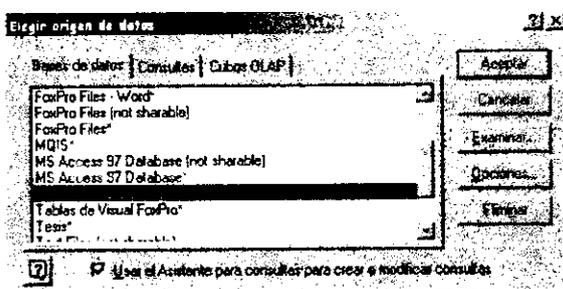


Ilustración IV-8 Tipo de Motor de BD

2.- Se indica la ubicación de la BD. Ilustración IV-9

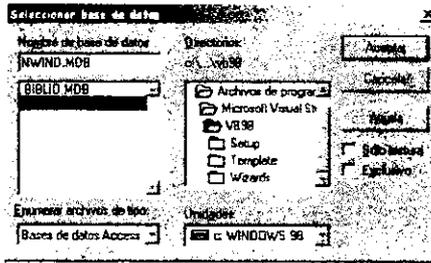


Ilustración IV-9 Ubicación de la BD

3.- Se indican las dimensiones y atributos de hecho que participan en la consulta.

Ilustración IV-10

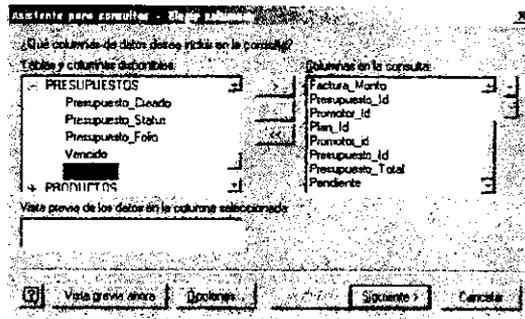


Ilustración IV-10 Atributos para la Consulta

4.- Se escoge la opción para crear un cubo para poder ver la información de forma multidimensional. Ilustración IV-11

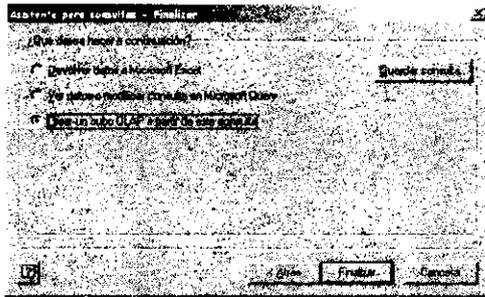


Ilustración IV-11 Opciones de Consulta

5.- Se indican los atributos de hecho. Ilustración IV-12, que en este caso es el monto de lo facturado, monto de lo presupuestado y el monto pendiente por facturar

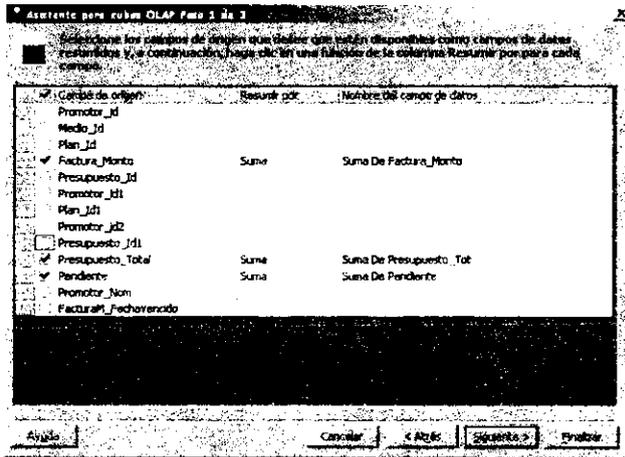


Ilustración IV-12 Atributos de Hecho

6.- Se indican las dimensiones y sus jerarquías, en este caso las dimensiones son el representante y el tiempo. Ilustración IV-13

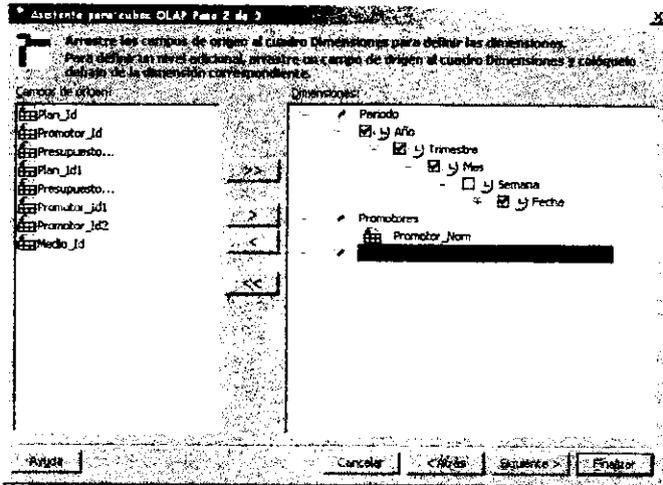


Ilustración IV-13 Dimensiones y Jerarquías

7.- Se diseña la consulta precisando que atributos van en las columnas, cuales en los renglones y el orden de los hechos. Ilustración IV-14

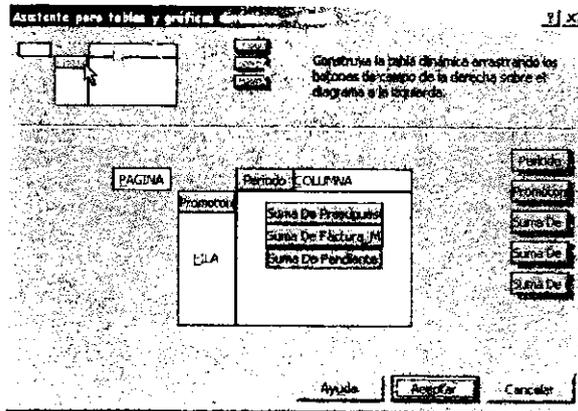


Ilustración IV-14 Diseño de la Consulta

Una vez hecho los puntos anteriores se presenta en forma automática la consulta en Excel. Ilustración IV-15

Promotor	Años	Trimestre	Total 1999	Total general
Espectacular Coacalco	12000	12000	24000	30000
Espectacular la Voz	12000	0	12000	18000
Industrial	2000	9000	11000	24000
Mundo del Radio	2000	0	2000	2000
Prensa la Verdad	1000	3000	4000	4000
Radio Estes	2000	0	2000	2000
Televisión de México	6000	28000	34000	58000
Televisión Europa	10000	18000	28000	61000
Total Suma De Presupuesto Tot	15000	41000	56000	125000
Total Suma De Factura Monto	12000	46500	58500	178000
Total Suma De Pendiente	0	0	0	53000

Ilustración IV-15 Reporte de Pagos Pendientes

Con el esquema de snow flake y este tipo de aplicaciones para consulta OLAP se pretende resolver que el usuario de los datos multidimensionales no este limitado al monitoreo del flujo de la facturación, si no que al implementar una aplicación que permita consultas dinámicas en base a datos multidimensionales para que él pueda hacer estudios profundos por ejemplo la probabilidad de que un cliente sea moroso o no, u otro análisis que el requiera en base a la información que se almacenará en el esquema presentado anteriormente. Se tiene la ventaja de mayor rapidez de respuesta en las consultas ya que, como se mencionó anteriormente, este tipo de BD no se ejecutan transacciones y son pocos los usuarios que accesan esta información, la mayoría de

veces solo el gerente o director del área lo hará, ya que ellos son los que toman las decisiones importantes.

CONCLUSIONES

La tendencia de la tecnología esta encaminada ya no solo a facilitar la vida de las personas si no aproximarlo a tomar mejores decisiones.

Dentro del mundo de las BD, se esta tratando de que los usuarios obtengan su información en forma dinámica, rápida y tal y como ellos piensan; con esto se está tratando de solucionar el problema de gasto innecesario de tiempo que se utiliza en interactuar y analizar la información, para tratar de encontrar información pura, es decir, información inmediata y con la semántica necesaria para la toma de decisiones.

Como se manejan ahora los reportes de la gente que toma las decisiones en la empresa, tienen un diseño de cierta forma fijo, por lo que cada cambio a ellos se necesita una persona experta para satisfacer dicho requerimiento; los sistemas OLAP permiten al usuario acomodar su información tal y como el la necesita al momento, para que el pueda tener una más amplia de cómo va caminando el negocio, además de que ellos pueden construir sus consultas dinámicamente, eliminando con esto la necesidad de requerir expertos para modificar la presentación de sus reportes.

A lo largo de la tesis se puede observar que las Bases de Datos Relacionales son excelentes para mantener la integridad de la información de la empresa, solo que a la hora de extraer información concreta para el análisis, resulta pesado y tardado, ya que no están optimizadas para ello. Por otro lado, en las Bases de Datos Multidimensionales se elimina la parte transaccional, con lo cual se optimiza el funcionamiento de las

consultas de tal manera que la información se obtiene rápida y dinámicamente en diferentes niveles de consolidación.

Por lo descrito anteriormente podemos concluir que el modelo de Base de Datos Multidimensionales es una tecnología poderosa que facilita la construcción de sistemas encaminados a la obtención inteligente de información y aunque es de nueva creación y su estructura resulta un tanto compleja, la infraestructura de cómputo con la que contamos actualmente nos hace pensar que no estamos lejanos de ser partícipes de ella.

INDICE DE ILUSTRACIONES

ILUSTRACIÓN I-1 REPRESENTACIÓN ESQUEMÁTICA DE UN SISTEMA DE INFORMACIÓN	2
ILUSTRACIÓN I-2 SMDB	3
ILUSTRACIÓN I-3 BASE DE DATOS DE TIPO RED	6
ILUSTRACIÓN I-4 BASE DE DATOS TIPO JERÁRQUICA	7
ILUSTRACIÓN I-5 MODELO RELACIONAL Y SUS ELEMENTOS	8
ILUSTRACIÓN I-6 ENTIDAD DE SINIESTROS NO NORMALIZADA	16
ILUSTRACIÓN I-7 EJEMPLO DE ELIMINACIÓN DE TUPLAS REPETIDAS	16
ILUSTRACIÓN I-8 ENTIDADES EN SEGUNDA FORMA NORMAL	18
ILUSTRACIÓN I-9 ENTIDADES EN TERCERA FORMA NORMAL	19
ILUSTRACIÓN I-10 ENTIDAD EN TERCERA FORMA NORMAL	20
ILUSTRACIÓN I-11 ENTIDADES EN CUARTA FORMA NORMAL	21
ILUSTRACIÓN I-12 EJEMPLO DE DEPENDENCIA DE REUNIÓN	22
ILUSTRACIÓN I-13 ENTIDADES EN 5FN	23
ILUSTRACIÓN I-14 COMPONENTES DE UNA BDOO	24
ILUSTRACIÓN II-1 INTEGRACIÓN DE LA INFORMACIÓN EN UNA BDM	29
ILUSTRACIÓN II-2 ACCESO A UNA BDM	31
ILUSTRACIÓN II-3 ELEMENTOS CONSOLIDADOS	31
ILUSTRACIÓN II-4 ROTACIÓN DE CUBO	33
ILUSTRACIÓN II-5 ROTACIÓN DE CUBO II	33
ILUSTRACIÓN II-6 JERARQUÍAS EN LAS DIMENSIONES	34
ILUSTRACIÓN II-7 MUESTRA DE CUBO	35
ILUSTRACIÓN II-8 JERARQUIZACIÓN DE DIMENSIONES	36
ILUSTRACIÓN II-9 ESQUEMA DE TIPO ESTRELLA	39
ILUSTRACIÓN II-10 ESQUEMA DE TIPO ESTRELLA DE LA ILUSTRACIÓN II-9	41
ILUSTRACIÓN II-11 ESQUEMA DE TIPO SNOWFLAKE	42
ILUSTRACIÓN II-12 EJEMPLO DE TABLA CONSOLIDADA POR REGIÓN Y PRODUCTO	43
ILUSTRACIÓN II-13 DATOS PARA ARMAR UNA SENTENCIA SELECT MULTIDIMENSIONAL	54
ILUSTRACIÓN II-14 SELECCIONA TIPO DE MOTOR DE BD	56
ILUSTRACIÓN II-15 ARCHIVO DE BD	56
ILUSTRACIÓN II-16 DIMENSIONES Y HECHOS	57
ILUSTRACIÓN II-17 OPCIONES DE CONSULTA	57
ILUSTRACIÓN II-18 ATRIBUTOS DE HECHOS	58
ILUSTRACIÓN II-19 JERARQUÍAS Y DIMENSIONES	58
ILUSTRACIÓN II-20 DISEÑO DE LA CONSULTA	59
ILUSTRACIÓN II-21 RESULTADO DE CONSULTA EN EXCEL 2000	59
ILUSTRACIÓN II-22 DRILL-DOWN EN UNA DIMENSIÓN	60
ILUSTRACIÓN III-1 CUBO DE VENTAS	63
ILUSTRACIÓN III-2 ATRIBUTOS Y JERARQUÍAS EN LAS DIMENSIONES	63
ILUSTRACIÓN III-3 ESPARCIDAD DE DATOS	64
ILUSTRACIÓN III-4 IDENTIFICACIÓN DE PROBLEMAS	66
ILUSTRACIÓN III-5 REPORTE DIMENSIONAL	67
ILUSTRACIÓN III-6 ENTIDADES NORMALIZADAS (VENTAS)	69
ILUSTRACIÓN III-7 ENTIDAD DENORMALIZADA	69
ILUSTRACIÓN III-8 ENTIDADES NORMALIZADAS	72
ILUSTRACIÓN III-9 ENTIDADES DENORMALIZADAS	72
ILUSTRACIÓN IV-1 RESUMEN ANUAL DE PRESUPUESTADO Vs. ACTUAL	81
ILUSTRACIÓN IV-2 REPORTE DETALLADO ANUAL PRESUPUESTADO Vs. ACTUAL	81
ILUSTRACIÓN IV-3 DETALLE DE PAGOS PENDIENTES	82

ILUSTRACIÓN IV-4 DIAGRAMA ENTIDAD-RELACIÓN ACTUAL DE "INTERMEDIARIA DE MEDIOS"	83
ILUSTRACIÓN IV-5 IDENTIFICACIÓN DE TABLA DE HECHOS Y DIMENSIONES	86
ILUSTRACIÓN IV-6 JERARQUÍAS EN LAS DIMENSIONES	86
ILUSTRACIÓN IV-7 DIAGRAMA SNOWFLAKE	87
ILUSTRACIÓN IV-8 TIPO DE MOTOR DE BD	88
ILUSTRACIÓN IV-9 UBICACIÓN DE LA BD	89
ILUSTRACIÓN IV-10 ATRIBUTOS PARA LA CONSULTA	89
ILUSTRACIÓN IV-11 OPCIONES DE CONSULTA	90
ILUSTRACIÓN IV-12 ATRIBUTOS DE HECHO	90
ILUSTRACIÓN IV-13 DIMENSIONES Y JERARQUÍAS	91
ILUSTRACIÓN IV-14 DISEÑO DE LA CONSULTA	91
ILUSTRACIÓN IV-15 REPORTE DE PAGOS PENDIENTES	92

BIBLIOGRAFÍA

1. Datamation OLAP And OLAP Server Definitions,
www.datamation.com/Plugin/issues/1996/april15/04bevalgls.htm, USA 1996
2. F. Korth Henry, Fundamentos de Base de Datos, McGraw Hill, México 1996.
3. J.Kelly Thomas, Dimensional Data Modeling,
www.sybase.com/mdbs/dimensionales/dimensional.html, USA 1999
4. Microsoft SQL Server 7.0 OLAP Services, Editorial Microsoft Corporation, USA 1998
5. OLAP, www.busintel.com/syn5.htm
6. OLAP, www2.computerworld.com/home/features.nsf/all/981130qs
7. Ortiz Ortiz Zulema, Modelo del Sistema de Contratación de Medios Para Arquitectura Cliente – Servidor: Visual Basic y SQL Server en una Empresa De Publicidad, México 1997, 48 p. Tesis (lic. Informática), UNAM, FES-C
8. Pendse Niegel, The OLAP Report, www.olapreport.com , USA 1999
9. Pilot Software, OLAP Report, www.pilotsw/olap/olap.html, USA 1998
10. Raden Niel, Multidimensional Modeling, www.netmar.com/~nraden/iw0196_1.html,
USA 1996.
11. S. Gill Harjinden, Datawarehousing, Prentince Hall, México 1996.
12. www.seagatesoftware.com/productos/i/library/reviews/content04.asp, USA 1999
13. Transforming OLTP Data to a Star Schema,
www.micosft.com/support/sql/cont.../transforming_oltp_data_to_star_schema.asp,
USA 1999