

4
2ej.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

"UNA METODOLOGIA PARA DISEÑO DE BASES DE DATOS
RELACIONALES USANDO EL MODELO ENTIDAD-RELACION"

T E S I S
QUE PARA OBTENER EL TITULO DE
M A T E M A T I C O
P R E S E N T A

MARIA DE LOS ANGELES BAUTISTA ZAMUDIO

DIRECTOR DE TESIS: M. EN C. AMPARO LOPEZ GAONA

MEXICO, D.F.

1992

FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está 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.

INDICE

INTRODUCCION.	5
---------------	---

CAPITULO 1. PANORAMA GENERAL DE LAS BASES DE DATOS.

INTRODUCCION.	9
SECCION 1.1	OBJETIVOS DE UNA BASE DE DATOS. 10
SECCION 1.2	COMPONENTES DE UN SISTEMA DE BASES DE DATOS. 14
SECCION 1.3	ARQUITECTURA DE UNA BASE DE DATOS. 18
SECCION 1.4	CLASIFICACION DE LOS MODELOS DE BASES DE DATOS. 20
SECCION 1.5	BASES DE DATOS RELACIONALES. 27
SECCION 1.6	NORMALIZACION. 32

CAPITULO 2. METODOLOGIA PARA DISEÑO DE BASES DE DATOS RELACIONALES USANDO EL MODELO ENTIDAD-RELACION.

INTRODUCCION.	43
SECCION 2.1	PANORAMA GENERAL DE LA METODOLOGIA. 44
SECCION 2.2	MODELO ENTIDAD-RELACION. 46
SECCION 2.3	DESARROLLO DEL PASO 1 DE LA METODOLOGIA. 57
SECCION 2.4	DESARROLLO DEL PASO 2 DE LA METODOLOGIA. 65
SECCION 2.5	DESARROLLO DEL PASO 3 DE LA METODOLOGIA. 77

CAPITULO 3.	APLICACION DE LA METODOLOGIA.	82
CONCLUSIONES.		97
BIBLIOGRAFIA.		100

INTRODUCCION

INTRODUCCION

En el pasado, la actividad de procesar datos relacionados con la industria o negocios, era realizada por un empleado cuyas herramientas de trabajo, eran una pluma para escribir y una colección de libros. Sin embargo, estos empleados sólo podían trabajar en una pequeña compañía, ya que conforme la compañía crecía, el tamaño de los libros se incrementaba, llegando a la necesidad de tener varios empleados para mantenerlos.

Posteriormente varios medios de mecanización fueron introducidos y se hizo eficiente uso de ellos, además el trabajo fue dividido en grupos.

Con el auge de las computadoras se suplantó el anterior método oficinesco y grandes archivos de datos pudieron ser procesados, mediante un programa, con el uso de cintas magnéticas de computadoras.

Posteriormente las organizaciones públicas y privadas en muchas naciones, fueron afectadas y beneficiadas cada vez más, por la tecnología para almacenar, manejar y procesar datos.

El costo de hardware y software para almacenamiento y recuperación de información disminuyó rápidamente, hasta el punto en que era más económico almacenar datos en archivos y bancos de datos, que en documentos.

Desde 1960 muchas organizaciones, gobiernos, industrias y educación fueron desarrollando grandes sistemas de información, por lo tanto, las organizaciones comenzaron a integrar varios sistemas de aplicación computarizados y archivos de datos, asociados en poderosas bases de datos, esto por la necesidad de tener independencia de datos entre programas de aplicación, tener habilidad para compartir datos entre los usuarios, no tener redundancia de información, tener alta eficiencia, seguridad e integridad es decir, el propósito por el que es importante utilizar bases de datos, es el de proporcionar independencia de datos,

tener compartimiento y almacenamiento de datos no redundantes, eficiencia, administración, control, seguridad etc, los cuales se explicarán con más detalle en el capítulo uno.

Para crear una base de datos se debe hacer un diseño previo ya que este permite reducir el número de dependencias de datos que se analizan, también reduce el número de objetos bajo consideración, quita redundancia, etc, es decir este permite ver y resolver la mayoría de los posibles problemas que pueden ocurrir en la base de datos antes de que esta se cree.

Por lo tanto el objetivo de esta tesis, es dar a conocer, una metodología para diseñar bases de datos, mediante el modelo entidad-relación, el cual proporciona una forma para que el diseñador, se comunique con el usuario, mediante la representación que utiliza el modelo, con el fin, de que el usuario, entienda fácilmente el diseño y pueda comprobar, que lo que solicitó, efectivamente es lo que el diseñador entendió.

El objetivo de esta metodología, es poder obtener las mejores relaciones que satisfagan lo mejor posible los objetivos de la base de datos y para esto la metodología se desarrollará paso por paso detalladamente, ejemplificando cada paso para que se entiendan mejor y por último se aplicará a un problema real para ver su utilidad.

Para cumplir los objetivos de este trabajo la tesis, está estructurada de la siguiente manera:

En el capítulo uno, se da a conocer un panorama general, sobre las bases de datos, sus componentes, su clasificación y con más detalle, lo que son las bases de datos relacionales y un método para normalizar relaciones, todo con el fin de proporcionar al lector, todos los conceptos, que se mencionan en la metodología y familiarizarlo con las bases de datos.

En el capítulo dos, se presenta lo que es la metodología, de tal forma, que este se estructura de la siguiente manera:

Primero se dan a conocer, los elementos que componen el modelo entidad-relación, los cuales son: entidades, atributos y relaciones y el como identificarlos y después se da a conocer la abstracción de categorías, tales como, la jerarquía de generalización y la jerarquía de subconjunto, las cuales representan una extensión del modelo entidad-relación y que permiten hacer una agrupación de entidades. Posteriormente, se proporcionan las reglas para transformar, el modelo entidad-relación extendido, al modelo relacional y por último, se explican un poco más, los pasos, que se deben seguir para normalizar las relaciones.

Por último, en el capítulo tres, se muestra cómo aplicar la metodología expuesta en los capítulos anteriores, en el planteamiento de la solución, a un problema en específico.

CAPITULO 1

CAPITULO I

PANORAMA GENERAL DE LAS BASES DE DATOS

INTRODUCCION

El objetivo de este capítulo, es introducir al lector, a los conceptos básicos sobre las bases de datos y en especial, a las bases de datos relacionales, ya que lo que se va a tratar en esta tesis, va a ser considerando este tipo.

Para ello, el capítulo está estructurado de manera que, en la sección 1.1 se verán los objetivos de una base de datos, en la sección 1.2 los componentes de un sistema de bases de datos, en la sección 1.3 la arquitectura de estas, en la sección 1.4 la clasificación de los modelos para ellas, en la sección 1.5 un desarrollo más detallado de las bases de datos relacionales y por último, se darán a conocer aspectos sobre el concepto de normalización.

Lo primero que se hará, es dar a conocer la definición de una base de datos.

Una base de datos es:

Un depósito de datos almacenados, que se puede considerar, como la unión de varios archivos de datos independientes, en donde se trata de eliminar la redundancia que existe entre ellos, y además se puede decir de esta, que los usuario pueden tener acceso a la misma parte de la base de datos y utilizarla con distintos propósitos.

Cuyos principales objetivos son los siguientes:

- 1) Independencia de datos.
- 2) Compartimiento y almacenamiento de datos no redundantes.
- 3) Relacionamiento.
- 4) Integridad.
- 5) Flexibilidad de acceso.
- 6) Seguridad.
- 7) Desempeño y eficiencia.
- 8) Administración y control.

- 1) El propósito de tener la independencia de datos, por ejemplo con respecto a los datos físicos, es poder tener la habilidad, para cambiar la organización física de los datos y no permitir hacer obsoletos los programas de aplicación.

Por ejemplo; cambios en la localidad de datos, en las uniones de los datos internos, estrategias de clase interna, en la representación física, es decir, en como los datos estan representados, en los depósitos de almacenamiento y rutas de acceso disponible.

El propósito de tenerla con respecto a los datos lógicos, es aislar las aplicaciones de los cambios, que puedan suceder en la organización lógica de las bases de datos usada.

Más globalmente, la independencia de datos; es aquella que denota

el aislamiento, a un programa de aplicación de la organización física o lógica y a las consideraciones, de almacenamiento de la estructura de datos y archivos que están en uso, además los programas no deberán estar sujetos, a la influencia externa. Generalmente, la falta de independencia de datos, es debida a los cambios que ocurren en la práctica, debido a los requerimientos imprevistos que surgen en el futuro.

- 2) El propósito de compartir y almacenar datos no redundantes, es facilitar aplicaciones, para compartir una base de datos integrada, la cual contenga, todos los datos necesarios para las aplicaciones y además, eliminar cuanto sea posible, la necesidad de almacenar datos redundantes.

Un camino hacia el compartimiento, es remover la redundancia, pero sin embargo, a veces puede ser necesaria alguna redundancia de datos, para evitar, excesivo tiempo de acceso a registros.

Específicamente, el compartimiento de datos, consiste en tener diferentes programas de aplicación compartidos independientemente, en una base de datos integrada y los cuales tienen que operar inadvertidos de la existencia de otro, la falta de este, es causada por los obstáculos de la dependencia de datos, tales como la inhabilidad que tiene un programa, para ver y compartir solamente un subconjunto de datos en un archivo utilizado por otro programa. Con respecto a la redundancia, esta causa inconsistencia en la precisión de datos, formatos y nombramientos.

La redundancia de registros y archivos, es debida, a que en el momento de nombrar y codificar datos para programas específicos, se agregan datos permanentemente a esos programas.

- 3) El propósito del relacionamiento, es poder definir relaciones entre entidades (las cuales como se mencionó anteriormente, representan un concepto del mundo real y que tienen un número determinado de atributos) en el nivel lógico exacto tan convenientemente como definir las entidades por ellos mismos, ya que estas son tan importantes e identificables como una entidad o atributo y de las que se debe tener la habilidad de derivar otras asociaciones.

4) Los principales propósitos, de tener la integridad, son: coordinar el acceso de datos para diferentes aplicaciones, propagar la actualización de valores a otras copias y mantener un alto grado de consistencia y corrección de datos.

El preservar la integridad, también contempla el mantener un registro en todos los accesos y cambios a cada dato en cada punto, es decir, si ocurre un cambio o acceso y sucede un error, la integridad debe ser recuperada.

5) El propósito de tener la flexibilidad de acceso, se refiere a la propiedad de acceder datos fácil y eficientemente, en una variedad de caminos.

6) El propósito de la seguridad, es poder asignar, controlar y eliminar los derechos de acceso (leer, cambiar, borrar, insertar, etc.) de algún usuario a algún dato o subconjunto definido de la base de datos ya que los datos, deben ser protegidos de la intrusión no autorizada por accidente o malicia de algún usuario o programa de aplicación.

Además, se puede mencionar, que el administrador de la base de datos, debe tener la facilidad, para asignar y controlar el acceso de los usuarios.

7) El propósito del desempeño y eficiencia, es poder manejar bases de datos de gran tamaño y poder dar atención a las necesidades de acceso a ellas.

Además, la integridad de la base de datos, depende mucho del buen desempeño y de la eficiencia, debido al gran tamaño de la base de datos.

También, la necesidad de tener una organización física, eficiente, es debido a que siempre se agregan más datos y relaciones a la base de datos.

Se puede mencionar, que una de las causas de conflictos de desempeño, es debida, a que muchos usuarios tienen diferentes requerimientos sobre el mismo dato, es decir, por ejemplo; puede

presentarse la situación de que algunos usuarios, sólo necesitan acceso de lectura de algunos datos, mientras que otro grupo necesita estar actualizando constantemente esos datos.

8) El propósito de la administración y control, es describir y controlar los datos, lo cual, es responsabilidad del administrador de bases de datos, el cual tiene que diseñar la base de datos global, las definiciones de los datos y el camino de transformación para el acceso de usuarios a la base de datos, además de determinar niveles de desempeño y eficiencia, mientras toma en cuenta el bien global.

Además, se puede decir, que la descripción y control de almacenamiento, debe ser llevada a cabo fuera de los programas de aplicación y debe ser declarada de manera común o compatible a todos los sistemas de aplicación y lenguajes de programación.

SECCION 1.2

COMPONENTES DE UN SISTEMA DE BASES DE DATOS

Lo primero que se hará en esta sección, es dar a conocer la definición de lo que es un sistema de base de datos.

Un sistema de bases de datos es: un sistema que tiene como propósito general, el registrar y mantener los datos, cuyo significado puede ser necesario, para los procesos de toma de decisiones, de una organización.

Un sistema de bases de datos, incluye cuatro componentes principales:

- 1) DATOS
- 2) HARDWARE
- 3) SOFTWARE
- 4) USUARIOS

1) Con respecto a los datos, se tiene que estos son, *una representación de hechos o ideas que estan adaptados a una interpretación humana*, los cuales, están almacenados en el sistema y se dividen en una o varias bases de datos.

A continuación, se da una definición de lo que es una base de datos, en términos distintos a los presentados en la sección anterior.

Una base de datos es:

Un conjunto de datos de operación (los cuales son datos que se refieren al funcionamiento de cualquier empresa, y no incluyen datos de entrada, ni de salida, ni cualquier otra información de índole transitoria) almacenados y utilizados por los sistemas de aplicación de cualquier organización comercial, científica, o de otra clase.

En donde con datos de entrada y datos de salida se hace referencia a lo siguiente:

Datos de entrada: *Son aquellos que se refieren a la información que alimenta al sistema desde afuera, pero que no forma parte de la base de datos.*

Datos de salida: *Son aquellos que se refieren a los mensajes e informes que salen del sistema, que tampoco forman parte de la base de datos, y que contienen información derivada de los datos de operación.*

Para que se entienda mejor el concepto de datos de operación, se presenta un ejemplo:

Se tiene el siguiente problema:

Una empresa desea almacenar información, sobre los proyectos que está desarrollando; las partes que se utilizan en esos proyectos; los proveedores que suministran estas partes; los depósitos donde estas se almacenan; y los empleados que trabajan en los proyectos.

Las entidades, de este problema, son:

- *D PROYECTOS
- *D PARTES
- *D PROVEEDORES
- *D DEPOSITOS
- *D EMPLEADOS.

En donde **entidad:** *es aquella, que designa cualquier objeto distinguible, que puede representarse en la base de datos.*

En otras palabras, es un objeto del mundo real acerca del cual, la información es registrada.

Además, una entidad, se compone de un conjunto de parejas, cada una formada por el nombre de un atributo y el valor del atributo.

Por ejemplo la entidad:

EMPLEADO

está compuesta por:

NOMBRE DEL ATRIBUTO : VALOR DEL ATRIBUTO
EMPLEADO-NUM : 291
NOMBRE : RAMIREZ
DEPARTAMENTO : ADMINISTRACION

Además existen asociaciones que relacionan a las entidades, por ejemplo; en el problema de la empresa hay una asociación entre PROVEEDORES y PARTES, ya que cada proveedor suministra ciertas partes y cada parte es abastecida por determinados proveedores y se tiene que estas asociaciones, las cuales se pueden considerar como tipos especiales de entidades, forman parte de los datos de operación, tanto como las entidades que se asocian.

2) Con respecto al hardware, se puede decir, que se compone de los dispositivos de almacenamiento secundario, es decir; los discos, tambores, cintas, etc, en donde se encuentra la base de datos, junto con los dispositivos asociados; tales como las unidades de control, los canales de comunicación, etc.

3) Con respecto al software, se puede mencionar, que existe un nivel entre el almacenamiento real de los datos y los usuarios del sistema, que recibe el nombre de sistema de administración de bases de datos, el cual se encarga de manejar, las solicitudes de acceso a la base de datos, hechas por los usuarios, además, protege a los usuarios contra los detalles a nivel de hardware y proporciona una vista de la base de datos, que está por encima del nivel de hardware y apoya las operaciones del usuario.

4) Con respecto a los usuarios, generalmente se consideran 3 clases:

- 1) **El programador de aplicaciones;** el cual se encarga de escribir programas de aplicación, que utilizan bases de datos.
- 2) **El usuario final;** el cual se encarga de acceder la base de datos desde una computadora.

3) El administrador de bases de datos; el cual se encarga del control general de las bases de datos, es decir tiene a su cargo entre otras cosas:

Crear la descripción original de la estructura de la base de datos.

Decidir que entidades, son de interés para la empresa y cuales datos, deben registrarse con respecto a esas entidades.

Decidir de que forma se van a representar los datos en la base de datos y cual va a ser la estrategia de acceso.

Conceder, a los usuarios, la autorización del acceso a la base de datos o partes de esta.

Definir los procedimientos de validación y las estrategias para respaldar y recuperar datos.

Hacer ajustes adecuados, cuando los requerimientos cambien.

Con respecto a la arquitectura, se puede mencionar, que el fin de esta, es mostrar el dominio de la organización de la base de datos, es decir, mostrar cómo puede ser vista la organización de los datos en la base de datos.

Por lo general la arquitectura se divide, en los tres niveles que a continuación se mencionan:

- 1) INTERNO: El cual se acerca más al almacenamiento físico, es decir, es el que se refiere, a la manera en como los datos se almacenan en los dispositivos de almacenamiento.
- 2) EXTERNO: El cual se acerca más a los usuarios, es decir, es el que se refiere a la manera en como cada usuario ve los datos, de acuerdo a sus intereses.
- 3) CONCEPTUAL: El cual es un nivel de mediación entre los 2 anteriores y que representa la estructura y los valores de los datos utilizados.

Por ejemplo; si se considera que el nivel externo se relaciona con las vistas de los usuarios, el nivel conceptual se puede considerar como el que define una vista de la comunidad de usuarios, es decir, habrá muchas vistas externas, cada una compuesta por una representación abstracta de alguna parte de la base de datos, habrá una vista conceptual, compuesta por una representación abstracta de la base de datos en su totalidad y habrá una vista interna, que represente la base de datos total, tal como está almacenada.

Lo que queda ahora por definir, es lo que son las vistas externas, internas y conceptuales.

VISTA EXTERNA: *Es el contenido de la base de datos, tal como lo ve un usuario determinado.*

VISTA CONCEPTUAL: *Es una representación del contenido total de la información de la base de datos, es decir, es una vista del contenido total de la base de datos.*

VISTA INTERNA: *Esta es una representación, del nivel más bajo de la base de datos en su totalidad.*

Además existen dos niveles de correspondencia, entre las vistas que son:

La correspondencia conceptual/interna, la cual define la correspondencia entre la vista conceptual y la base de datos almacenada.

La correspondencia externa/conceptual, la cual define la correspondencia entre una vista externa determinada y la vista conceptual.

Se tiene, que el principal método para el diseño de base de datos, es la construcción de modelos, los cuales representan la estructura de la base de datos y se puede decir, que son abstracciones del mundo real. En otras palabras, un modelo de datos: es una colección de conceptos matemáticamente definidos, que ayudan a considerar y expresar los objetos, atributos y relaciones entre objetos y sus correspondientes operaciones, es decir, proveen una descripción conceptual de los datos.

Además, se tiene que los tres principales tipos de modelos de bases de datos, los cuales difieren en la manera en que permiten al usuario ver y manipular las asociaciones son:

- 1) JERARQUICO.
- 2) RED.
- 3) RELACIONAL

En general, se puede decir, que un modelo de datos, consiste de 2 elementos, los cuales son:

- ***) La notación matemática, para expresar los datos y las relaciones.
- ***) Las operaciones sobre los datos, que nos sirven para expresar preguntas y otras manipulaciones sobre los datos.

En un panorama general, se puede decir, que los modelos de datos jerárquicos y de red, representan los objetos en registros que están organizados, usando relaciones binarias uno a muchos, como nodos en árboles y redes respectivamente y el modelo de datos relacional, está basado, en el concepto matemático, de una relación, que es un conjunto de tuplas.

1) Con respecto al modelo de bases de datos de tipo jerárquico, lo primero que se dará a conocer, es el concepto de jerarquía. Una jerarquía es; una colección de árboles, cada uno de los cuales está formado por nodos y enlaces que sólo van en la dirección de hijo a padre, es decir aquí cada nodo hijo tiene sólo un nodo padre.

Además, aquí se utiliza el registro de tipo lógico, para hablar de las jerarquías y se tiene, que un nodo representa un registro de tipo lógico y que los hijos de un nodo, representan conjuntos de registros que están asociados, con cada registro del padre y además existe el número de serie, que representa el lugar del registro.

Pero se tiene que en este enfoque jerárquico, hay pérdida de simetría, lo cual, es un inconveniente, ya que obliga a que el usuario dedique tiempo para resolver problemas, que se introducen por este tipo de estructura. Además se presentan anomalías, en relación con cada una de las 3 operaciones básicas (que son: insertar, suprimir y actualizar) debido a que se trata, de una situación de muchos a muchos, pero que, no se presentarían, en una situación de uno a muchos.

La operación básica en este modelo, es la de recorrer un árbol en la dirección de padre a hijo.

Y por último, se tiene, que si se necesita cambiar la descripción de una entidad, se enfrenta uno con el problema de tener que buscar en toda la vista, para hallar cada ocurrencia de la entidad, o también se puede topár, con la posibilidad de introducir inconsistencias.

Ejemplo:

A continuación, se muestra un árbol que representa la información a cerca del personal de una organización.

La información que se presenta se refiere a:

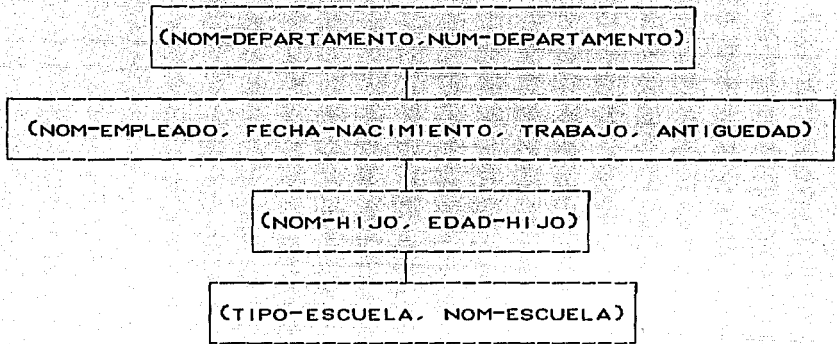
Que empleados trabajan en cada departamento.

Cuales son los hijos de cada empleado.

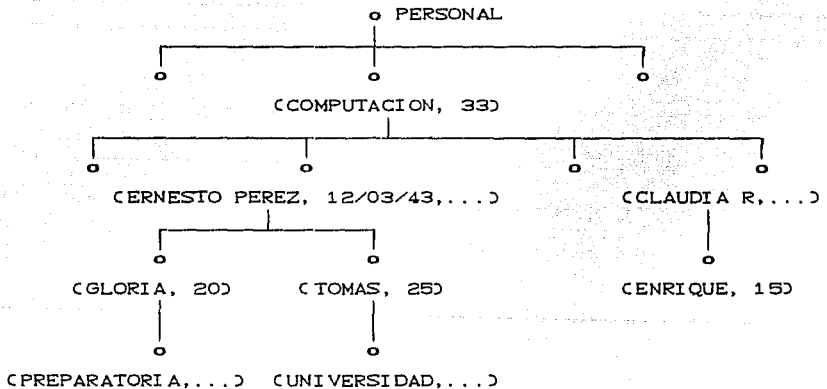
Que educación tiene cada hijo.

NOMBRE DEL ARBOL: PERSONAL

EL ARBOL ES:



INSTANCIAS DE LA BASE DE DATOS JERARQUICA:



2) Con respecto al modelo de bases de datos de tipo red, se tiene que los datos se representan por registros y sus enlaces, es decir su representación se hace por medio de una gráfica dirigida, en donde los nodos de la gráfica, corresponden a los registros y si hay un enlace entre 2 registros, digamos R1 y R2, y este es un enlace muchos a uno,

entonces se representará por un arco, que va del nodo R1 al nodo R2 y además los nodos y los arcos se etiquetan, con los nombres de sus registros y enlaces.

Aquí, en lugar de referirse a un conjunto de entidades, se hace referencia a un conjunto de registros de tipo lógico, el cual es esencialmente una relación y en donde el registro de tipo lógico en el modelo relacional, representa una tupla, y el esquema de relación se denomina formato de registro lógico, en este modelo, además las componentes de cada registro se denominan campos.

Además, este tipo, es una estructura, en donde una ocurrencia de registro determinada, puede tener cualquier número de superiores inmediatos o dependientes inmediatos.

La desventaja principal de este tipo de bases de datos, es su excesiva complejidad, en la estructura de datos.

Y una de las operaciones de este tipo de modelo de bases de datos es:

SELECCION : la cual, es sobre los registros de tipo lógico y que es similar a la selección del modelo relacional.

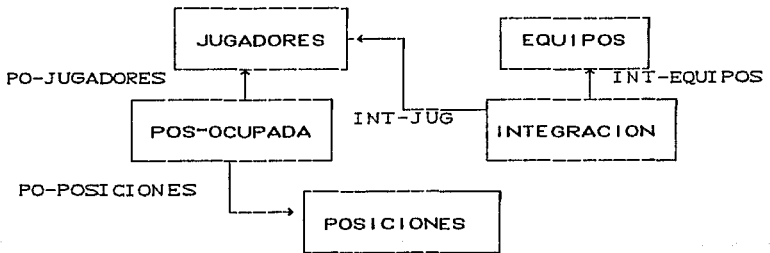
Ejemplo:

A continuación, se muestra una gráfica que representa la información acerca de los equipos y jugadores de beisbol.

La información que se presenta se refiere a:

Los equipos en los que ha estado cada jugador.

Las posiciones que ha ocupado cada jugador.



3) Con respecto al modelo de bases de datos relacional, se puede mencionar que formalmente está basado en principios matemáticos, incluyendo la teoría de relaciones.

Además una característica de este tipo de bases de datos, es que toda la información de la base de datos, tanto entidades como asociaciones se representa en forma de tablas.

A continuación se proporcionará la definición de lo que es una relación.

Una relación; es una tabla de 2 dimensiones, con N columnas o atributos, cada una compuesta de un conjunto de valores de un atributo y que se refieren a un dominio (el cual, es un depósito de valores, del cual se sacan los que aparecen en una determinada columna), y compuesta de un conjunto de tuplas o renglones.

Al conjunto de nombres de atributos de una relación, se le llama esquema de la relación y se puede representar de la siguiente manera:

$REL(A_1, A_2, \dots, A_N)$

en donde : REL es el nombre de la relación y

A_1, A_2, \dots, A_N son los nombres de los atributos.

Para aclarar esta definición, se explicará mediante un ejemplo:

RELACION LLAMADA EMPLEADO

FORMADA POR LOS ATRIBUTOS :

EMP-NUM (NUMERO DE IDENTIFICACION
DEL EMPLEADO)

EMP-NOMBRE (NOMBRE DEL EMPLEADO)

SUCURSAL (EN DONDE EL EMPLEADO
TRABAJA)

En donde, la representación notacional o esquema de relación es:

EMPLEADO (EMP-NUM, EMP-NOMBRE, SUCURSAL)

Y la tabla o relación es:

EMPLEADO

EMP-NUM	EMP-NOMBRE	SUCURSAL
E1	BAUTISTA	REVES
E2	BAUTISTA	LAGUNILLA
E3	LOPEZ	REVES
E4	RAMIREZ	----

Además, las tablas tienen las siguientes características:

- 1) En cada columna o atributo, el dominio, está asignado a distinto nombre.
- 2) Cada renglón es único, no puede haber renglones repetidos.
- 3) Los renglones y las columnas pueden ser ordenadas, sin afectar el contenido de los datos.
- 4) Cada tupla de la relación, debe tener una llave (la cual es un atributo simple o atributo compuesto) por medio de la cual la tupla puede ser identificada de manera única y diferenciada de otras tuplas de la relación.
- 5) Cada renglón, representa una entidad o asociación entre entidades.
- 6) Cada columna, tiene una etiqueta que la describe, la cual proporciona algunos aspectos del objeto que representa.

La colección de esquemas de relación, que se utiliza para representar la información, se denomina esquema de bases de datos y los valores actuales de la correspondiente relación, se denominan base de datos relacional.

Además, se puede decir, que una llave primaria: es el conjunto de atributos, que identifican las tuplas y que generalmente debe ser, aquél con el menor número de atributos.

En el ejemplo anterior se tiene que la llave primaria es EMP-NUM.

Y algunas operaciones en este tipo de modelo de base de datos son:

SELECCION : La cual consiste, de escoger cierto número de renglones de una tabla, que cumplan cierta condición.

REUNION : La cual consiste, en conectar los datos de 2 o más relaciones.

PROYECCION : La cual consiste, en seleccionar un número determinado de columnas, es decir formar subtablas.

Como el propósito de esta tesis, es presentar una metodología sobre el diseño de bases de datos relacionales, se procederá a desarrollar mas a fondo las estructuras de datos relacionales

Primero, se definirá una relación, en términos más precisos que la anterior definición.

RELACION: Si D_1, D_2, \dots, D_n son conjuntos no necesariamente distintos, se dice que R es una relación sobre estos n conjuntos si es un conjunto de tuplas (las cuales generalmente representan entidades del mundo real) $\langle d_1, d_2, \dots, d_n \rangle$ de tal forma que d_1 pertenece a D_1, \dots, d_n pertenece a D_n .

En donde los conjuntos D_1, D_2, \dots, D_n son los dominios de R .

Además, se tiene que el grado de la relación R es n , en donde con grado se hace referencia al número de dominios de R .

Al número de tuplas de R se denomina como la cardinalidad de la relación.

Para evitar confusiones, se mencionará la diferencia entre dominio y atributo.

Un atributo (el cual es una propiedad de una entidad) representa el uso de un dominio dentro de la relación.

Un dominio, es un depósito de valores, del cual se sacan los que aparecen en una determinada columna, en otras palabras, es un conjunto de valores, para la columna asociada y que usualmente es un conjunto de enteros, números reales o cadenas de caracteres.

De esta forma, se tiene, que a las relaciones de grado 1 se les denominan relaciones unarias.

Las relaciones de grado 2 son relaciones binarias.

Las relaciones de grado 3 son relaciones ternarias.

Las relaciones de grado n son relaciones n-arias.

A continuación se da un ejemplo, de una relación llamada PARTE, la cual es de grado 5, en el cual se involucran estos conceptos.

La relación como se mencionó anteriormente, se representará en forma de tabla, en donde cada renglón, representa una tupla de la relación:

PARTE

P-NUM	P-NOM	COLOR	PESO	PAIS
P1	TUERCA	ROJO	10	MEXICO
P2	PERNO	VERDE	20	BRASIL
P3	TORNILLO	AZUL	20	ARGENTINA
P4	TORNILLO	ROJO	15	MEXICO

En donde los 5 dominios, son conjuntos de valores que representan, respectivamente, números, nombres, colores, pesos y lugares en donde se almacenan las partes.

Además se tiene que la cardinalidad de la relación *PARTE* es 4.

Por último, se puede observar en la siguiente figura, que se han declarado 5 dominios *NUMERO-PARTE*, *NOMBRE-PARTE*, *COLOR*, *PESO* y *LUGAR* y una relación *PARTE* la cual se define con 5 atributos *P-NUM*, *P-NOM*, *COLOR*, *PESO* y *PAIS* y además se especifica que cada uno de los atributos se extrae de un dominio correspondiente.

DOMINIO NUMERO-PARTE	CARACTER(6)
DOMINIO NOMBRE-PARTE	CARACTER(20)
DOMINIO COLOR	CARACTER(6)
DOMINIO PESO	NUMERICO(6)
DOMINIO LUGAR	CARACTER(20)

RELACION PARTE

(P-NUM : DOMINIO NUMERO-PARTE
P-NOM : DOMINIO NOMBRE-PARTE
COLOR : DOMINIO COLOR
PESO : DOMINIO PESO
PAIS : DOMINIO LUGAR)

Continuando, con este tipo de bases de datos se tiene, que todas las relaciones deben satisfacer, que cada valor de atributo en cada tupla, sea atómico, es decir en cada intersección de un renglón y una columna de la tabla siempre haya exactamente un valor.

Además, se permite que haya valores nulos, es decir, valores que representan algo desconocido o inaplicable como por ejemplo los retardos de un empleado, que está de vacaciones.

Y como anteriormente se dijo, toda relación tiene una llave primaria, tal vez compuesta, que no es redundante y en donde con llave primaria se hace referencia a:

Aquel atributo cuyos valores son únicos dentro de una relación y que se utilizan para identificar las tuplas de la relación, la cual sirve de identificador único de entidades.

Una llave primaria compuesta, es aquella, que esta compuesta de 2 o más atributos.

Además, se debe mencionar que en una relación, se puede encontrar a veces más de una combinación de atributos, que posean la propiedad de identificación única, es decir mas de una llave candidata, en donde con llave candidata, se hace referencia a un conjunto de atributos, que identifican a cada tupla.

Ahora se explicará, este tipo de llaves, por medio de un ejemplo.

En la siguiente figura, se muestra una relación *PROVEEDOR*.

Aquí se observa, que cada proveedor tiene un número de proveedor único y un nombre de proveedor único, por lo tanto, se tienen como llaves candidatas *S-NUM* y *S-NOM* pero en este caso *S-NUM*, se considerará como la llave primaria de la relación, además *S-NOM* frecuentemente, recibe el nombre de llave alterna.

PROVEEDOR

S-NUM	S-NOM	ESTADO	CIUDAD
S1	SALAZAR	20	LONDRES
S2	JARAMILLO	10	PARIS
S3	BERNAL	30	PARIS
S4	CAICEDO	20	LONDRES
S5	ALDANA	30	ATENAS

Además, se puede decir, que una relación en este tipo de bases de datos, tiene 2 componentes:

EXTENSION: *La cual, es el conjunto de tuplas, que aparecen en la relación y que cambia según las tuplas, se creen, destruyen y actualizen, es decir, son los valores de los atributos que forman el cuerpo de la relación.*

COMPRESION: *La cual, es la parte permanente de la relación, y corresponde a lo que se especifica en el esquema de relación (que está integrado por el nombre de la relación y los nombres de los atributos) y el cual debe satisfacer las siguientes reglas de integridad:*

1) REGLA DE INTEGRIDAD DE LA RELACION: la cual dice que ningún valor de una llave primaria, debe ser nulo, ya que todas las entidades deben tener alguna identificación única.

2) REGLA DE INTEGRIDAD DE REFERENCIA: la cual dice que si D es un dominio primario, es decir un *dominio específico en donde se define alguna llave primaria formada de un solo atributo*; R_1 y R_2 relaciones; X, Y atributos de R_1 y R_2 respectivamente; además X se define sobre D y R_2 tiene llave primaria definida también sobre D, entonces cada valor de X debe ser o nulo o igual a Y, donde Y por ejemplo, es el valor de la llave primaria de alguna tupla de R_2 .

A un atributo como X se le suele llamar llave foránea.

Tanto las llaves foráneas, como las primarias, proporcionan un medio, para representar asociaciones entre tuplas.

Por lo tanto la comprensión debe incluir la especificación de atributos que constituyen la llave primaria, pero esta especificación comprende la restricción de no admitir valores nulos aquí y además debe incluir la especificación de todas las llaves foráneas de la relación.

Por último, como se mencionó en la sección anterior, algunas de las operaciones que se pueden realizar son:

SELECCION : La cual consiste, de escoger cierto número de renglones de una tabla, que cumplan cierta condición.

La cual en términos matemáticos equivale a lo siguiente:

$S_F(R)$ denota la selección de una relación R y la cual especifica el conjunto de tuplas T de R, que satisfagan el hecho de que las componentes de cada una de estas hagan verdadera la fórmula F.

Donde F es una fórmula que envuelve:

- 1) Operandos que son constantes o números.
- 2) Operadores de comparación: <, =, >, ≤, ≠, ≥.
- 3) Operadores lógicos: ^, v, ¬.

PROYECCION : La cual consiste, en seleccionar un número determinado de columnas, es decir formar subtablas.

La cual en términos matemáticos equivale a lo siguiente:

$\Pi_{i_1, i_2, \dots, i_m}(R)$ denota la proyección de una relación R, de grado K, en las componentes i_1, i_2, \dots, i_m (donde las i_j son enteros distintos cuyos valores están entre 1 y K) que es el conjunto de tuplas formadas por las columnas $C_1, C_2, C_3, \dots, C_m$ tal que hay alguna tupla formada por las columnas CC_1, CC_2, \dots, CC_K en R para la cual $C_l = CC_{i_j}$ para $l=1, 2, \dots, m$.

En el proceso de diseño, se hace uso de herramientas, que sirven para hacer un buen diseño y una de ellas, es la teoría de la normalización, la cual está basada en el concepto de formas normales.

Como en el proceso para definir las formas normales, se hace uso de términos tales como dependencias funcionales (DF) y dependencias multivaluadas (DMV) primero se dará una definición de ellos.

Dependencia funcional :

Sea R una relación y sean X, Y atributos de R, entonces se dice que Y es funcionalmente dependiente de X si y sólo si cada valor de X en R tiene asociado a él exactamente un valor de Y en R (en cualquier instante), en otras palabras Y es funcionalmente dependiente de X si y sólo si siempre que 2 tuplas de R sean iguales en sus valores de X, también sean iguales en sus valores de Y, es decir, si el valor de algunos atributos Y están determinados por el valor de otros atributos X.

Dependencia funcional completa :

Sea R una relación y sean X, Y atributos de R, entonces se dice que Y es funcionalmente dependiente de X en forma completa, si es funcionalmente dependiente de X y no existe un subconjunto propio X_1 de los atributos que forman parte de X tal que Y sea funcionalmente dependiente de X_1 .

Dependencia multivaluada :

Sea R una relación y sean X, Y, Z atributos (posiblemente compuestos) de R, entonces se dice, que la dependencia multivaluada (DMV) $X \twoheadrightarrow Y$ se cumple en R si y sólo si el conjunto de valores de Y que corresponden a un par dado en R, dependen sólo del valor de X y es independiente del valor de Z, es decir, X multidetermina Y si para los valores dados de X hay un conjunto de cero o más valores asociados para Y y este conjunto de valores de Y no está conectado con cualquier valor de los atributos de R-X-Y.

En otras palabras, una dependencia multivaluada, ocurre, cuando un valor de un atributo, determina un conjunto de valores múltiples.

Además, las dependencias multivaluadas, pueden existir, solamente si la relación R tiene al menos 3 atributos, y también se puede decir, que las dependencias multivaluadas, se consideran una generalización de las dependencias funcionales.

Además, tenemos que si R es una relación y X, Y, Z son atributos de R, entonces, la dependencia multivaluada $X \twoheadrightarrow Y$ se cumple si y sólo si también se cumple la dependencia multivaluada $X \twoheadrightarrow Z$.

Ahora, se darán a conocer algunas técnicas de descomposición de relaciones, sin pérdidas, (con esto se hace referencia a que en el momento de descomponer las relaciones no se pierda o aumente la información) para ayudar al diseño de bases de datos.

1) Una relación que está en primera forma normal (es decir, que los valores en los dominios no son compuestos) y no en segunda (es decir, que los atributos que no forman parte de la llave primaria no dependen en forma completa de esta) se puede reducir siempre a un conjunto de relaciones equivalentes a ella y esto se logra, reemplazando las relaciones por proyecciones adecuadas.

En general, si R es una relación con los atributos X, Y, Z posiblemente compuestos y estos satisfacen la dependencia funcional $X \rightarrow Y, Z$ entonces R se puede descomponer sin pérdida en las proyecciones R1 con atributos X, Y y R2 con atributos X, Z. Lo mismo ocurre para la dependencia multivaluada.

Ejemplo:

Se tiene la siguiente relación:

MATERIA	PROFESOR	LIBRO
MATEMATICAS	BERRIOS	TIT1
MATEMATICAS	LOYOLA	TIT1
MATEMATICAS	LOPEZ	TIT1
MATEMATICAS	BERRIOS	TIT2

MATERIA	PROFESOR	LIBRO
MATEMATICAS	LOYOLA	TIT2
MATEMATICAS	LOPEZ	TIT2
QUIMICA	AHULD	TIT3
QUIMICA	AHULD	TIT4

Aquí, se observa que PROFESOR multidepende de MATERIA y también LIBRO multidepende de MATERIA, por lo tanto la relación, se puede descomponer en las siguientes proyecciones:

MATERIA	PROFESOR	MATERIA	LIBRO
MATEMATICAS	BERRIOS	MATEMATICAS	TIT1
MATEMATICAS	LOYOLA	MATEMATICAS	TIT2
MATEMATICAS	LOPEZ	QUIMICA	TIT3
QUIMICA	AHULD	QUIMICA	TIT4

- 2) Sea R una relación y sean X, Y, Z atributos de R, si R satisface las dependencias funcionales $X \rightarrow Y$, $Y \rightarrow Z$ entonces se puede descomponer mejor en sus proyecciones sobre X, Y y Y, Z y no en aquellas sobre X, Y y X, Z.

Lo mismo ocurre, si se trata de dependencias multivaluadas.

Ejemplo:

Se tiene la siguiente relación:

NUM-VENDEDOR	NOM-TIENDA	COD-TIEND
v1	AURRERA	03
v2	BLANCO	05
v3	GIGANTE	02
v4	BLANCO	05

Aquí, se observa que NOM-TIENDA depende funcionalmente de NUM-VENDEDOR y COD-TIEND depende funcionalmente de NOM-TIENDA, es decir, existe una dependencia transitiva,

por lo tanto la relación, se puede descomponer en las siguientes proyecciones:

NUM-VENDEDOR	NOM-TIENDA	NOM-TIENDA	COD-TIEND
V1	AURRERA	AURRERA	03
V2	BLANCO	BLANCO	05
V3	GIGANTE	GIGANTE	02
V4	BLANCO		

Frecuentemente, sucede que, durante el proceso de reducción, una relación se puede descomponer en varias formas diferentes.

Cuando hay más de una posibilidad para descomponer una relación, el concepto de proyecciones independientes, ayuda a seleccionar la mejor descomposición.

Sea R una relación y sean R1 y R2 las proyecciones de R entonces R1 y R2 son independientes si y sólo si toda dependencia funcional, en R se puede deducir lógicamente de las de R1 y R2 y los atributos comunes a R1 y R2 forman una llave candidata al menos para una de las dos.

A continuación, se da a conocer la definición de las formas normales:

Una relación R, está en primera forma normal (1FN) si y sólo si todos los dominios contienen únicamente valores atómicos, es decir, los valores en los dominios no son compuestos.

Ejemplo :

Se tiene la siguiente relación :

NUM-VENDEDOR	NOM-TIENDA	CODIG-LIB	CANT-LIB
V1	AURRERA	c2	10
V1	AURRERA	c4	15
V2	BLANCO	c3	20
V3	GIGANTE	c1	20
V3	GIGANTE	c2	15
V3	GIGANTE	c3	20

LLAVE : NUM-VENDEDOR + CODIG-LIB.

Aquí, como todos los dominios son atómicos, entonces la relación está en primera forma normal.

Una relación R está en segunda forma normal (2FN) si y sólo si R está en primera forma normal y cada atributo no primo (el cual es un atributo que no forma parte de la llave primaria) es completamente dependiente de la llave primaria.

Ejemplo:

Se tiene la siguiente relación o tabla:

NUM-EMP	SUCURSAL	CLAVE-SUC
E1	REYES	01
E2	VIGA	02
E3	LAGUNILLA	05
E4	LAGUNILLA	05

LLAVE : NUM-EMP.

Aquí, todos los dominios son atómicos y los atributos no primos que son SUCURSAL y CLAVE-SUC, dependen funcionalmente en forma completa de la llave NUM-EMP, ya que no existe un subconjunto de la llave del cual también dependan, por lo tanto la relación, está en segunda forma normal.

Una relación R, está en tercera forma normal (3FN) si y sólo si está en segunda forma normal y todo atributo no primo depende de la llave primaria pero no transitivamente.

Ejemplo:

Se tiene la siguiente relación :

NUM-EMP	SUCURSAL
E1	REYES
E2	VIGA
E3	LAGUNILLA
E4	LAGUNILLA

LLAVE : NUM-EMP.

Aquí, el atributo no primo que es SUCURSAL, depende sólo de la llave primaria y lo hace no transitivamente, por lo tanto la relación, está en tercera forma normal.

Una relación R está en forma normal de Boyce/Codd (FNBC) si y sólo si cada determinante (el cual es un atributo, tal vez compuesto, del cual depende funcionalmente en forma completa algún otro atributo) es una llave candidata.

Ejemplo:

Se tiene la siguiente relación :

NUM-EMP	SUCURSAL
E1	REYES
E2	VIGA
E3	LAGUNILLA
E4	LAGUNILLA

LLAVE : NUM-EMP.

Aquí, el determinante de la relación es NUM-EMP, ya que de este atributo depende funcionalmente en forma completa el atributo SUCURSAL y como además es la única llave candidata, entonces la relación está en forma normal de Boyce/Codd.

Una relación R está en cuarta forma normal (4FN) si y sólo si (siempre que exista una dependencia multivaluada, digamos $X \twoheadrightarrow Y$) todos los atributos de R dependen funcionalmente de X, es decir, las únicas dependencias funcionales o dependencias multivaluadas en R son de la

forma $K \rightarrow X$, es decir la dependencia funcional de una llave candidata K a algún otro atributo X .

Ejemplo:

Se tiene la siguiente relación :

TEMA	AUTOR
ANATOMIA	CABALLERO
ANATOMIA	GAMALIEL
ANATOMIA	PRADO
FISIOLOGIA	RAMIREZ

Aquí, el atributo **AUTOR**, multidepende del atributo **TEMA**, por lo tanto la relación, está en cuarta forma normal.

Por último, después de dar a conocer lo anterior, lo que sigue es resumir el proceso para transformar una relación a un conjunto de relaciones que sean equivalentes a ella, pero mejores, ya que este proceso se hace para reducir la redundancia y evitar problemas con las operaciones de actualización.

Los pasos son los siguientes:

- 1) Eliminar las dependencias funcionales no completas, de la relación que está en primera forma normal, tomando sus proyecciones, para obtener un conjunto de relaciones en segunda forma normal.
- 2) Eliminar las dependencias transitivas, de las anteriores relaciones en segunda forma normal, tomando sus proyecciones, para obtener un conjunto de relaciones en tercera forma normal.
- 3) Eliminar las dependencias funcionales, donde el determinante no sea una llave candidata de las anteriores relaciones, en tercera forma normal, tomando sus proyecciones, para obtener un conjunto de relaciones en forma normal de Boyce/Codd.

4) Eliminar las dependencias multivaluadas, que no sean dependencias funcionales de las relaciones anteriores en forma normal Boyce/Codd, tomando sus proyecciones, para obtener un conjunto de relaciones en cuarta forma normal.

Ejemplo :

Se tiene la siguiente relación :

NUM-VENDEDOR	NOM-TIENDA	COD-TIEND	CODIG-LIB	CANT-LIB
V1	AURRERA	03	c2	10
V1	AURRERA	03	c4	15
V2	BLANCO	05	c3	20
V3	GIGANTE	02	c1	20
V3	GIGANTE	02	c2	15
V3	GIGANTE	02	c3	20
V4	BLANCO	05	c2	15

LLAVE : NUM-VENDEDOR + CODIG-LIB.

Aquí, como todos los dominios son atómicos, entonces la relación está en primera forma normal.

Pero se tienen las siguientes dependencias:

NOM-TIENDA depende funcionalmente de NUM-VENDEDOR.

COD-TIEND depende funcionalmente de NUM-VENDEDOR.

CANT-LIB depende funcionalmente de NUM-VENDEDOR + CODIG-LIB.

NOM-TIENDA depende funcionalmente de NUM-VENDEDOR + CODIG-LIB.

Por lo tanto como NOM-TIENDA, depende funcionalmente tanto de NUM-VENDEDOR como de NUM-VENDEDOR + CODIG-LIB y el primero es subconjunto del segundo, entonces NOM-TIENDA no depende funcionalmente en forma completa de NUM-VENDEDOR + CODIG-LIB por lo tanto se tienen que hacer las siguientes proyecciones:

P1

NUM-VENDEDOR	NOM-TIENDA	COD-TIEND
v1	AURRERA	03
v2	BLANCO	05
v3	GIGANTE	02
v4	BLANCO	05

P2

NUM-VENDEDOR	CODIG-LIB	CANT-LIB
v1	c2	10
v1	c4	15
v2	c3	20
v3	c1	20
v3	c2	15
v3	c3	20
v4	c2	15

Estas relaciones ya están en segunda forma normal, pero en la relación P1 existe una dependencia transitiva, ya que NOM-TIENDA depende funcionalmente de NUM-VENDEDOR y COD-TIEND depende funcionalmente de NOM-TIENDA, por lo tanto se tienen que hacer las siguientes proyecciones:

Q1

NUM-VENDEDOR	NOM-TIENDA
v1	AURRERA
v2	BLANCO
v3	GIGANTE
v4	BLANCO

Q2

NOM-TIENDA	COD-TIEND
AURRERA	03
BLANCO	05
GIGANTE	02

Estas relaciones están en tercera forma normal y como NUM-VENDEDOR es el determinante de Q1 y NOM-TIENDA es el determinante de Q2 debido a que de estos atributos dependen funcionalmente en forma completa los atributos NOM-TIENDA y COD-TIEND respectivamente y como, además, son las únicas llaves candidatas, entonces las relaciones Q1 y Q2 están en forma normal de Boyce/Codd.

CAPITULO 2

CAPITULO 2

METODOLOGIA PARA DISEÑO DE BASES DE DATOS RELACIONALES USANDO EL MODELO ENTIDAD-RELACION

INTRODUCCION

El propósito de este capítulo, es dar a conocer, una metodología para diseñar bases de datos relacionales, utilizando el modelo entidad-relación con constructores extendidos, para ello, el capítulo se conforma de la siguiente manera:

En la sección 2.1, se dará a conocer el panorama general de lo que es la metodología, en la sección 2.2 se describe, lo que es el modelo entidad-relación y sus constructores extendidos, en la sección 2.3 se desarrolla detalladamente el primer paso de la metodología, en la sección 2.4 se desarrolla detalladamente el segundo paso y por último en la sección 2.5 se toca un poco más a fondo, la normalización que corresponde al tercer paso de dicha metodología.

En este capítulo se da a conocer, el diseño de bases de datos relacionales, utilizando el modelo entidad-relación y el modelo relacional.

Para empezar, se puede mencionar que el modelo entidad-relación es una herramienta de comunicación entre el diseñador y el usuario, durante el análisis de requerimientos y las fases de diseño conceptual y que es más fácil de representar y de entender debido a que en la aproximación de arriba hacia abajo (que es una manera de diseñar una base de datos) se usa el concepto de abstracción, es decir, las entidades se utilizan como una abstracción de los datos y además, el enfoque en la relación entre entidades reduce el número de objetos considerados y simplifica el análisis. Ahora, se darán a conocer, los pasos de la metodología del diseño de base de datos relacional.

Los pasos son los siguientes:

PASO 1:

MODELADO DE REQUERIMIENTOS UTILIZANDO EL MODELO ENTIDAD-RELACION EXTENDIDO

Este paso consiste, en analizar y modelar los requerimientos de datos, utilizando un diagrama de entidad-relación extendido, el cual contiene semántica para relaciones opcionales, relaciones ternarias y subtipos o categorías y los cuales en el procesamiento de requerimientos se especifican utilizando expresiones de lenguaje natural y además las vistas que se obtienen de varias fuentes, se integran dentro de una vista global de la base de datos total.

PASO 2:

TRANSFORMACION DEL MODELO ENTIDAD-RELACION EXTENDIDO A RELACIONES

Este paso consiste, en que a cada relación y a sus entidades asociadas se les transforme a un conjunto de relaciones candidatas y esto se logra, basandose en una categorización de constructores del modelo entidad-relación extendido y en un conjunto de reglas de transformación.

Además, en este paso, se tratan de eliminar las relaciones redundantes.

PASO 3:

NORMALIZACION DE RELACIONES

En este paso, las relaciones candidatas asociadas, junto con todas las dependencias tanto funcionales (las cuales se derivan del diagrama de entidad-relación extendido y representan las dependencias, entre las llaves de las entidades) como multivaluadas (las cuales son las que representan, las dependencias entre atributos llaves y no llaves de las entidades, que se derivan de la especificación de requerimientos) derivadas, son normalizadas al grado más alto deseado, utilizando técnicas de normalización. También se analizan las redundancias para poder eliminarlas.

Además, se puede mencionar, que existe una metodología, que simplifica el diseño de bases de datos relacionales y el cual, reduce el número de dependencias de los datos, que se analizan y esto se hace introduciendo un paso de diseño conceptual, en el modelo relacional común, que consiste en capturar una representación exacta de la realidad, utilizando el modelo entidad-relación extendido, además, con la introducción de este paso, la integridad de los datos se preserva, debido a la normalización de las relaciones candidatas, que se forman en la transformación del modelo entidad-relación extendido.

Antes de dar a conocer, los pasos de la metodología, se empezará con la descripción, del modelo entidad-relación.

En este modelo, se representa la información, en términos de entidades, atributos y relaciones.

Inicialmente, en este modelo, se considerarán las siguientes 3 clases de objetos :

- 1) **ENTIDADES** *las cuales usualmente, denotan una persona, lugar, cosa o evento de interés informativo, o en términos generales, son aquellas que designan cualquier objeto distinguible, que pueda representarse en una base de datos.*

Las entidades, se pueden distinguir por la fuerza de su atributo de identificación y se pueden clasificar en 2 tipos:

- a) **ENTIDADES FUERTES** *las cuales, tienen identificadores, que sólo determinan la existencia de la entidad.*
- b) **ENTIDADES DEBILES** *las cuales, derivan su existencia de los atributos de identificación de otras entidades.*

Más específicamente, las entidades débiles difieren de las fuertes, solamente por el hecho, de que necesitan llaves de otras entidades, para establecer su singularidad, de otra forma las dos tienen las mismas propiedades de transformación (las cuales se mencionan en la sección 2.4) y no se necesitan reglas especiales.

Ejemplo:

Si alguien está interesado en tener información acerca de un profesor, entonces se puede crear la entidad **PROFESOR** que tenga como identificador al atributo denominado **R.F.C** y otros atributos tales como: **NOMBRE, ANTIGUEDAD, ETC.**

2) ATRIBUTOS *los cuales son propiedades descriptivas, que detallan las entidades.*

Estos se dividen en 2 tipos :

a) **IDENTIFICADORES** *los cuales son usados, para distinguir de manera única a las entidades.*

b) **DESCRIPTORES** *los cuales son usados, para describir las entidades.*

Ejemplo:

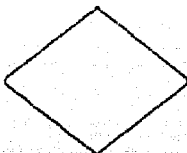
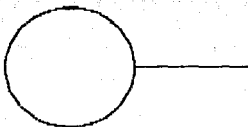
Si se quiere que una entidad se refiera a los estudiantes, pero que cada entidad sea identificada por el número de cuenta de cada estudiante y además esta se pueda describir con la fecha de nacimiento, nombre y sexo entonces se debe considerar la entidad **ESTUDIANTE**, cuyo atributo identificador sea **NUM-CUENTA** y atributos descriptores sean **FECHA-NACIMIENTO, NOMBRE, SEXO.**

3) RELACIONES *las cuales representan la asociación, en el mundo real, entre una o varias entidades.*

Además, estas tienen significado semántico, el cual se indica por la conexión entre las ocurrencias de una entidad (uno a uno, uno a muchos, y muchos a muchos) y debido a la participación en esta conexión los miembros de las entidades pueden ser opcionales u obligatorios.

Por ejemplo; la entidad **PERSONA**, en una relación que indique quienes son esposos, es opcional, ya que una persona puede o no puede tener un esposo o esposa.

Los diagramas para representar entidades, atributos y relaciones son los siguientes :

REPRESENTACION**CONCEPTO****ENTIDAD FUERTE****ENTIDAD DEBIL****RELACION****ATRIBUTO DESCRIPTOR****ATRIBUTO IDENTIFICADOR**

Este modelo que se usa como un esquema conceptual de representación a veces presenta dificultades, porque en el momento de integrar las vistas, se requiere del uso de conceptos de abstracción, tales como la generalización, además la integridad de los datos, involucra valores nulos en los atributos y también puede ocurrir que las relaciones de grado mayor a 2, en el momento de representarlas en su forma binaria sean incorrectas y debido a esto, se crea el modelo entidad-relación extendido, el cual proporciona una representación para estos conceptos.

En este modelo, se introducen categorías de abstracción dentro del modelo entidad-relación y esto se hace al agregar 2 tipos de objetos:

1) JERARQUIA DE SUBCONJUNTO *la cual especifica la superposición de subconjuntos.*

Una definición formal, es la siguiente:

Si E_1 y E_2 son entidades, entonces E_1 es un subconjunto de E_2 si cada ocurrencia de E_1 es también una ocurrencia de E_2 .

Es decir, una jerarquía de subconjunto, es cuando cada ocurrencia de una entidad puede también ser, una ocurrencia de otras entidades, que son subconjuntos de superposición.

Por ejemplo; la entidad **EMPLEADO**, puede tener las siguientes clasificaciones: **EMPLEADOS QUE SON EJECUTIVOS** y **EMPLEADOS QUE ATIENDEN VENTANILLAS**.

2) JERARQUIA DE GENERALIZACION *la cual especifica, la no superposición de subconjuntos.*

Una definición formal, es la siguiente:

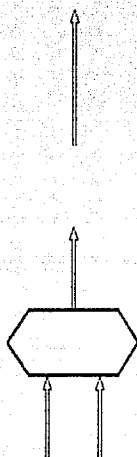
Si E, E_1, E_2, \dots, E_n son entidades entonces E es generalización de E_1, \dots, E_n si cada ocurrencia de E es también una ocurrencia de una y sólo una de las entidades E_1, E_2, \dots, E_n .

Es decir, una jerarquía de generalización, es cuando una entidad, es dividida en diferentes valores de un atributo común. Por ejemplo la entidad **EMPLEADO** es una generalización de **PROFESOR, ADMINISTRATIVO, INTENDENTE, TECNICO**.

Los diagramas de jerarquía de generalización y subconjunto son los siguientes:

REPRESENTACION

CONCEPTO



JERARQUIA DE
SUBCONJUNTO

JERARQUIA DE
GENERALIZACION

Para poder entender mejor los diagramas de entidad-relación extendido, se dará a conocer la clasificación de sus constructores.

Los constructores son los siguientes:

1) GRADO DE RELACION.

2) CONEXION DE UNA RELACION.

3) CLASE DE MIEMBROS EN UNA RELACION.

4) CLASE DE OBJETOS DE ENTIDADES Y RELACIONES.

1) Con respecto al grado de relación, se tiene que este indica el número de entidades asociadas en la relación.

Por ejemplo; una relación n -aria es de grado n .

En la figura 1 se muestra la representación de relaciones unarias, binarias y ternarias las cuales tienen grado 1, 2 y 3 respectivamente; así como un ejemplo de aplicación.

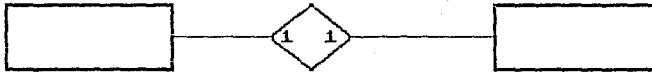
RELACION: UNARIA

GRADO: 1



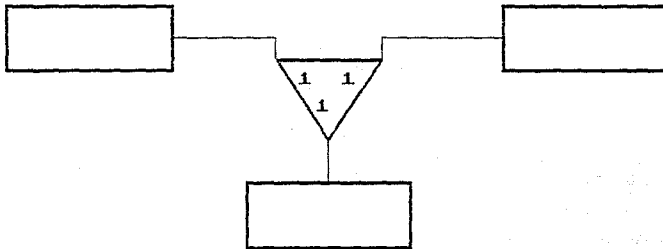
RELACION: BINARIA

GRADO: 2



RELACION: TERNARIA

GRADO: 3



EJEMPLOS :

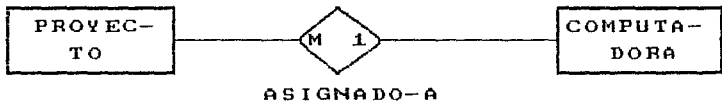
RELACION: UNARIA

RELACION QUE INDICA QUE EMPLEADO ESTA CASADO CON OTRO EMPLEADO DE LA MISMA COMPAÑIA



RELACION: BINARIA

RELACION QUE INDICA QUE PROYECTOS ESTAN ASIGNADOS A CADA COMPUTADORA



RELACION: TERNARIA

RELACION QUE INDICA LOS CONOCIMIENTOS QUE UTILIZA CADA PROFESOR PARA LOS PROYECTOS EN LOS QUE PARTICIPA.

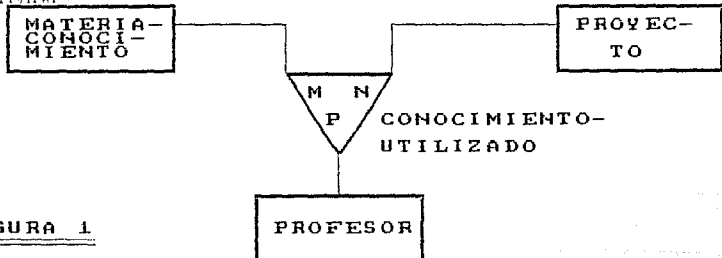


FIGURA 1

2) Con respecto a la conexión de una relación, se tiene que esta especifica las ocurrencias de una entidad en la relación.

Los valores para una conexión son uno o muchos.

La cardinalidad de la conexión, es el número asociado con el término *MUCHOS*.

En la siguiente figura, se muestran los constructores básicos para la conexión (uno a uno, uno a muchos y muchos a muchos) la esquina del rombo, en donde aparece una letra (M, N, P, etc) tanto en la relación unaria como la binaria, representa el lado *MUCHOS*, mientras que la esquina, en donde aparece un 1, representa el lado *UNO*.

Las relaciones n-arias para $n > 2$, están representadas por un polígono de n lados y cada esquina del polígono de n lados conecta una entidad además al igual que en las anteriores relaciones, cuando una esquina tiene una letra (M, N, P, etc) esta representa *MUCHOS* y una esquina con un 1 denota *UNO*.

CONEXION: 1:1 (UNO a UNO)



CONEXION: 1:M (UNO a MUCHOS)



CONEXION: M:N (MUCHOS a MUCHOS)



EJEMPLOS:

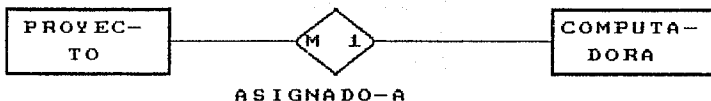
CONEXION:
1:1

RELACION QUE INDICA CUAL ES EL ENCARGADO, QUE COORDINA CADA LABORATORIO



CONEXION:
M:1

RELACION QUE INDICA TODOS LOS PROYECTOS ASIGNADOS A CADA COMPUTADORA



CONEXION:
M:N

RELACION QUE INDICA TODOS LOS ESTUDIANTES, QUE ESTAN INSCRITOS EN CADA UNA DE LAS MATERIAS



Una entidad en una relación ternaria, se considera con valor *UNO*, si sólo una ocurrencia de ella, puede asociarse con una ocurrencia de cada una de las otras 2 entidades asociadas; y se considera con valor *MUCHOS*, si más de una ocurrencia de ella, puede ser asociada con una ocurrencia de cada una de las otras dos entidades asociadas, por ejemplo en la relación *CONOCIMIENTO-UTILIZADO* en la figura 1 se asocian las entidades *MATERIA-CONOCIMIENTO*, *PROYECTO Y PROFESOR* cada entidad en este ejemplo se considera con valor *MUCHOS*, es decir aquí cada profesor con una materia de conocimiento dada, puede trabajar en muchos proyectos.

Con respecto a la relación ternaria, se puede mencionar, que esta no puede ser reducida a una relación binaria equivalente, si la relación está en 4FN por ejemplo *CONOCIMIENTO-UTILIZADO* en la figura 1 está en 4FN y no puede ser descompuesto esto se ve en la siguiente figura :

Relación:

CONOCIMIENTO-UTILIZADO

NUM-PROFESOR	MAT-CON	NUM-PROYECTO
20	03	42
20	04	42
20	10	23
20	02	23

Pero, sin embargo, *CONOCIMIENTO-DETERMINADO*, que tiene la misma, representación de entidad-relación, que *CONOCIMIENTO-UTILIZADO*, no estará en 4FN, si sucede, que todas las materias conocidas de un profesor, pueden ser utilizados en todos los proyectos trabajados por ese profesor. En tal caso la relación *CONOCIMIENTO-DETERMINADO*, puede ser descompuesta en 2 relaciones binarias, *MUCHOS a MUCHOS* entre *PROFESOR* y *MATERIA-CONOCIMIENTO* y *PROFESOR* y *PROYECTO*, cada una de estas 2 nuevas relaciones representan una relación en 4FN.

Esto se observa en la siguiente figura:

Relación:

CONOCIMIENTO-DETERMINADO

NUM-PROFESOR	MAT-CON	NUM-PROYECTO
20	03	42
20	10	42
20	03	23
20	10	23

Esta relación, se puede descomponer, en las siguientes dos relaciones:

Relación:

PROF-CON

NUM-PROFESOR	MAT-CON
20	10
20	03

Relación:

PROF-PROY

NUM-PROFESOR

20

20

NUM-PROYECTO

42

23

- 3) Con respecto a la clase de miembros en una relación, se puede mencionar que es aquella que especifica, si el lado *UNO* o *MUCHOS*, en una relación, es obligatoria u opcional.

Por ejemplo, si una ocurrencia de la entidad en el lado *UNO*, debe existir siempre, entonces la entidad es obligatoria, pero cuando una ocurrencia de esa entidad, no deba existir necesariamente, esta se considera como opcional y similarmente ocurre con el lado *MUCHOS* de una relación, ya que es obligatoria, si por lo menos una ocurrencia de la entidad, tiene que existir y es opcional de otra manera.

En los diagramas, la clase de miembros opcional, está definida, por una marca en la línea de conexión, entre la entidad y la relación y esto se muestra en la siguiente figura. La clase de miembros, está involucrada en la existencia de dependencia, en un sistema del mundo real, por ejemplo, si una entidad independiente o fuerte, está asociada con una entidad dependiente o débil, esta no puede ser opcional.

EJEMPLO:

CLASE DE MIEMBROS:
OBLIGATORIO

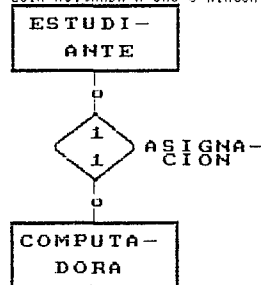


CLASE DE MIEMBROS:
OPCIONAL



CLASE DE MIEMBROS:
OPCIONAL

RELACION QUE INDICA SI UNA COMPUTADORA
ESTA ASIGNADA A UNO O NINGUN ESTUDIANTE



4) Con respecto a la clase de objetos de entidades y relaciones, se tiene que los objetos básicos, son las relaciones n -arias, con sus entidades asociadas y los objetos resultantes son:

La jerarquía de generalización, la cual implica, que los subconjuntos son una partición completa, de tal manera, que están desarticulados y si se combinan integran el conjunto completo.

La jerarquía de subconjunto, la cual implica, que los subconjuntos están superpuestos.

SECCION 2.3

DESARROLLO DEL PASO 1 DE LA METODOLOGIA

En esta sección, se desarrolla detalladamente el paso 1 de la metodología, que consiste en el:

MODELADO DE REQUERIMIENTOS UTILIZANDO EL MODELO ENTIDAD-RELACION EXTENDIDO

PASO 1.1:

Objetivos del análisis de requerimientos.

Los objetivos del análisis de requerimientos son:

- 1) Delinear los requerimientos de datos de la organización.
- 2) Describir la información acerca de los objetos y su asociación, con el fin de modelar esos requerimientos de datos.
- 3) Determinar los tipos de transacciones, que se piensan realizar en la base de datos. Para esto se usa el modelo entidad-relación extendido, el cual describe esos objetos y sus relaciones y además se utilizan expresiones de lenguaje natural, para describir las transacciones.

Una de las características del modelo entidad-relación extendido, es que incrementa la habilidad de los diseñadores, para captar correctamente los requerimientos de datos reales, ya que se necesita, un enfoque con detalle semántico, en la relación de los datos.

Es decir, la semántica de entidad-relación extendida, permite transformaciones directas de entidades y relaciones, en por lo menos una relación 1FN y especifica normas para coacción de integridad. También las técnicas de abstracción, tales, como la generalización, proporcionan herramientas, que sirven para la integración de las vistas de los usuarios, para definir, un esquema conceptual global.

PASO 1.2:

Normas para la clasificación de entidades y atributos.

Las normas para la clasificación de entidades y atributos son las siguientes:

1) Las entidades tienen información descriptiva y los atributos no la tienen. De esta manera, si hay información descriptiva acerca de un objeto, el objeto deberá ser clasificado, como una entidad y si sólo se necesita identificar un objeto, entonces el objeto, deberá ser clasificado como atributo.

Por ejemplo; si tenemos el objeto *ESTUDIANTE* y este tiene información descriptiva, tal como *NOMBRE*, *EDAD* y *CARRERA* entonces, *ESTUDIANTE* debe ser clasificado, como una entidad, pero si sólo se necesita el número de cuenta del estudiante para identificarlo, entonces *ESTUDIANTE*, debe ser clasificado, como un atributo.

2) Los atributos multivaluados, deben ser clasificados como entidades.

Por ejemplo; si varios valores de un atributo descriptor, corresponden a un valor de un atributo identificador, entonces el atributo descriptor, debe ser clasificado como una entidad, en lugar de como un atributo.

Por ejemplo; si tenemos que un banco, puede ser localizado en varias delegaciones y la entidad *BANCO* tiene como atributo identificador *NOMBRE-BANCO* y como atributo descriptor *DELEGACION*, entonces *DELEGACION*, puede ser clasificada como una entidad.

3) Un atributo, que tiene una relación *MUCHOS* a *UNO* con una entidad, debe convertirse en una entidad. Es decir, si un atributo descriptor en una entidad, tiene una relación *MUCHOS* a *UNO* con otra entidad, el atributo descriptor, deberá ser clasificado, como una entidad. Por ejemplo si se definen 2 entidades tales como:

BANCO (con atributo identificador *NUM-BANCO* y
atributos descriptores : *CIUDAD* y *GERENTE*)

ESTADO

debido, a que hay una relación *MUCHOS* a *UNO* entre *CIUDAD* y *ESTADO*, *CIUDAD* debe ser clasificado, como una entidad.

4) Los atributos deben ser asignados, a entidades, que describen mejor. Por ejemplo; si se quiere saber el número de teléfono del domicilio de cada estudiante entonces, el atributo *NUM-TELEFONO*, puede ser, un atributo de la entidad *DOMICILIO*, en lugar de la entidad *ESTUDIANTE*.

5) Los identificadores compuestos, deben ser evitados tanto como sea posible. Por ejemplo; si una entidad ha sido definida con un identificador compuesto (es decir, que el identificador este formado por dos o más atributos) y las componentes del identificador son todos identificadores de otras entidades, entonces, lo recomendable es eliminar esta entidad y al objeto definirlo después como una relación, pero si la entidad ha sido definida con un identificador compuesto, y las componentes del identificador no son identificadores de otras entidades, entonces hay dos opciones:

- 1) Eliminar la entidad y definir nuevas entidades con, las componentes del identificador compuesto, las cuales se convertirán en identificadores de cada una de las nuevas entidades y posteriormente, se definirá una relación, para representar este objeto.
- 2) Conservar la entidad con el identificador compuesto, si es razonablemente natural.

Por ejemplo; si la entidad *PUBLICACIONES* se define con un identificador compuesto *ESCRITOR* y *LIBRO*, entonces puede eliminarse la entidad *PUBLICACIONES* y definir otras dos nuevas entidades, que son: *ESCRITOR* y *LIBRO*, posteriormente se puede definir la relación entre *ESCRITOR* y *LIBRO* para representar el objeto *PUBLICACIONES*. De otra forma, por ejemplo, si la entidad *DOMICILIO* se define con un identificador compuesto *CALLE* y *NUMERO* entonces, en este caso, parece más adecuado conservar esta entidad, ya que la definición de una entidad *NUMERO* no es muy común.

PASO 1.3:

Identificación de la jerarquía de generalización y jerarquía de subconjunto.

Para identificar la jerarquía de generalización y jerarquía de subconjunto se tiene lo siguiente:

Si hay una jerarquía de generalización o subconjunto entre entidades, entonces, lo que se tiene que hacer, es reasignar atributos a las entidades relevantes.

Se deben poner identificadores y descriptores genéricos, en la entidad genérica y poner identificadores y descriptores específicos, en las entidades subconjunto, para entender mejor esto, supongamos que las siguientes entidades, fueron identificadas en el modelo entidad-relación extendido:

ENTIDAD	IDENTIFICADOR	DESCRIPTORES
EMPLEADO	NUM-EMPLEADO	NOM-EMPLEADO FECHA DE NACIMIENTO TITULO DE TRABAJO SALARIO CONOCIMIENTO
PROFESOR	NUM-EMPLEADO	NOM-EMPLEADO MATERIA
ADMINISTRATIVO	NUM-EMPLEADO	NOM-EMPLEADO FECHA DE NACIMIENTO SALARIO PUESTO
TECNICO	NUM-EMPLEADO	NOM-EMPLEADO CONOCIMIENTO EXPERIENCIA
INTENDENTE	NUM-EMPLEADO	NOM-EMPLEADO SALARIO ANTIGUEDAD

Aquí se observa, que *EMPLEADO* es una generalización de *PROFESOR*, *ADMINISTRATIVO*, *TECNICO* e *INTENDENTE*, por lo tanto, se tiene que hacer una reasignación de atributos a las entidades.

ENTIDAD	IDENTIFICADOR	DESCRIPTORES
EMPLEADO	NUM-EMPLEADO	NOM-EMPLEADO FECHA DE NACIMIENTO TITULO DE TRABAJO SALARIO
PROFESOR	NUM-EMPLEADO	MATERIA
ADMINISTRATIVO	NUM-EMPLEADO	PUESTO
TECNICO	NUM-EMPLEADO	CONOCIMIENTO EXPERIENCIA
INTENDENTE	NUM-EMPLEADO	ANTIGUEDAD

PASO 1.4:

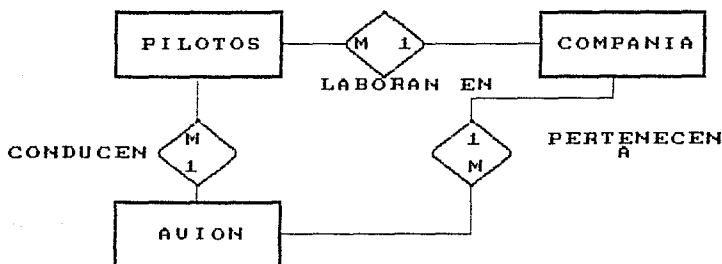
Normas para definir relaciones:

Con respecto a las relaciones se sabe, que se tienen que especificar : GRADO, CONEXION, CLASE DE MIEMBROS y ATRIBUTOS.

Algunas normas, que definen a las relaciones son :

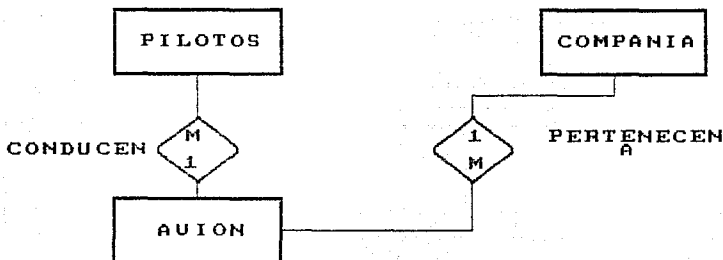
1) Las relaciones redundantes, deberán ser eliminadas, en donde por relaciones redundantes, se debe entender, que es cuando dos o más relaciones, son usadas para representar el mismo concepto y además se considera, que no hay redundancia, si dos o más relaciones son asignadas entre las mismas dos entidades pero las dos relaciones tienen diferente significado.

Además se debe mencionar, que un caso importante de redundancia, es la dependencia transitiva, que se muestra en el siguiente ejemplo:



En esta figura se puede ver que:

CONDUCEN, es una relación *MUCHOS a UNO* entre *PILOTOS* y *AVION*
PERTENECEN-A es una relación *MUCHOS a UNO* entre *AVION* y *COMPANIA*
LABORAN-EN es una relación *MUCHOS a UNO* entre *PILOTOS* y *COMPANIA* y
ambos *CONDUCEN* y *PERTENECEN-A* son obligatorias, por lo tanto
LABORAN-EN es redundante y debe ser eliminada quedando así:



2) Las relaciones ternarias, deben ser definidas cuidadosamente.

Una relación ternaria entre 3 entidades, debe ser definida, sólo cuando el concepto de asociación, no puede ser representado por varias relaciones binarias entre esas entidades.

Por ejemplo; si hay una asociación entre las entidades *PROFESOR*, *ALUMNO* y *PROYECTO* y si cada alumno, puede trabajar en muchos proyectos y además, puede trabajar bajo la instrucción de varios profesores para algunos de esos proyectos y si además cada profesor puede instruir a muchos alumnos en algún proyecto, entonces en este caso, se pueden definir 2 relaciones binarias, en lugar de una relación ternaria ver figura 2CA).

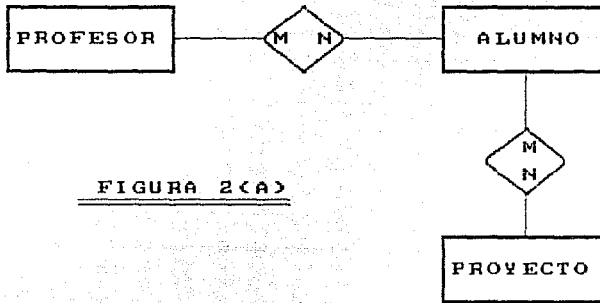


FIGURA 2(A)

Si cada alumno, puede trabajar en varios proyectos y además trabajar bajo la instrucción de varios profesores, pero si se tiene, que para cada proyecto, el alumno trabaja bajo la instrucción de exactamente un profesor, entonces debe ser definida la relación ternaria figura 2(B).

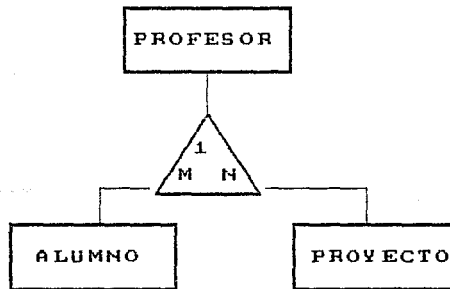


FIGURA 2(B)

Se debe mencionar, que en las relaciones ternarias, el significado de la conexión es importante, ya que por ejemplo en la figura 2(B), se muestra, que para un par de ocurrencias dadas de ALUMNO y PROYECTO, hay solamente una ocurrencia correspondiente de PROFESOR y sin

embargo, para un par de ocurrencias dadas de *PROFESOR* y *ALUMNO* puede haber muchas ocurrencias correspondientes de *PROYECTO*.

Frecuentemente, cuando el diseño es largo y más de una persona está involucrada en el análisis de requerimientos, surgen varias vistas de los datos y las relaciones, las cuales deben ser unidas, en una vista global para eliminar la redundancia e inconsistencia del modelo.

Para poder integrar las vistas, se requiere del uso, de herramientas semánticas del modelo entidad-relación extendido, tales como la identificación de sinónimos, reunión y generalización.

En esta sección, se desarrolla detalladamente, el paso dos de la metodología, que es la:

TRANSFORMACION DEL MODELO ENTIDAD-RELACION EXTENDIDO A RELACIONES

El objetivo de este paso, es dar a conocer las reglas, para transformar el modelo entidad-relación extendido, al modelo relacional.

Las transformaciones básicas, dan como resultado, tres tipos de relaciones:

1) La entidad-relación que contiene la misma información, que la entidad original. Esta transformación ocurre para las entidades, que están en relaciones binarias, que son *MUCHOS a MUCHOS*; para las entidades, que están en el lado *UNO*, de relaciones binarias *UNO a MUCHOS*; para una de las entidades, que forman parte de las relaciones binarias *UNO a UNO*; para las entidades de relaciones unarias que son *MUCHOS a MUCHOS*; y para las entidades de alguna relación ternaria o de grado mayor, jerarquía de generalización o jerarquía de subconjunto.

2) La entidad-relación con llave fija, que procede de la entidad padre. Esta transformación, ocurre para la entidad hijo en el lado *MUCHOS* de relaciones binarias, que son (padre) *UNO a MUCHOS* (hijo); para una de las entidades, en una relación binaria (padre) *UNO a UNO* (hijo) y para cada entidad en una relación unaria, que es *UNO a UNO* o *UNO a MUCHOS*.

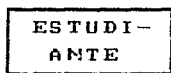
Estrategias fundamentales para la transformación son: mantener la relación padre-hijo, poniendo la llave de la entidad padre en la relación hijo, otra estrategia, está basada en la eficiencia: es decir, añadir la llave foránea, a la relación con menos n-adas.

3) La relación-r con las llaves foráneas de todas las entidades. Esta transformación, ocurre para relaciones, binarias y unarias que son *MUCHOS a MUCHOS*, y todas las relaciones, que son de grado mayor o igual tres.

Además se debe mencionar, que los valores nulos, sólo se permiten en llaves foráneas en una entidad-relación, si estas provienen de alguna entidad opcional, pero no son permitidos, en llaves foráneas, que provengan de alguna entidad obligatoria, ni en llaves foráneas, que vayan a formar parte de una relación-r, por lo tanto, las siguientes reglas servirán para manipular los valores nulos, en las transformaciones:

A) En las relaciones unarias, se tiene lo siguiente:

Una entidad completamente opcional o completamente obligatoria, en una relación *UNO a UNO*, debe transformarse en una entidad-relación y en ambos casos, en el obligatorio (figura 3a) y en el opcional (figura 3b) la llave de la entidad debe aparecer como llave foránea, en la entidad-relación. Además en ambos casos, los atributos llave, son tomados del mismo dominio pero con diferentes nombres, para poder lograr unicidad.



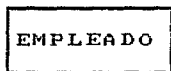
RELACION QUE INDICA QUE ESTUDIANTES
FORMAN PARTE DE CADA EQUIPO PARA
HACER LAS TAREAS
CADA ESTUDIANTE, TRABAJA EN EQUIPO
CON EXACTAMENTE OTRO ESTUDIANTE

RELACION:

ESTUDIANTE(NUM-CUENTA,...,
NUM-EQ-TAR)

FIGURA 3a

EQUIPO-TAR



RELACION QUE INDICA LOS EMPLEADOS
QUE CONTRAJERON MATRIMONIO CON
OTRO EMPLEADO DE LA MISMA COMPAÑIA
UN EMPLEADO PUEDE TENER A OTRO
EMPLEADO COMO SU ESPOSO O ESPOSA

RELACION:

EMPLEADO(NUM-EMP,...,NUM-EMP-ES)

FIGURA 3b

CASADO-CON

Una entidad en una relación *UNO a MUCHOS*, se debe transformar en una entidad-relación, que además debe contener la llave foránea que proviene de dicha entidad, esto debe hacerse para los 2 casos opcional (figura 3c) y obligatorio (figura 3d) pero sólo en el caso opcional se acepta asignación de valores nulos para la llave foránea.



FIGURA 3c

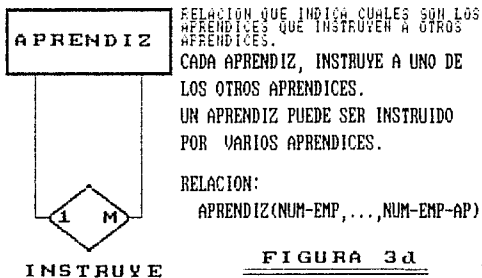


FIGURA 3d

Una relación *MUCHOS a MUCHOS*, debe transformarse en una relación-r, la cual debe estar formada por las llaves foráneas que provienen de la entidad, tanto en el caso opcional (figura 3e) como en el obligatorio y la entidad debe transformarse en una entidad-relación, con el mismo contenido de información que la entidad original.

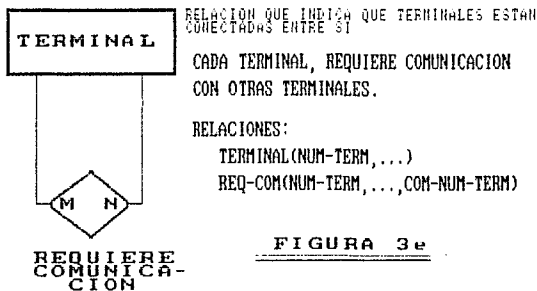
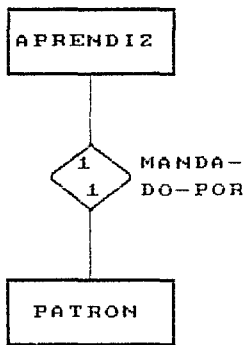


FIGURA 3e

B) En las relaciones binarias, se tiene lo siguiente:

En las relaciones *UNO a UNO*, cuando ambas entidades son obligatorias, (figura 4a) cada entidad debe convertirse en una entidad-relación y la llave de una de las entidades, debe aparecer como llave foránea en la entidad-relación que corresponde a la otra entidad, pero esta llave no debe tener valores nulos.



RELACION QUE INDICA CUAL ES EL PATRON QUE MANDA A CADA APRENDIZ.

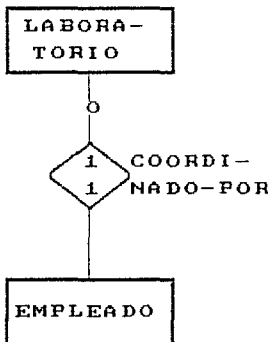
CADA APRENDIZ, TIENE UN PATRON Y CADA PATRON MANDA A UN APRENDIZ

RELACIONES:

APRENDIZ(NUM-EMPLEADO,...,NUM-EMP-PATRON)
 PATRON(NUM-EMP-PATRON,...)

FIGURA 4a

Si una de las entidades es opcional, (figura 4b) la entidad-relación que se obtiene de esta entidad es la que debe contener la llave foránea, de la otra entidad.



RELACION QUE INDICA QUE EMPLEADO COORDINA CADA LABORATORIO.

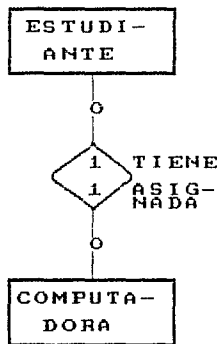
CADA LABORATORIO DEBE TENER UN COORDINADOR UN EMPLEADO PUEDE SER COORDINADOR DE A LO MUCHO UN LABORATORIO.

RELACIONES:

LABORATORIO(NUM-LAB,...,NUM-EMP)
 EMPLEADO(NUM-EMP,...)

FIGURA 4b

Si ambas entidades son opcionales, (figura 4c) una de las entidades-relación que proviene de una de las entidades, debe contener la llave foránea de la otra entidad, pero aquí si se permiten los valores nulos en la llave.



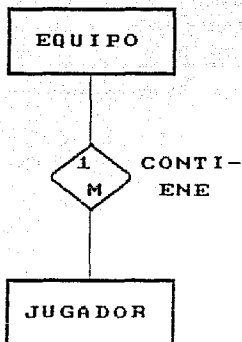
RELACION QUE INDICA QUE ESTUDIANTES TIENEN ASIGNADAS UNA COMPUTADORA
 ALGUNAS COMPUTADORAS ESTAN ASIGNADAS A LOS ESTUDIANTES. CADA COMPUTADORA ES ASIGNADA A EXACTAMENTE UN ESTUDIANTE PERO NO NECESARIAMENTE CADA ESTUDIANTE TIENE ASIGNADA UNA COMPUTADORA.

RELACIONES:

ESTUDIANTE(NUM-CUENTA, ..., NUM-COMP)
 COMPUTADORA(NUM-COMP, ...)

FIGURA 4c

En una relación *UNO a MUCHOS*, puede ocurrir que la entidad en el lado *MUCHOS* sea obligatoria u opcional y también puede ser obligatoria (figura 4d) u opcional (figura 4e) la entidad en el lado *UNO* y en todos los casos, la llave foránea, debe aparecer en la entidad-relación, que proviene de la entidad del lado *MUCHOS*, la cual, representa la entidad hijo y la llave foránea sólo puede tener valores nulos, si la entidad del lado *UNO* es opcional, además la entidad del lado *UNO* se debe convertir en una entidad-relación que contenga la misma información que ella.



RELACION QUE INDICA LOS JUGADORES QUE FORMAN CADA EQUIPO

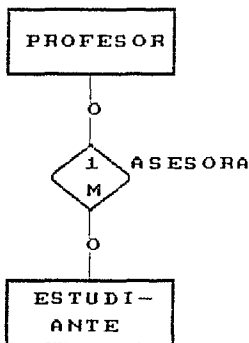
CADA JUGADOR PERTENECE A EXACTAMENTE UN EQUIPO
CADA EQUIPO CONTIENE MUCHOS JUGADORES.

RELACIONES:

EQUIPO(NUM-EQ,...)

JUGADOR(NUM-JUG,...,NUM-EQ)

FIGURA 4d



RELACION QUE INDICA LOS PROFESORES QUE ASESORAN A LOS ESTUDIANTES

UN PROFESOR PUEDE ASESORAR A VARIOS ESTUDIANTES Y CADA ESTUDIANTE PUEDE ESTAR ASESORADO POR A LO MAS UN PROFESOR.

RELACIONES:

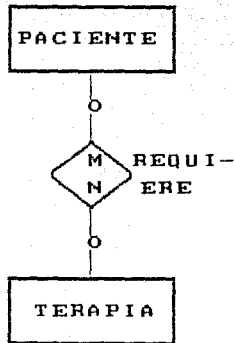
PROFESOR(EMP-NUM,...)

ESTUDIANTE(NUM-CUENTA,...,EMP-NUM)

FIGURA 4e

Una relación *MUCHOS a MUCHOS*, debe transformarse en una relación-r, la cual debe contener llaves primarias de ambas entidades, y esta transformación se hace tanto para el caso opcional (figura 4f) como para el obligatorio y las entidades se deben convertir en entidades-relación que contengan la misma información que cada entidad.

En este caso no es posible fijar las llaves foráneas, ya que la propiedad de *MUCHOS*, está en ambas direcciones.



RELACION QUE INDICA LAS TERAPIAS QUE REQUIERE CADA PACIENTE

CADA TERAPIA, PUEDE SER REQUERIDA POR VARIOS O NINGUN PACIENTE Y CADA PACIENTE PUEDE REQUERIR MUCHAS O NINGUNA TERAPIA.

RELACIONES:

TERAPIA(NUM-TEP,...)

PACIENTE(NUM-PAC,...)

REQUIERE(NUM-TEP,NUM-PAC)

FIGURA 4f

C) En las relaciones n-arias, se tiene lo siguiente:

Una relación n-aria, tiene $n+1$ variedades posibles de conexión las cuales son: una cuando todos los n lados tienen conexión *UNO*, otra es cuando $n-1$ lados tiene conexión *UNO* y un lado tiene conexión *MUCHOS*, otra es cuando $n-2$ lados tiene conexión *UNO* y dos lados tienen conexión *MUCHOS* y así, hasta que todos los lados tienen conexión *MUCHOS*.

Las cuatro posibles relaciones ternarias; son mostradas en las siguientes figuras. Todas son transformadas, para crear una relación-r, que contenga las llaves primarias de las n entidades, sin embargo, en cada caso, el significado de las llaves es diferente. Por ejemplo, cuando todos los lados de la relación tienen conexión *UNO* (figura 5a) la relación-r, consiste de tres posibles llaves candidatas distintas, esto es por el hecho de que la relación se describe, por tres dependencias funcionales.

Además aquí también se tiene que si la entidad en el lado *UNO*, es opcional entonces las llaves foráneas pueden tener valores nulos, lo cual no ocurre si se habla del caso obligatorio.

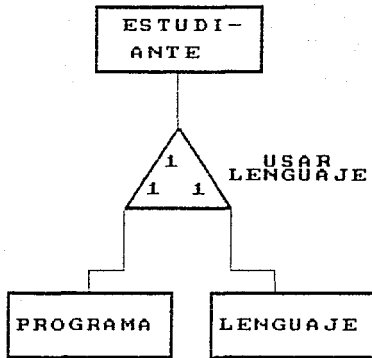


FIGURA 5a

RELACION QUE INDICA EL LENGUAJE UTILIZADO POR CADA ESTUDIANTE PARA CADA PROGRAMA
 UN ESTUDIANTE UTILIZA UN LENGUAJE PARA DETERMINADO PROGRAMA. DIFERENTES ESTUDIANTES UTILIZAN DIFERENTES LENGUAJES PARA EL MISMO PROGRAMA. DIFERENTES ESTUDIANTES PUEDEN UTILIZAR EL MISMO LENGUAJE PARA DIFERENTES PROGRAMAS.

RELACIONES:

ESTUDIANTE(NUM-CUENTA, ...)

PROGRAMA(NOM-PROGRAMA, ...)

LENGUAJE(NUM-LENG, ...)

USAR-LENGUAJE(NUM-CUENTA, NOM-PROGRAMA, NUM-LENG)

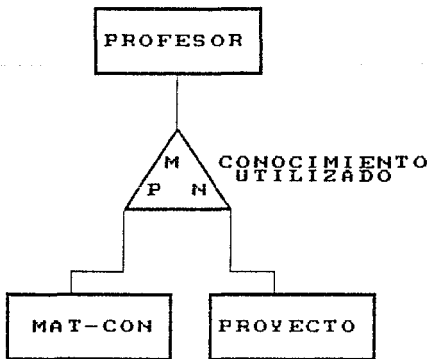
DEPENDENCIAS FUNCIONALES:

NUM-CUENTA, NOM-PROGRAMA → NUM-LENG

NUM-LENG, NOM-PROGRAMA → NUM-CUENTA

NUM-CUENTA, NUM-LENG → NOM-PROGRAMA

Quando todos los lados de la relación tienen conexión *MUCHOS*, (figura 5b) la relación-r, debe estar integrada por las llaves de cada una de las entidades, las cuales componen la única llave candidata, si es que la relación no tiene sus propios atributos. Además, todas las entidades se convierten en entidades-relación, con el mismo contenido de información que provienen de ellas.



RELACION QUE INDICA LAS MATERIAS DE CONOCIMIENTO QUE UTILIZA CADA PROFESOR PARA CADA PROYECTO

LOS PROFESORES UTILIZAN DIFERENTES MATERIAS DE CONOCIMIENTO PARA CADA PROYECTO.

RELACIONES:

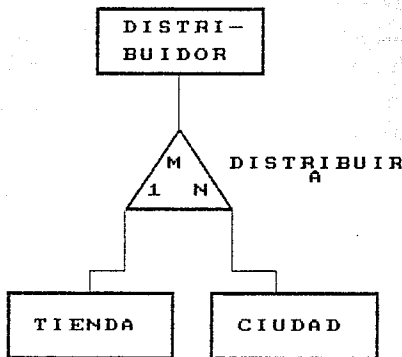
PROFESOR(NUM-PROFESOR, ...)

MAT-CON(NUM-MAT, ...)

PROYECTO(NOM-PROYECTO, ...)

CON-UTIL(NUM-PROFESOR, NUM-MAT, NOM-PROYECTO)

FIGURA 5b



RELACION QUE INDICA A QUE TIENDA DISTRIBUYE CADA DISTRIBUIDOR EN LAS CIUDADES

LOS DISTRIBUIDORES, DISTRIBUYEN A UNA O VARIAS TIENDAS, PERO SOLO PUEDEN DISTRIBUIR A LO MUCHO, A UNA TIENDA EN UNA CIUDAD DETERMINADA.

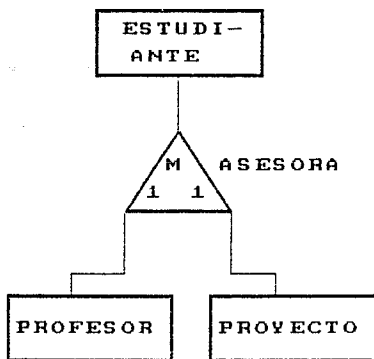
RELACIONES:

DISTRIBUIDOR(NUM-DIST,...)
 TIENDA(NUM-TIENDA,...)
 CIUDAD(NOM-CIUDAD,...)
 DIST-A(NUM-DIST,NOM-CIUDAD,NUM-TIENDA)

DEPENDENCIAS FUNCIONALES:

NUM-DIST,NOM-CIUDAD-->NUM-TIENDA

FIGURA 5c



RELACION QUE INDICA QUE PROFESOR ASESORA A CADA ESTUDIANTE EN CADA PROYECTO

LOS ESTUDIANTES TRABAJAN EN SUS PROYECTOS, BAJO EL ASESORAMIENTO DE PROFESORES. UN PROFESOR NO PUEDE ASESORAR A UN DETERMINADO ESTUDIANTE, EN MAS DE UN PROYECTO. UN ESTUDIANTE NO PUEDE TRABAJAR EN UN DETERMINADO PROYECTO BAJO EL ASESORAMIENTO DE VARIOS PROFESORES.

RELACIONES:

ESTUDIANTE(NUM-CUENTA,...)
 PROFESOR(NUM-PROF,...)
 PROYECTO(NOM-PROY,...)
 ASESORA(NUM-PROF,NUM-CUENTA,NOM-PROY)

DEPENDENCIAS FUNCIONALES:

NUM-CUENTA,NUM-PROF-->NOM-PROY
 NUM-CUENTA,NOM-PROY-->NUM-PROF

FIGURA 5d

D) En las jerarquías de generalización y subconjunto, se tiene lo siguiente:

En la jerarquía de generalización, (la cual es resultado de los subconjuntos desarticulados y que es producida, de la partición de la entidad genérica en diferentes valores de un atributo común) la entidad-relación que se obtiene de la entidad genérica, debe contener la llave de dicha entidad y todos los atributos comunes, incluyendo el atributo común usado para la partición, pero esto sólo se considera cuando el atributo para la partición existe actualmente y si no existe, entonces debe ser creado y cada entidad-relación de cada subconjunto, debe contener, la llave de la entidad genérica y los atributos específicos, para ese subconjunto, un ejemplo de esto se presenta en la figura 6.

Además si ocurre un cambio y este incluye la llave, entonces la entidad-relación del subconjunto así como la entidad-relación del conjunto, debe actualizarse, ya que el cambio para un atributo sin llave afecta, ya sea al conjunto o a una entidad-relación de subconjunto.

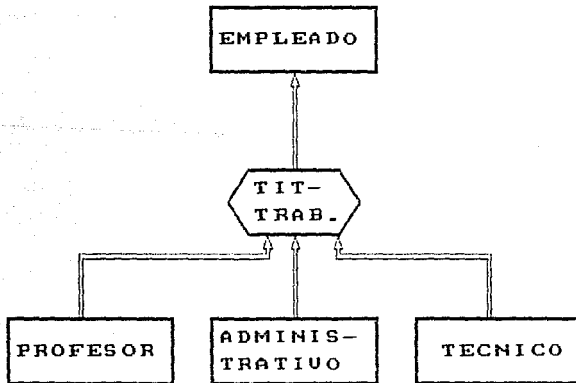


FIGURA 6

DIFERENTES TIPOS DE EMPLEADOS FORMAN UNA PARTICION DE VALORES DE UN ATRIBUTO COMUN QUE ES EL DE TIT-TRAB.

RELACIONES:

EMPLEADO(NUM-EMP, TIT-TRAB, ATRIBUTOS COMUNES)

PROFESOR(NUM-EMP, ATRIBUTOS ESPECIFICOS)

ADMINIS(NUM-EMP, ATRIBUTOS ESPECIFICOS)

TECNICO(NUM-EMP, ATRIBUTOS ESPECIFICOS)

{ JERARQUIA DE }
{ GENERALIZACION }

En la jerarquía de subconjunto, que es resultado de los subconjuntos de superposición, los cuales son producidos, por la partición de la entidad genérica en valores diferentes de atributos, (figura 7) la transformación, produce entidades-relación separadas, para la entidad genérica y cada una de las entidades de subconjunto. La llave de cada entidad-relación, es la llave de la entidad genérica y además la entidad-relación de la entidad genérica, contiene sólo atributos comunes y la entidad-relación de cada entidad de subconjunto, contiene atributos específicos para esa entidad de subconjunto. De esta manera, las reglas de transformación para los subconjuntos desarticulados y superpuestos son el mismo.

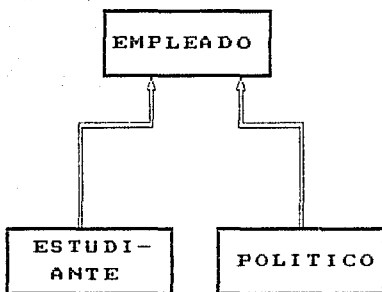


FIGURA 7

<JERARQUIA DE SUBCONJUNTO>

EMPLEADOS CON SITUACIONES ESPE-
CIALES, SON MOSTRADAS COMO SUB-
CONJUNTOS DE SUPERPOSICION, CON
BASE EN LA PARTICION DE VALORES
DE ATRIBUTOS DIFERENTES.

RELACIONES:

EMPLEADO(NUM-EMP, ATRIBUTOS
COMUNES)
ESTUDIANTE(NUM-EMP, ATRIBUTOS
ESPECIFICOS)
POLITICO(NUM-EMP, ATRIBUTOS
ESPECIFICOS)

La abstracción de reunión, puede ocurrir, entre las entidades o atributos que están relacionados en sólo una entidad.

La reunión entre las entidades, por ejemplo definidas en la relación *PARTE DE* es un caso especial de la colección de relaciones binarias *UNO a UNO* o *UNO a MUCHOS* y puede ser transformada como anteriormente se mencionó. Por ejemplo *CARRO*, puede representar la entidad entera, mientras *MOTOR*, *LLANTAS*, *FRENOS*, etc representan sus partes y cada parte es una entidad con sus propios atributos.

Por último se puede mencionar, que la introducción de una llave foránea, en una entidad-relación, no influye en el incremento de dependencias funcionales en una relación normalizada, si el diagrama de entidad-relación extendido es correcto, es decir si todos los atributos, son asociados con las entidades correctas.

Pero si el diagrama de entidad-relación extendido, no es correcto, entonces se pueden incrementar las dependencias funcionales y causar alguna desnormalización, pero después, tal transformación, puede normalizarse fácilmente, utilizando la técnica desarrollada en el capítulo anterior.

SECCION 2.5

DESARROLLO DEL PASO 3 DE LA METODOLOGIA

En esta sección, se detalla un poco más, el paso tres de la metodología que es:

LA NORMALIZACION.

Con respecto a la normalización, en el capítulo anterior ya se mencionó, que la normalización de relaciones candidatas, se realiza analizando, las dependencias funcionales y las dependencias multivaluadas asociadas con estas relaciones.

Se puede mencionar, que los pasos para la normalización, son los siguientes:

PASO 1:

Derivar las dependencias funcionales primarias del diagrama de entidad-relación extendido.

Se puede decir, que las dependencias funcionales, representan las dependencias, entre varias entidades y por otro lado, representan las dependencias entre los elementos de los datos, que comprenden una sola entidad.

En la siguiente tabla, se muestra el tipo de dependencias funcionales primarias, que se derivan de cada tipo de constructores de entidad-relación extendido, que se definieron en la sección 2.2.

GRADO	CONEXION	DEPENDENCIAS FUNCIONALES PRIMARIAS
UNARIA	1 A 1	LLAVE(UNO A)--->LLAVE(UNO B) LLAVE(UNO B)--->LLAVE(UNO A)
	1(OP) A 1(OP)	LLAVE(UNO A)--->LLAVE(UNO B) LLAVE(UNO B)--->LLAVE(UNO A)
	1(OP) A MUCHOS	LLAVE(MUCHOS)--->LLAVE(UNO)
	1 A MUCHOS	LLAVE(MUCHOS)--->LLAVE (UNO)
	MUCHOS A MUCHOS	LLAVE COMPUESTA--->O

GRADO	CONEXION	DEPENDENCIAS FUNCIONALES PRIMARIAS
BINARIA	1 A 1	LLAVE(UNO A)--->LLAVE(UNO B) LLAVE(UNO B)--->LLAVE(UNO A)
	1 A 1(OP)	LLAVE(UNO A)--->LLAVE(UNO B) LLAVE(UNO B)--->LLAVE(UNO A)
	1(OP) A 1(OP)	LLAVE(UNO A)--->LLAVE(UNO B) LLAVE(UNO B)--->LLAVE(UNO A)
	1 A MUCHOS	LLAVE(MUCHOS)--->LLAVE(UNO)
	1(OP) A MUCHOS	LLAVE(MUCHOS)--->LLAVE(UNO)
	MUCHOS A MUCHOS	LLAVE COMPUESTA--->O
TERNARIA	1 A 1 A 1	LLAVE(A).LLAVE(B)--->LLAVE(C) LLAVE(A).LLAVE(C)--->LLAVE(B) LLAVE(B).LLAVE(C)--->LLAVE(A)
	1 A 1 A MUCHOS	LLAVE(UNO A).LLAVE(MUCHOS)--> LLAVE(UNO B) LLAVE(UNO B).LLAVE(MUCHOS)--> LLAVE(UNO A)
	1 A MUCHOS A MUCHOS	LLAVE(MUCHOS A).LLAVE(MUCHOS B)---> LLAVE(UNO)
	MUCHOS A MUCHOS A MUCHOS	LLAVE COMPUESTA--->O
JERARQUIA DE GENERALIZACION		DEPENDENCIA FUNCIONAL SECUNDARIA
JERARQUIA DE SUBCONJUNTO		DEPENDENCIA FUNCIONAL SECUNDARIA

Basandose en las transformaciones de esta tabla, se resumen los tipos básicos de dependencias funcionales primarias derivables, de los constructores del modelo entidad-relación extendido y que son los siguientes:

- 1) LLAVE(LADO MUCHOS)--->LLAVE(LADO UNO)
- 2) LLAVE(LADO UNO A)--->LLAVE(LADO UNO B)
- 3) LLAVE(LADO MUCHOS A).LLAVE(LADO MUCHOS B)--->
LLAVE(LADO UNO)
- 4) LLAVE(LADO UNO A).LLAVE(LADO MUCHOS)--->
LLAVE(LADO UNO B)
- 5) LLAVE(LADO UNO A).LLAVE(LADO UNO B)--->
LLAVE(LADO UNO C)
- 6) LLAVE COMPUESTA--->0

Los tipos 1 y 2 representan una llave fija foránea, funcionalmente determinada, por la llave primaria, en una relación unaria o binaria. Los tipos 3 a 5, se aplican solamente, a relaciones ternarias y el tipo 6 se aplica, a todas las relaciones de cualquier grado, en el cual la relación está representada, como una llave total.

PASO 2:

Examinar todas las relaciones candidatas para obtener las dependencias funcionales secundarias y multivaluadas.

Cada relación candidata, debe ser examinada, para determinar, qué dependencias existen entre las llaves primarias, las llaves foráneas y los atributos, que no son llave.

Si los constructores de entidad-relación extendido, no incluyen atributos, que no son llaves, la especificación de los requerimientos de datos, deberá ser consultada.

PASO 3:

Normalizar todas las relaciones candidatas y eliminar las redundancias.

Cada relación candidata, ahora, quizás tenga alguna dependencia funcional primaria, dependencia funcional secundaria y dependencia multivaluada asociada con ella.

Estas dependencias, determinan el grado de normalización de la relación. Y para incrementar el grado de normalización, ahora se puede aplicar, la técnica desarrollada en el capítulo uno, a cada una de las relaciones.

El hecho de eliminar la redundancia de datos, es para minimizar el espacio de almacenamiento y para no sacrificar la integridad de los datos.

Con el fin de eliminar relaciones, se dará a conocer la definición, de lo que es una relación subtotal ya que si una relación B, es subtotal de otra relación A, entonces B puede ser eliminada:

Si A, B son relaciones, entonces se dice que B es subtotal de A, cuando todos los atributos de B, están también contenidos en A y todas las dependencias de datos en B, también suceden en A.

También se tiene, que una relación, es subtotal de otra, si es la unión de otras dos relaciones.

Tenemos como caso trivial, que alguna relación, que contiene solamente una llave compuesta y no contiene atributos, que no son llave, esta es automáticamente subtotal de alguna otra relación, que contiene los mismos atributos llave, porque la llave compuesta, es la más débil forma de dependencia de datos, pero si, sin embargo, las relaciones A y B representan casos genérico y específico respectivamente de entidades de la abstracción de la jerarquía de generalización o subconjunto y A es subtotal de B porque B no tiene atributos específicos adicionales, entonces se debe analizar, la información adicional para decidir si se elimina o no a B.

Cuando la relación subtotal, es debida a la unión de otras 2 relaciones, la eliminación de una relación subtotal, puede traer la consecuencia, de no poder recuperar la eficiencia, aunque el almacenamiento y los costos de actualización disminuyan y entonces, este cambio se debe analizar, en el diseño físico, respecto al procesamiento de requerimientos, para poder determinar, si es conveniente la eliminación de la relación subtotal.

En general, se puede mencionar, que las relaciones candidatas, son buenos estimadores del esquema final y normalmente requieren, muy poco refinamiento.

CAPITULO 3

CAPITULO 3

APLICACION DE LA METODOLOGIA

En este capítulo, se presenta un problema real, al cual se le aplica la metodología descrita en el capítulo anterior.

PASO 1:

MODELADO DE REQUERIMIENTOS UTILIZANDO EL MODELO ENTIDAD-RELACION EXTENDIDO.

El problema es el siguiente:

PROBLEMA:

En la unidad de posgrado de la Facultad de Ciencias, se desea construir una base de datos, con el fin de que se tenga toda la información, acerca de la trayectoria académica de cada profesor, para ello, se desea conocer lo siguiente:

- *D Materias impartidas desde 1987.
Se debe indicar la materia y el período en que se impartió.
- *D Jurados en los que ha participado desde 1987.
Se debe indicar, de quién fue jurado, en qué fecha fue el examen, qué tipo de examen se realizó y para qué tipo de grado se presentó dicho examen.
- *D Tesis que ha dirigido desde 1987.
Se debe indicar a quién dirigió la tesis, en qué fecha fue el examen y para qué tipo de grado se presentó dicho examen.
- *D Area en la que labora.
- *D Area en la que desea impartir clases y/o dirigir tesis y/o asesorar.
Además, se debe indicar, en qué nivel desea impartirla.

- *D) Domicilio y Teléfono.
- *D) Nombre, R.F.C.
- *D) Lugar de trabajo y Teléfono.
- *D) Grados obtenidos, indicando cuándo y en qué institución se obtuvieron.

También se quiere tener, la siguiente información acerca de los alumnos:

- *D) Materias en las que se ha inscrito, cuándo las tomó, con quién y si las acreditó.
- *D) Quienes han sido sus directores de tesis.
- *D) Quienes han sido sus sinodales.
Además, se debe indicar, en qué fecha se realizó el examen, qué tipo de examen hizo y para qué tipo de grado se presentó dicho examen.
- *D) El número de cursos, que llevó en licenciatura correspondientes a cada área.
- *D) Nombre, número de cuenta, fecha de nacimiento.
- *D) Domicilio y Teléfono.

Con respecto a los términos:

TIPO DE EXAMEN, se está haciendo referencia a si el grado se obtuvo con examen general o con tesis.

TIPO DE GRADO se está haciendo referencia a si el examen que se presentó fue para obtener el grado de Licenciatura, Maestría o Doctorado.

NIVEL se está haciendo referencia a si el nivel en el que se está disponible para impartir clases o dirigir tesis o asesorar es Licenciatura, Maestría o Doctorado.

Además, esta información se debe regir bajo los siguientes supuestos:

- 1) Un profesor labora en exactamente, un trabajo.
- 2) En cada trabajo, pueden laborar muchos profesores.
- 3) Cada profesor o estudiante, habita en exactamente un domicilio.
- 4) Un domicilio, puede ser habitado por uno o muchos profesores o estudiantes.
- 5) Un estudiante, lleva cursos de muchas áreas.
- 6) Muchos estudiantes, llevan cursos de cada área.
- 7) Un profesor tiene una o muchas áreas de trabajo.
- 8) En cada área, trabajan muchos profesores.
- 9) Un profesor puede estar disponible en muchas áreas o ninguna.
- 10) En cada área, hay uno o muchos profesores disponibles.
- 11) Un profesor imparte una o muchas materias.
- 12) Una materia es impartida por muchos profesores.
- 13) Un profesor obtiene un grado en, exactamente, una institución.
- 14) Un profesor pudo obtener distintos grados en la misma institución.
- 15) Dada una institución, se puede obtener un determinado grado, por muchos profesores.
- 16) Un estudiante se inscribe en una materia, con uno o varios profesores.
- 17) Un profesor inscribe a muchos estudiantes en una materia.
- 18) Un estudiante puede estar inscrito con un profesor en varias materias.
- 19) Un profesor puede dirigir a uno o muchos estudiantes, para obtener un grado determinado.

- 20) Un estudiante obtiene un grado, dirigido por exactamente un profesor.
- 21) Un estudiante puede ser dirigido por el mismo profesor, para obtener varios grados.
- 22) Un profesor puede ser jurado de varios estudiantes, para obtener un grado.
- 23) Un estudiante para obtener un grado, necesita que varios profesores sean jurado.
- 24) Un estudiante puede tener de jurado, al mismo profesor, para obtener distintos grados.

De acuerdo con la metodología, lo primero, que se tiene que hacer, es identificar las entidades y los atributos.

Las entidades y atributos, que se obtuvieron de este problema son:

- *D PROFESOR cuyos atributos son:
R.F.C. NOMBRE.
- *D ESTUDIANTE cuyos atributos son:
NUM-CUENTA. NOMBRE. FECHA-NACIMIENTO.
- *D AREA cuyos atributos son:
NUMERO-AREA. NOMBRE-AREA.
- *D LUGAR-TRABAJO cuyos atributos son:
CLAVE-LUGAR. LUGAR. TELEFONO.
- *D INSTITUCION cuyos atributos son:
CLAVE-INSTIT. NOMBRE-INSTIT.
- *D DOMICILIO cuyos atributos son:
CALLE. NUMERO. COLONIA. TELEFONO.
- *D EXAMEN cuyos atributos son:
CLAVE-EXAMEN. FECHA-EXAMEN. TIPO-EXAMEN. TIPO-GRADO.

El siguiente paso, es identificar las jerarquías de subconjunto y generalización.

En este problema, se encuentra una jerarquía de subconjunto, ya que se identifican las siguientes entidades:

*D) **MATERIA** cuyos atributos son:

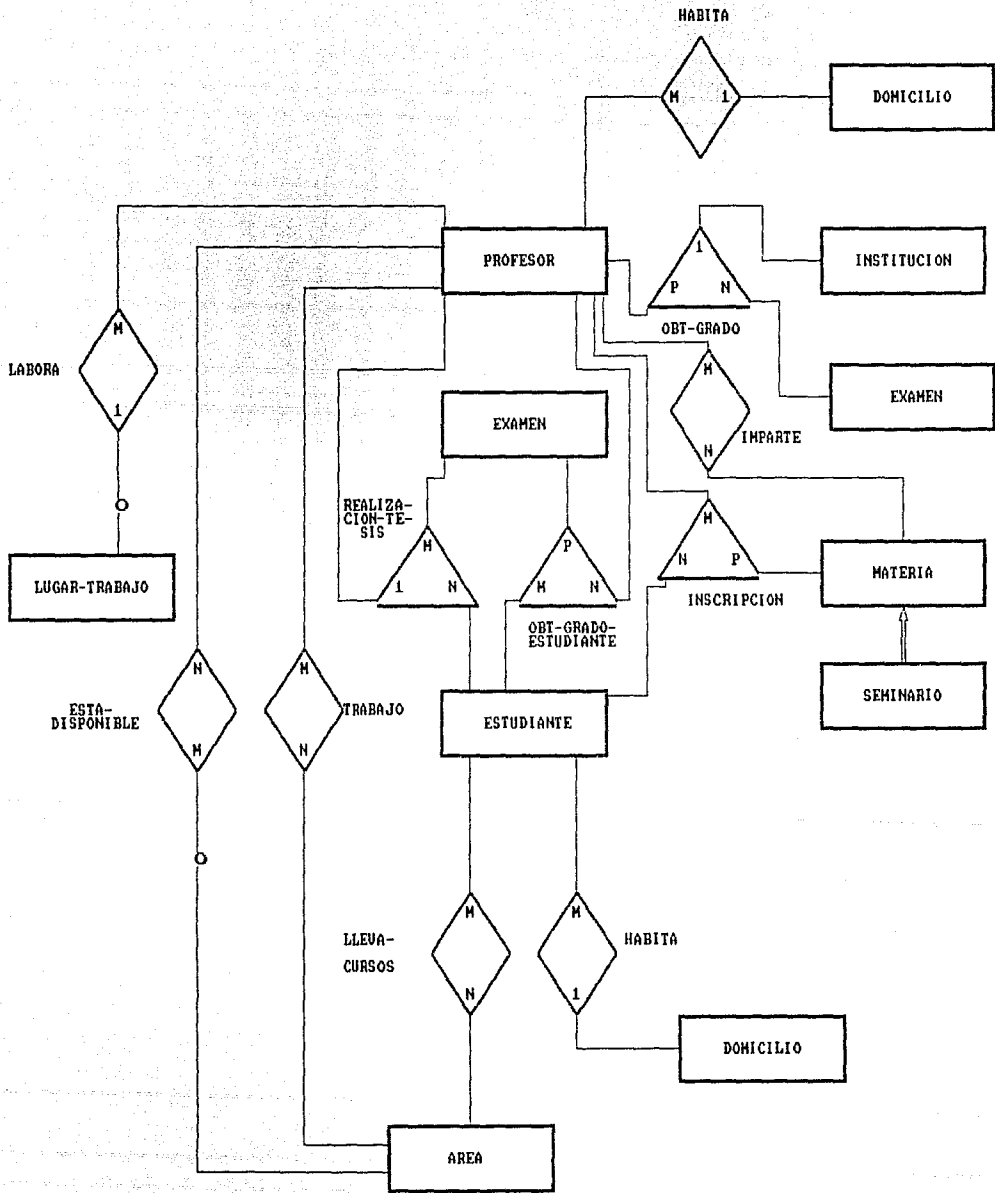
CLAVE-MATERIA, NOMBRE-MATERIA.

*D) **SEMINARIO** cuyos atributos son:

CLAVE-MATERIA, NOMBRE-MATERIA, SUBTITULO, PERIODO.

De las cuales se observa, que la entidad **SEMINARIO** es un subconjunto de la entidad **MATERIA**, por lo tanto, se tiene que hacer una reasignación de atributos, es decir se tienen que poner atributos identificadores y descriptores genéricos, en la entidad genérica y poner atributos identificadores y descriptores específicos, en la entidad subconjunto, es decir, la entidad **MATERIA**, que es la entidad genérica, tendrá como atributo identificador a **CLAVE-MATERIA** y atributo descriptor genérico a **NOMBRE-MATERIA** y la entidad **SEMINARIO**, que es la entidad subconjunto, tendrá como atributo identificador a **CLAVE-MATERIA** y atributos descriptores específicos a **SUBTITULO** y **PERIODO**.

El siguiente paso, es definir las relaciones entre las entidades, las cuales, se muestran en el siguiente diagrama de entidad-relación y posteriormente se explica, porque se proponen dichas relaciones.



Se definen las siguientes relaciones binarias:

LABORA, una relación *UNO a MUCHOS*, opcional, que asocia a las entidades LUGAR-TRABAJO y PROFESOR, debido a los supuestos 1 y 2.

HABITA, una relación *UNO a MUCHOS*, que asocia a las entidades DOMICILIO y PROFESOR o ESTUDIANTE, debido a los supuestos 3 y 4.

LLEVA-CURSOS, una relación *MUCHOS a MUCHOS*, que asocia a las entidades ESTUDIANTE y AREA, debido a los supuestos 5 y 6.

TRABAJO, una relación *MUCHOS a MUCHOS*, que asocia a las entidades PROFESOR y AREA, debido a los supuestos 7 y 8.

ESTA-DISPONIBLE, una relación *MUCHOS a MUCHOS*, opcional, que asocia a las entidades PROFESOR y AREA, debido a los supuestos 9 y 10.

IMPARTE, una relación *MUCHOS a MUCHOS*, que asocia a las entidades PROFESOR y MATERIA, debido a los supuestos 11 y 12.

Se definen las siguientes relaciones ternarias:

OBT-GRADO, una relación ternaria *MUCHOS a MUCHOS a UNO*, que asocia a las entidades PROFESOR, EXAMEN E INSTITUCION, debido a los supuestos 13, 14 y 15.

INSCRIPCION, una relación *MUCHOS a MUCHOS a MUCHOS*, que asocia a las entidades PROFESOR, ESTUDIANTE y MATERIA, debido a los supuestos 16, 17 y 18.

REALIZACION-TESIS. una relación *UNO a MUCHOS a MUCHOS*, que asocia a las entidades **PROFESOR, ESTUDIANTE y EXAMEN**, debido a los supuestos **19, 20 y 21**.

OBT-GRADO-ESTUDIANTE. una relación *MUCHOS a MUCHOS a MUCHOS*, que asocia a las entidades **PROFESOR, ESTUDIANTE y EXAMEN**, debido a los supuestos **22, 23 y 24**.

PASO 2:

TRANSFORMACION DEL MODELO ENTIDAD-RELACION EXTENDIDO A RELACIONES.

El siguiente paso, es transformar las entidades y relaciones del modelo entidad-relación, en relaciones del modelo relacional.

Las entidades:

AREA, MATERIA, LUGAR-TRABAJO, DOMICILIO, EXAMEN, E INSTITUCION se transforman en relaciones, con la misma información que la entidad, debido a lo siguiente:

AREA, forma parte de relaciones binarias *MUCHOS a MUCHOS*.

MATERIA, forma parte de una relación binaria *MUCHOS a MUCHOS* y también forma parte de una relación ternaria.

LUGAR-TRABAJO, está en el lado uno de una relación binaria *UNO a MUCHOS*.

DOMICILIO, está en el lado uno de una relación binaria *UNO a MUCHOS*.

INSTITUCION, forma parte de una relación ternaria.

EXAMEN, forma parte de relaciones ternarias.

Es decir, las relaciones son las siguientes:

- A) AREA(NUMERO-AREA, NOMBRE-AREA).
- B) MATERIA(CLAVE-MATERIA, NOMBRE-MATERIA).
- C) LUGAR-TRABAJO(CLAVE-LUGAR, LUGAR, TELEFONO).
- D) DOMICILIO(CALLE, NUMERO, COLONIA, TELEFONO).
- E) INSTITUCION(CLAVE-INSTIT, NOMBRE-INSTIT).
- F) EXAMEN(CLAVE-EXAMEN, FECHA-EXAMEN, TIPO-EXAMEN, TIPO-GRADO).

Las entidades:

SEMINARIO se transforma en una relación, que estará compuesta por la llave, de la entidad genérica y sus atributos específicos, es decir, la relación está compuesta por **CLAVE-MATERIA**, que es la llave de la entidad genérica, la cual es **MATERIA** y los atributos específicos, **SUBTITULO** y **PERIODO**, es decir, la relación que se obtiene es la siguiente:

G) **SEMINARIO**(**CLAVE-MATERIA**, **SUBTITULO**, **PERIODO**).

ESTUDIANTE se transforma en una relación, pero agregándole la llave de la entidad **DOMICILIO**, ya que está en una relación muchos a uno, por lo tanto, la relación que se obtiene es la siguiente:

H) **ESTUDIANTE**(**NUM-CUENTA**, **NOMBRE**, **FECHA-NACIMIENTO**, **CALLE**, **NUMERO**).

PROFESOR se transforma en una relación, pero agregándole la llave de la entidad **LUGAR-TRABAJO**, ya que está en una relación muchos a uno, que es la relación **LABORA**, además se le tiene que agregar la llave de la entidad **DOMICILIO**, ya que está en otra relación muchos a uno, que es la relación **HABITA**, por lo tanto, la relación que se obtiene es la siguiente:

I) **PROFESOR**(**R.F.C.**, **NOMBRE**, **CLAVE-LUGAR**, **CALLE**, **NUMERO**).

Las relaciones binarias **MUCHOS a MUCHOS**, es decir **ESTA-DISPONIBLE**, **IMPARTE**, **LLEVA-CURSOS** y **TRABAJO**, así como las relaciones ternarias **OBT-GRADO**, **REALIZACION-TESIS**, **OBT-GRADO-ESTUDIANTE** e **INSCRIPCION**, se transforman en relaciones, las cuales se compondrán de las llaves de las entidades que están relacionadas y de sus propios atributos.

Las relaciones son las siguientes:

- J> OBT-GRADO(R.F.C. CLAVE-INSTIT, CLAVE-EXAMEN).
- K> REALIZACION-TESIS(R.F.C. NUM-CUENTA, CLAVE-EXAMEN).
- L> OBT-GRADO-ESTUDIANTE(R.F.C. NUM-CUENTA, CLAVE-EXAMEN).
- M> INSCRIPCION(R.F.C. CLAVE-MATERIA, NUM-CUENTA, PERIODO, ACREDITACION).
- N> ESTA-DISPONIBLE(R.F.C. NUMERO-AREA, NIVEL, TIPO-DISP).
- O> IMPARTE(R.F.C. CLAVE-MATERIA, PERIODO).
- P> LLEVA-CURSOS(NUMERO-AREA, NUM-CUENTA, CANTIDAD).
- Q> TRABAJO(NUMERO-AREA, R.F.C).

PASO 3:

NORMALIZACION DE RELACIONES.

El último paso a desarrollar, es la normalización de relaciones. Primero, se derivan las dependencias funcionales primarias, del diagrama de entidad-relación, esto se puede hacer basandose en la tabla, de tipos básicos de dependencias funcionales primarias, que se mostró en la sección 2.5.

Las dependencias funcionales primarias son:

R.F.C ---> CALLE, NUMERO (la cual se obtuvo de la relación I)
NUM-CUENTA ---> CALLE, NUMERO (la cual se obtuvo de la relación H)
R.F.C, CLAVE-EXAMEN ---> CLAVE-INSTIT (la cual se obtuvo de la relación J)
R.F.C ---> CLAVE-LUGAR (la cual se obtuvo de la relación I)
R.F.C, NUM-CUENTA, CLAVE-EXAMEN ---> O (la cual se obtuvo de la relación L)
NUM-CUENTA, CLAVE-EXAMEN ---> R.F.C (la cual se obtuvo de la relación K)
R.F.C, NUM-CUENTA, CLAVE-MATERIA ---> O (la cual se obtuvo de la relación M)
R.F.C, NUMERO-AREA ---> O (la cual se obtuvo de la relación Q)
R.F.C, NUMERO-AREA ---> O (la cual se obtuvo de la relación N)
R.F.C, CLAVE-MATERIA ---> O (la cual se obtuvo de la relación O)
NUM-CUENTA, NUMERO-AREA ---> O (la cual se obtuvo de la relación P)

El siguiente paso, es derivar las dependencias funcionales secundarias, de las relaciones candidatas.

Las dependencias funcionales secundarias son:

R.F.C ---> NOMBRE (la cual se obtuvo de la relación I)
NUM-CUENTA ---> NOMBRE (la cual se obtuvo de la relación H)

CLAVE-MATERIA ---> NOMBRE-MATERIA (la cual se obtuvo de la
relación B)

NUMERO-AREA ---> NOMBRE-AREA (la cual se obtuvo de la relación
A)

CLAVE-LUGAR ---> LUGAR (la cual se obtuvo de la relación C)

LUGAR ---> TELEFONO (la cual se obtuvo de la relación C)

CLAVE-INSTIT ---> NOMBRE-INSTIT (la cual se obtuvo de la
relación E)

CALLE, NUMERO ---> TELEFONO (la cual se obtuvo de la relación
D)

CALLE, NUMERO ---> COLONIA (la cual se obtuvo de la relación
D)

CALLE ---> COLONIA (la cual se obtuvo de la relación D)

El último paso por realizar, es normalizar las relaciones.

En las dependencias listadas anteriormente, se observa qué hay una dependencia transitiva qué es:

CLAVE-LUGAR ---> LUGAR ---> TELEFONO, por lo tanto, se tiene qué hacer, la proyección de la relación LUGAR-TRABAJO, obteniéndose las siguientes relaciones:

LUGAR-TRABAJO(CLAVE-LUGAR, LUGAR)
TELEFONO-TRABAJO(LUGAR, TELEFONO)

También se observa, que existe una dependencia funcional no completa en la relación DOMICILIO, ya que COLONIA, depende funcionalmente de CALLE, NUMERO y además, depende también de un subconjunto de CALLE, NUMERO que es CALLE, por lo tanto, se tiene que hacer la proyección de la relación obteniéndose las siguientes relaciones:

DOMICILIO(CALLE, NUMERO, TELEFONO)
CALLE-COL(CALLE, COLONIA)

Por lo tanto, después de llevar a cabo los pasos de la metodología, las relaciones que se obtuvieron, finalmente, son las siguientes:

AREA(NUMERO-AREA, NOMBRE-AREA).

MATERIA(CLAVE-MATERIA, NOMBRE-MATERIA).

INSTITUCION(CLAVE-INSTIT, NOMBRE-INSTIT).

EXAMEN(CLAVE-EXAMEN, FECHA-EXAMEN, TIPO-EXAMEN, TIPO-GRADO).

SEMINARIO(CLAVE-MATERIA, SUBTITULO, PERIODO).

ESTUDIANTE(NUM-CUENTA, NOMBRE, FECHA-NACIMIENTO, CALLE, NUMERO).

PROFESOR(R.F.C, NOMBRE, CLAVE-LUGAR, CALLE, NUMERO).

OBT-GRADO(R.F.C, CLAVE-INSTIT, CLAVE-EXAMEN).

REALIZACION-TESIS(R.F.C, NUM-CUENTA, CLAVE-EXAMEN).

OBT-GRADO-ESTUDIANTE(R.F.C, NUM-CUENTA, CLAVE-EXAMEN).

TRABAJO(NUMERO-AREA, R.F.C).

ESTA-DISPONIBLE(R.F.C, NUMERO-AREA, NIVEL, TIPO-DISP).

IMPARTE(R.F.C, CLAVE-MATERIA, PERIODO).

LLEVA-CURSOS(NUMERO-AREA, NUM-CUENTA, CANTIDAD).

INSCRIPCION(R.F.C, CLAVE-MATERIA, NUM-CUENTA, PERIODO,

ACREDITACION).

LUGAR-TRABAJO(CLAVE-LUGAR, LUGAR).

TELEFONO-TRABAJO(LUGAR, TELEFONO).

DOMICILIO(CALLE, NUMERO, TELEFONO).

CALLE-COL(CALLE, COLONIA).

CONCLUSIONS

CONCLUSIONES

Uno de los objetivos de la tesis, fue presentar una introducción a las bases de datos, de manera que se proporcionaron aspectos importantes acerca de ellas, así como se presentó, una técnica de normalización, la que se desarrolló paso por paso y se ejemplificó para su entendimiento, la cual se considera de suma importancia, para darle el toque final a la estructura de la base de datos obtenida por la metodología y como el objetivo principal, fue el de presentar la metodología para diseñar bases de datos relacionales, utilizando el modelo entidad-relación, esta también se presentó detalladamente paso por paso y se trató de ejemplificar cada uno de ellos, para proporcionar mayor comprensión, además de que también, se presentaron con detalle los elementos que componen al modelo y por último, se aplicó a un problema real, para ver más explícitamente, como llevar a cabo, cada paso de dicha metodología.

Con respecto al capítulo uno, en la sección de modelos, sólo se proporcionaron algunas características, para distinguir los tres modelos, no adentrándose mucho en los tres, sólo en el relacional, que es el de interés.

De acuerdo a los resultados obtenidos, de la aplicación de dicha metodología, se puede concluir, que esta permite obtener relaciones candidatas buenas, que requieren de poco refinamiento, es decir, proporciona relaciones lo mejor normalizadas posibles, las cuales se obtienen, de llevar a cabo el paso uno y paso dos de la metodología y este resultado se logra, si el modelo entidad-relación obtenido está bien definido.

Además si después de llevar a cabo los pasos anteriores, se presenta alguna relación con bajo grado de normalización, la cual se detecta, en el paso tres de la metodología, entonces dichas relaciones se tratarán de normalizar mejor, por un método de normalización, como el que se

mencionó en el capítulo uno y cuyo objetivo, es el de reducir redundancias y evitar problemas con las operaciones de actualización.

También, se puede mencionar que, además, uno de los fines para aplicar dicha metodología, es que mediante el modelo entidad-relación, el usuario de la base de datos pueda ver, de forma simple y amigable, si sus requerimientos, fueron captados adecuadamente por el diseñador, es decir, que el propósito de utilizar dicho modelo, es proporcionar, la comunicación usuario-diseñador.

Con respecto a esto, se puede decir que los diagramas proporcionan una manera sencilla de representación, que se entiende fácilmente, siempre y cuando los nombres de las entidades, atributos y relaciones sean explícitos, es decir, que el nombre ilustre bien los objetos que representa.

Por último, de acuerdo a la experiencia adquirida se puede decir que, el paso más importante de la metodología (para llevar a cabo un buen diseño) es el uno, ya que si este se realiza bien las relaciones que se obtendrán estarán lo mejor normalizadas posibles, también se puede mencionar que las transformaciones que se realizan en el paso dos son la parte más fácil de llevar a cabo y por último se debe mencionar que el llevar a cabo el paso tres requiere de mucho cuidado, ya que con este se va a dar el toque final, es decir, es la última oportunidad para mejorar la mayoría de las relaciones con bajo grado de normalización.

Además, con respecto al paso uno se puede decir que, en particular la identificación de los objetos cuesta trabajo y con mayor dificultad la identificación de las jerarquías por lo que se tiene que poner mucha atención y dedicación a todo esto, ya que como se mencionó anteriormente, es la parte fundamental para obtener un buen diseño.

BIBLIOGRAFIA

BIBLIOGRAFIA

- ENTITY-RELATIONSHIP APPROACH.
THE USE OF ER CONCEPT IN KNOWLEDGE REPRESENTATION.
PETER P. CHEN.
IEEE CS PRESS/NORTH HOLLAND.
- DATA BASE MANAGEMENT SYSTEMS.
ALFONSO F. CARDENAS.
ALLYN AND BACON, SEGUNDA EDICION, 1985.
- DISEÑO DE BASES DE DATOS.
GIO WIEDERHOLD.
MC. GRAW-HILL, SEGUNDA EDICION, 1988.
- INTRODUCCION A LOS SISTEMAS DE BASES DE DATOS.
C. J. DATE.
ADDISON-WESLEY IBEROAMERICANA, TERCERA EDICION, 1986.
- PRINCIPLES OF DATABASE SYSTEMS.
JEFFREY D. ULLMAN.
COMPUTER SCIENCE PRESS, SEGUNDA EDICION, 1982.
- THE THEORY OF RELATIONAL DATABASES.
DAVID MAIER.
COMPUTER SCIENCE PRESS, 1983.
- ON THE DEVELOPMENT OF DATA MODELS.
MICHAEL L. BRODIE.
COMPUTER CORPORATION OF AMERICA.
- RELATIONAL DATABASE DESIGN.
DAVID R. McCLANAHAN.
DBMS, OCTUBRE 1991.