



Universidad Nacional del Nordeste
Facultad de Ciencias Exactas, Naturales y Agrimensura
Trabajo de Adscripción
Sistemas de BD en
Dispositivos Móviles
y su Integración con las BD Tradicionales



Adscripta: Sofia J. Vallejos - L.U.: 37.032
Materia: Diseño y Administración de Datos
Director: Mgter. David Luis la Red Martínez

Licenciatura en Sistemas de Información
Corrientes - Argentina

2009

Índice General

1	Gestión de Datos	1
1.1	Sistemas de información	1
1.1.1	Introducción	1
1.1.2	Tipos de sistemas de información	4
2	Sistemas de Base de Datos	7
2.1	Concepto y Objetivos	7
2.2	Elementos de un sistema de base de datos	7
2.3	Estructura de una base de datos	8
3	Sistemas Gestores de Bases de Datos	11
3.1	Objetivo de los SGBD	11
3.2	Niveles de abstracción de una base de datos	12
3.2.1	Introducción	12
3.2.2	Independencia en los esquemas	13
3.2.3	Funciones	14
3.2.4	Estandarización	15
3.2.5	Recursos humanos de las bases de datos	16
3.2.6	Componentes del SGBD	18
3.2.7	Funcionamiento del SGBD	20
3.3	Tipos de SGBD	21
3.3.1	Introducción	21
3.3.2	Modelo jerárquico	23
3.3.3	Modelo en red (Codasy1)	23
3.3.4	Modelo relacional	23
3.3.5	Modelo de bases de datos orientadas a objetos	24
3.3.6	Bases de datos objeto relacionales	25
4	EL Futuro de las Bases de Datos	27

4.1	Evolución y Nuevos Retos	27
4.1.1	Tratamiento de datos multimedia. Del tiempo, de la seguridad de la incertidumbre	31
4.1.2	Bases de Datos Temporales	33
4.1.3	SGBD Orientados a objetos multinivel	35
4.1.4	Metodologías para diseño de bases de datos seguras . . .	37
4.1.5	Bases de Datos Difusas	38
5	Dist., Integración e Interoperabilidad	39
5.1	Introducción	39
5.2	Bases de datos distribuidas	40
5.3	Bases de datos federadas y multibases de datos	43
5.4	Bases de datos móviles	45
5.5	Bases de datos e infopistas	46
6	Conclusiones	51
	Bibliografía	53

Índice de Figuras

3.1	Modelo de referencia de las facilidades de usuario.	19
3.2	Esquema del funcionamiento de un SGBD.	20
3.3	Modelos de datos utilizados en el desarrollo de una BD.	22
4.1	Modelo de referencia para SGBD multimedia.	34
5.1	Distintas posibilidades de SGBD distribuidas.	42
5.2	Taxonomía para la compartición de información global.	45
5.3	SGBD Móviles como extensión de los SGBD Distribuidos.	46
5.4	Metodología RMM.	49

Capítulo 1

Gestión de Datos

1.1 Sistemas de información

1.1.1 Introducción

La necesidad de gestionar datos

En el mundo actual existe una cada vez mayor demanda de datos. Esta demanda siempre ha sido patente en empresas y sociedades, pero en estos años la demanda se ha disparado más debido al acceso multitudinario a las redes integradas en Internet y a la aparición de pequeños dispositivos (móviles y PDA) que también requieren esa información.

En informática se conoce como dato a cualquier elemento informativo que tenga relevancia para un usuario. Desde su nacimiento, la informática se ha encargado de proporcionar herramientas que faciliten la gestión de los datos.

Antes de la aparición de las aplicaciones informáticas, las empresas tenían como únicas herramientas de gestión de datos a los cajones, carpetas y fichas en las que se almacenaban los datos. En este proceso manual, el tipo requerido para manipular estos datos era enorme. Sin embargo el proceso de aprendizaje era relativamente sencillo ya que se usaban elementos que el usuario reconocía perfectamente.

Por esa razón, la informática ha adaptado sus herramientas para que los

elementos que el usuario maneja en el ordenador se parezcan a los que utilizaba manualmente.

Por eso en informática se sigue hablado de ficheros, formularios, carpetas, directorios, etc.

La Empresa como Sistema

La definición de sistema es “Conjunto de cosas que ordenadamente relacionadas entre sí contribuyen a un determinado objeto”.

La clientela fundamental del profesional de la informática es la empresa. La empresa se puede entender como un sistema formado por diversos objetos: el capital, los recursos humanos, los inmuebles, los servicios que presta, etc.

El sistema completo que forma la empresa, por otra parte, se suele dividir en los siguientes subsistemas:

- Subsistema productivo. También llamado subsistema real o físico. Representa la parte de la empresa encargada de gestionar la producción de la misma.
- Subsistema financiero. Encargado de la gestión de los bienes económicos de la empresa.
- Subsistema directivo. Encargado de la gestión organizativa de la empresa.

Hay que hacer notar que cada subsistema se asocia a un departamento concreto de la empresa.

Sistemas de Información

Los sistemas que aglutinan los elementos que intervienen para gestionar la información que manejan los subsistemas empresariales es lo que se conoce como Sistemas de Información. Se suele utilizar las siglas SI o IS (de Information Server) para referirse a ello).

Realmente un sistema de información sólo incluye la información que nos interesa de la empresa y los elementos necesarios para gestionar esa información.

Un sistema de información genérico está formado por los siguientes elementos:

- Recursos físicos. Carpetas, documentos, equipamiento, discos, etc.
- Recursos humanos. Personal que maneja la información.
- Protocolo. Normas que debe cumplir la información para que sea manejada (formato de la información, modelo para los documentos, etc). Aunque es el elemento en el que menos se repara, es uno de los más importantes, no tener un protocolo garantiza un desastre.

Las empresas necesitan implantar estos sistemas de información debido a la necesidad de gestionar las diferentes acciones que realiza una empresa. En cada proceso empresarial se genera una serie de datos de interés para la misma que deben ser manipulados adecuadamente por el sistema de información.

El Software y la Crisis del Software

Con la aparición de la informática, se empezó a implantar sistemas de información donde la mayor parte de la carga de esos sistemas se la llevaba el ordenador. Empezó a crearse software específico para la manipulación de los sistemas de información.

Inicialmente ese software no constaba de demasiadas instrucciones, pero al aumentar las necesidades de gestión de datos, los programadores tuvieron que empezar a codificar miles y miles de instrucciones.

Tras esos primeros años, ocurrió un problema que se conoció como la crisis del software. El problema es que el código creció espectacularmente de tamaño de modo que nuevas necesidades por parte de la empresa desembocaban en una cantidad interminable de correcciones en el software, lo que le hacía poco productivo (se gastaba demasiado en él).

La solución a esta crisis (todavía no del todo resuelta por culpa de una mala concienciación profesional en el sector) es la aportación de personal más

cualificado y de metodologías profesionales. A este punto es al que se dedica la ingeniería del software.

Componentes de un sistema de información electrónico

En el caso de una gestión electrónica de la información (lo que actualmente se considera un sistema de información), los componentes son:

- Datos. Se trata de la información relevante que almacena y gestiona el sistema de información.
- Hardware. Equipamiento físico que se utiliza para gestionar los datos.
- Software. Aplicaciones que permiten el funcionamiento adecuado del sistema.
- Recursos humanos. Personal que maneja el sistema de información.

1.1.2 Tipos de sistemas de información

En la evolución de los sistemas de información ha habido dos puntos determinantes, que han formado los dos tipos fundamentales de sistemas de información.

Sistemas de información orientados al proceso

En estos sistemas de información se crean diversas aplicaciones (software) para gestionar diferentes aspectos del sistema. Cada aplicación realiza unas determinadas operaciones. Los datos de dichas aplicaciones se almacenan en archivos digitales dentro de las unidades de almacenamiento del ordenador (a veces en archivos binarios, o en hojas de cálculo, ...).

Cada programa almacena y utiliza sus propios datos de forma un tanto caótica. La ventaja de este sistema (la única ventaja), es que los procesos eran independientes por lo que la modificación de uno no afectaba al resto. Pero tiene grandes inconvenientes:

- Coste de almacenamiento elevado. Al almacenarse varias veces el mismo dato, se requiere más espacio en los discos. Luego se agotarán antes.
- Datos redundantes. Ya que se repiten continuamente.
- Probabilidad alta de inconsistencia en los datos. Ya que un proceso cambia sus datos y no el resto. Por lo que el mismo dato puede tener valores distintos según qué aplicación acceda a él.
- Dificil modificación en los datos. Debido a la probabilidad de inconsistencia. Para que ésta no exista, cada modificación se debe repetir en todas las copias del dato (algo que normalmente es imposible).
- Tiempos de procesamiento elevados. Al no poder optimizar el espacio de almacenamiento.

A estos sistemas se les llama sistemas de ficheros. Se consideran también así a los sistemas que utilizan programas ofimáticos (como Word y Excel) para gestionar sus datos. De hecho estos sistemas producen los mismos (si no más) problemas.

Sistemas de información orientados a los datos. Bases de datos

En este tipo de sistemas los datos se centralizan en una base de datos común a todas las aplicaciones. En esos sistemas los datos se almacenan en una única estructura lógica que es utilizable por las aplicaciones. A través de esa estructura se accede a los datos que son comunes a todas las aplicaciones.

Ventajas

- Independencia de los datos y los programas y procesos. Esto permite modificar los datos sin modificar el código de las aplicaciones.
- Menor redundancia. No hace falta tanta repetición de datos. Sólo se indica la forma en la que se relacionan los datos.
- Integridad de los datos. Mayor dificultad de perder los datos o de realizar incoherencias con ellos.

- Mayor seguridad en los datos. Al permitir limitar el acceso a los usuarios. Cada tipo de usuario podrá acceder sólo a los datos que está autorizado.
- Datos más documentados. Gracias a los metadatos que permiten describir la información de la base de datos.
- Acceso a los datos más eficiente. La organización de los datos produce un mejor resultado en el rendimiento.
- Menor espacio de almacenamiento. Gracias a una mejor estructuración de los datos.

Desventajas

- Instalación costosa. El control y administración de bases de datos requiere de un software y hardware poderoso.
- Requiere personal cualificado. Debido a la dificultad de manejo de este tipo de sistemas.
- Implantación larga y difícil. Debido a los puntos anteriores. La adaptación del personal es mucho más complicada y lleva bastante tiempo.
- Ausencia de estándares reales. Lo cual significa una excesiva dependencia hacia los sistemas comerciales del mercado. Aunque, hoy en día, una buena parte de esta tecnología está aceptada como estándar de hecho.

Capítulo 2

Sistemas de Base de Datos

2.1 Concepto y Objetivos

Como se ha comentado anteriormente, cuando los datos de un sistema de información se almacenan en una única estructura, se llama base de datos.

Una base de datos es una colección de datos almacenados en un soporte informático permanente de forma que sea posible obtener la relación entre los datos a través de un esquema conceptual que oculte la física real de los datos.

El objetivo es que las aplicaciones puedan acceder a los datos sin necesidad de conocer exactamente cómo están almacenados los datos. Esto se consigue con un esquema conocido como esquema conceptual. Cualquier desarrollador que conozca ese esquema puede acceder a los datos desde cualquier aplicación.

2.2 Elementos de un sistema de base de datos

Los elementos de un sistema de base de datos son los mismos que los de un sistema de información. Se comentan a continuación para concretar un poco más la información que ya poseíamos:

- Hardware. Máquinas en las que se almacenan las bases de datos. Incorporan unidades de almacenamiento masivo para este fin.

- Software. Es el sistema gestor de bases de datos. La aplicación que permite el manejo de la base de datos. [4]
- Datos. Incluyen los datos que se necesitan almacenar y los metadatos que son datos que sirven para describir lo que se almacena en la base de datos.
- Usuarios. Personas que manipulan los datos del sistema. Hay tres categorías:
 1. Usuarios finales. Aquellos que utilizan datos de la base de datos para su trabajo cotidiano que no tiene por qué tener que ver con la informática. Normalmente no utilizan la base de datos directamente, sino que utilizan aplicaciones creadas para ellos a fin de facilitar la manipulación de los datos. Estos usuarios sólo acceden a ciertos datos.
 2. Desarrolladores. Analistas y programadores encargados de generar aplicaciones para los usuarios finales.
 3. Administradores. También llamados DBA (Data Base Administrator), se encargan de gestionar las bases de datos.

Hay que tener en cuenta que las necesidades de los usuarios son muy diferentes en función del tipo de usuario que sean: a los finales les interesa la facilidad de uso, a los desarrolladores la potencia y flexibilidad de los lenguajes incorporados del sistema de bases de datos, a los administradores herramientas de gestión avanzada para la base de datos.

2.3 Estructura de una base de datos

Las bases de datos están compuestas (como ya se han comentado), de datos y de metadatos. Los metadatos son datos (valga la redundancia) que sirven para especificar la estructura de la base de datos; por ejemplo qué tipo de datos se almacenan (si son texto o números o fechas), qué nombre se le da a cada dato (nombre, apellidos), cómo están agrupados, cómo se relacionan.

De este modo se producen dos visiones de la base de datos:

- Estructura lógica o conceptual. Indica la composición y distribución teórica de la base de datos. La estructura lógica sirve para que las aplicaciones puedan utilizar los elementos de la base de datos sin saber realmente cómo se están almacenando.
- Estructura física. Es la estructura de los datos tan cual se almacenan en las unidades de disco. La correspondencia entre la estructura lógica y la física se almacena en la base de datos (en los metadatos).

Capítulo 3

Sistemas Gestores de Bases de Datos

3.1 Objetivo de los SGBD

Un sistema gestor de bases de datos o SGBD (aunque se suele utilizar más a menudo las siglas DBMS procedentes del inglés, Data Base Management System) es el software que permite a los usuarios procesar, describir, administrar y recuperar los datos almacenados en una base de datos.

En estos Sistemas se proporciona un conjunto coordinado de programas, procedimientos y lenguajes que permiten a los distintos usuarios realizar sus tareas habituales con los datos, garantizando además la seguridad de los mismos.

El éxito del SGBD reside en mantener la seguridad e integridad de los datos.

Lógicamente tiene que proporcionar herramientas a los distintos usuarios. Entre las herramientas que proporciona están:

- Herramientas para la creación y especificación de los datos. Así como la estructura de la base de datos.
- Herramientas para administrar y crear la estructura física requerida en las unidades de almacenamiento.

- Herramientas para la manipulación de los datos de las bases de datos, para añadir, modificar, suprimir o consultar datos.
- Herramientas de recuperación en caso de desastre.
- Herramientas para la creación de copias de seguridad.
- Herramientas para la gestión de la comunicación de la base de datos.
- Herramientas para la creación de aplicaciones que utilicen esquemas externos de los datos.
- Herramientas de instalación de la base de datos.
- Herramientas para la exportación e importación de datos.

3.2 Niveles de abstracción de una base de datos

3.2.1 Introducción

En cualquier sistema de información se considera que se pueden observar los datos desde dos puntos vista:

- Vista externa. Esta es la visión de los datos que poseen los usuarios del Sistema de Información.
- Vista física. Esta es la forma en la que realmente están almacenados los datos.

En un Sistema de ficheros, los usuarios ven los datos desde las aplicaciones creadas por los programadores. Esa vista pueden ser formularios, informes visuales o en papel.

Pero la realidad física de esos datos, tal cual se almacenan en los discos, no la ven. Esa visión está reservada a los administrados.

En el caso de los Sistemas de Base de datos, se añade una tercera vista, que es la vista conceptual. Esa vista se sitúa entre la física y la externa. Se habla pues en Bases de datos de la utilización de tres esquemas para representar los datos:

- Esquema físico

Representa la forma en la que están almacenados los datos. Esta visión sólo la requiere el administrador. El administrador la necesita para poder gestionar más eficientemente la base de datos.

En este esquema es donde aparecen las unidades de disco, archivos y carpetas del sistema.

- Esquema conceptual

Se trata de un esquema teórico de los datos en la que figuran organizados en estructuras reconocibles del mundo real y en el que también aparece la forma de relacionarse los datos.

Esta estructura es utilizada por el desarrollador, que necesita conocerla para al crear aplicaciones saber cómo acceder. Todas las aplicaciones que se crean sobre la base de datos utilizan este esquema.

Realmente cuando se habla del diseño de la base de datos, se suele referir a este esquema. El esquema conceptual lo realiza el diseñador de la base de datos.

- Esquema externo

Se trata de la visión de los datos que poseen los usuarios finales. Esa visión es la que obtienen a través de las aplicaciones. Las aplicaciones creadas por los desarrolladores abstraen la realidad conceptual de modo que el usuario no conoce las relaciones entre los datos, como tampoco conoce todos los datos que realmente se almacenan.

Realmente cada aplicación produce un esquema externo diferente (aunque algunos pueden coincidir) o vistas de usuario. El conjunto de todas las vistas de usuario es lo que se denomina esquema externo global.

3.2.2 Independencia en los esquemas

Lo interesante de estos esquemas es que trabajan independientemente. Cada usuario utiliza el esquema que necesita sin saber nada del resto. Esa independencia se traduce en:

- Independencia física de los datos. Aunque el esquema físico cambie, el esquema conceptual no debe verse afectado. En la práctica esto significa que aunque la base de datos se almacene en otro disco duro, la parte conceptual de la base de datos no debe verse afectada. No habrá que modificar dicho esquema.
- Independencia lógica de los datos. Significa que aunque se modifique los datos del esquema conceptual, las vistas que poseen las aplicaciones (los esquemas externos) no se verán afectados.

3.2.3 Funciones

- Los SGBD tienen que realizar tres tipos de funciones para ser considerados válidos.

Función de descripción o definición

Permite al diseñador de la base de datos crear las estructuras apropiadas para integrar adecuadamente los datos. Esta función es la que permite definir las tres estructuras de la base de datos (relacionadas con sus tres esquemas):

- Estructura interna.
- Estructura conceptual.
- Estructura externa.

Esta función se realiza mediante el lenguaje de descripción de datos o DDL. Mediante ese lenguaje:

- Se definen las estructuras de datos.
- Se definen las relaciones entre los datos.
- Se definen las reglas que han de cumplir los datos.

Función de manipulación

Permite modificar y utilizar los datos de la base de datos. Se realiza mediante el lenguaje de modificación de datos o DML. Mediante ese lenguaje se puede:

- Añadir datos.
- Eliminar datos.
- Modificar datos.
- Buscar datos.

Actualmente se suele distinguir la función de buscar datos respecto del resto. Para lo cual se proporciona un lenguaje de consulta de datos o DQL.

Función de control

Mediante esta función los administradores poseen mecanismos para determinar las visiones de los datos permitidas a cada usuario, además de proporcionar elementos de creación y modificación de esos usuarios.

Se suelen incluir aquí las tareas de copia de seguridad, carga de ficheros, auditoria protección ante ataques externos, configuración del sistema.

- El lenguaje que implementa esta función es el lenguaje de control de datos o DCL.

3.2.4 Estandarización

Es uno de los aspectos que todavía sigue pendiente. Desde la aparición de los primeros gestores de base de datos se intentó llegar a un acuerdo para que hubiera una estructura común para todos ellos, a fin de que el aprendizaje y manejo de este software fuera más provechoso y eficiente.

El acuerdo nunca se ha conseguido del todo, no hay estándares aceptados del todo.

Sí hay unas cuentas propuestas de estándares que sí funcionan como tales.

Estructuras operacionales

Actualmente casi todos los sistemas gestores de base de datos poseen también la misma idea operacional en la que se entiende que la base de datos se almacena en un servidor y hay una serie de clientes que pueden acceder a los datos del mismo. Las posibilidades son:

- Estructura Cliente-Servidor. Estructura clásica, la base de datos y su SGBD están en un servidor al cual acceden los clientes. El cliente posee software que permite al usuario enviar instrucciones al SGBD en el servidor y recibir los resultados de estas instrucciones. Para ello el software cliente y el servidor deben utilizar software de comunicaciones en red.
- Cliente multi-servidor. Ocurre cuando los clientes acceden a datos situados en más de un servidor. También se conoce esta estructura como base de datos distribuida. El cliente no sabe si los datos están en uno o más servidores, ya que el resultado es el mismo independientemente de dónde se almacenan los datos. En esta estructura hay un servidor de aplicaciones que es el que recibe las peticiones y el encargado de traducirlas a los distintos servidores de datos para obtener los resultados. Una posibilidad muy extendida hoy en día es la posibilidad Cliente/Servidor Web/Servidor de datos, el cliente se conecta a un servidor mediante un navegador web y desde las páginas de este ejecuta las consultas. El servidor web traduce esta consulta al servidor (o servidores) de datos.

3.2.5 Recursos humanos de las bases de datos

Intervienen (como ya se ha comentado) muchas personas en el desarrollo y manipulación de una base de datos. Habíamos seleccionado cuatro tipos de usuarios (administradores/as, desarrolladores, diseñadores/as y usuarios/as). Ahora vamos a desglosar aún más esta clasificación.

Informáticos

Lógicamente son los profesionales que definen y preparan la base de datos. Pueden ser:

- Directivos/as. Son los organizadores y coordinadores del proyecto a desarrollar. Esto significa que son los encargados de decidir los recursos que se pueden utilizar, planificar el tiempo y las tareas, la atención al usuario y de dirigir las entrevistas y reuniones pertinentes.
- Analistas. Son los encargados de controlar el desarrollo de la base de datos aprobada por la dirección. Son además los diseñadores de la base de datos (especialmente de los esquemas interno y conceptual) y los coordinadores de la programación de la misma.
- Administradores/as de las bases de datos. Definen la seguridad de la base de datos y gestionan las copias de seguridad y la gestión física de la base de datos. Los analistas suelen tener esta funcionalidad cuando la base de datos está creada.
- Desarrolladores/as o programadores/as. Encargados de la realización de las aplicaciones de usuario de la base de datos.
- Equipo de mantenimiento. Encargados de dar soporte a los usuarios en el trabajo diario (suelen incorporar además tareas administrativas).

Usuarios

- Expertos/as. Utilizan el lenguaje de manipulación de datos (DML) para acceder a la base de datos. Son usuarios que utilizan la base de datos para gestión avanzada de decisiones.
- Habituales. Utilizan las aplicaciones creadas por los desarrolladores para consultar y actualizar los datos. Son los que trabajan en la empresa a diario con estas herramientas y el objetivo fundamental de todo el desarrollo de la base de datos.
- Ocasionales. Son usuarios que utilizan un acceso mínimo a la base de datos a través de una aplicación que permite consultar ciertos datos. Serían por ejemplo los usuarios que consultan el horario de trenes a través de Internet.

3.2.6 Componentes del SGBD

Estructura multicapa

Un SGBD está en realidad formado por varias capas que actúan como interfaces entre el usuario y los datos. El propio ANSI/X3/SPARC introdujo una mejora de su modelo en 1988 a través de un grupo de trabajo llamado UFTG (User Facilities Task Group, grupo de trabajo para las facilidades de usuario). Este modelo toma como objeto principal, al usuario habitual de la base de datos y orienta el funcionamiento de la base de datos de modo que este usuario ignora el funcionamiento externo.

Desde esta óptica para llegar a los datos hay que pasar una serie de capas que poco a poco van entrando más en la realidad física de la base de datos. Esa estructura se muestra en la figura 3.1 de la página 19.

Núcleo

El núcleo de la base de datos es el encargado de traducir las operaciones que le llegan a instrucciones ejecutables por el sistema operativo en el lenguaje que éste último requiera.

Diccionario de datos

Se trata del elemento que posee todos los metadatos. Gracias a esta capa las solicitudes de los clientes se traducen en instrucciones que hacen referencia al esquema interno de la base de datos. La capa de acceso a datos es la que permite comunicar a las aplicaciones de usuario con el diccionario de datos a través de las herramientas de gestión de datos que incorpore el SGBD.

Facilidades de usuario

Son las herramientas que proporciona el SGBD a los usuarios para permitir un acceso más sencillo a los datos. Actúan de interfaz entre el usuario y la base de datos, y son el único elemento que maneja el usuario.



Figura 3.1: Modelo de referencia de las facilidades de usuario.

3.2.7 Funcionamiento del SGBD

En la figura 3.2 de la página 20 presenta el funcionamiento típico de un SGBD:

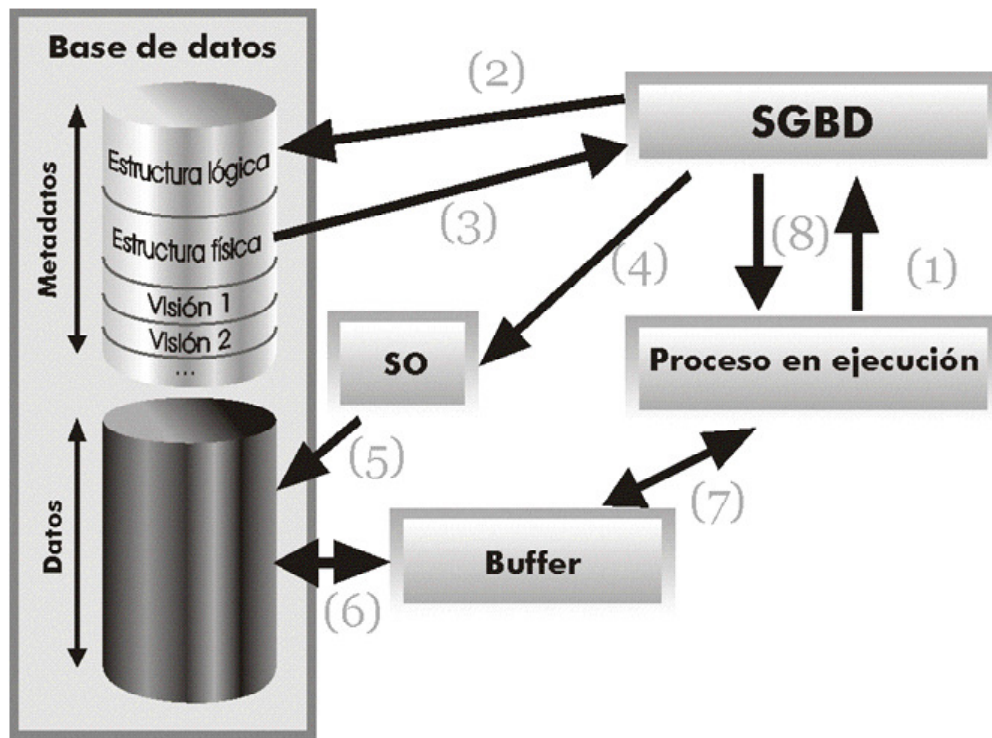


Figura 3.2: Esquema del funcionamiento de un SGBD.

El esquema anterior reproduce la comunicación entre un proceso de usuario que desea acceder a los datos y el SGBD:

- 1) El proceso lanzado por el usuario llama al SGBD indicando la porción de la base de datos que se desea tratar.
- 2) El SGBD traduce la llamada a términos del esquema lógico de la base de datos. Accede al esquema lógico comprobando derechos de acceso y la traducción física.
- 3) El SGBD obtiene el esquema físico.
- 4) El SGBD traduce la llamada a los métodos de acceso del Sistema Ope-

rativo que permiten acceder a los datos requeridos.

5) El Sistema Operativo accede a los datos tras traducir las órdenes dadas por el SGBD.

6) Los datos pasan del disco a una memoria intermedia o buffer. En ese buffer se almacenarán los datos según se vayan recibiendo.

7) Los datos pasan del buffer al área de trabajo del usuario (ATU) del proceso del usuario.

8) El SGBD devuelve indicadores en los que manifiesta si ha habido errores o advertencias a tener en cuenta. Esto se indica al área de comunicaciones del proceso de usuario. Si las indicaciones son satisfactorias, los datos de la ATU serán utilizables por el proceso de usuario.

3.3 Tipos de SGBD

3.3.1 Introducción

Como se ha visto en los apartados anteriores, resulta que cada SGBD puede utilizar un modelo diferente para los datos. Por lo que hay modelos conceptuales diferentes según que SGBD utilicemos.

No obstante existen modelos lógicos comunes, ya que hay SGBD de diferentes tipos.

En la realidad el modelo ANSI se modifica para que existan dos modelos internos: el modelo lógico (referido a cualquier SGBD de ese tipo) y el modelo propiamente interno (aplicable sólo a un SGBD en particular). De hecho en la práctica al definir las bases de datos desde el mundo real hasta llegar a los datos físicos se pasa por los esquemas que se muestran en la figura 3.3 de la página 22.

Por lo tanto la diferencia entre los distintos SGBD está en que proporcionan diferentes modelos lógicos.

Diferencias entre el modelo lógico y el conceptual:

- El modelo conceptual es independiente del DBMS que se vaya a utilizar. El lógico depende de un tipo de SGBD en particular.

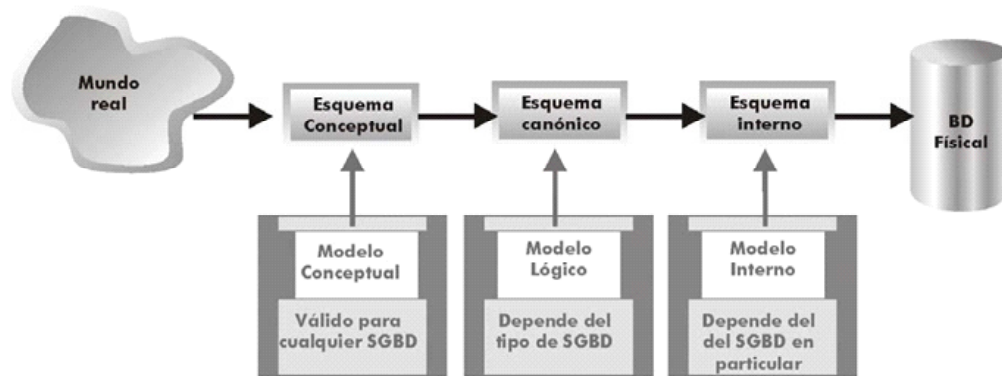


Figura 3.3: Modelos de datos utilizados en el desarrollo de una BD.

- El modelo lógico es más cercano al ordenador.
- El modelo conceptual es más cercano al usuario, el lógico es el encargado de establecer el paso entre el modelo informático y el modelo físico del sistema.

Algunos ejemplos de modelos conceptuales son:

Sistemas Gestores de Bases de Datos:

- Modelo E/R.
- Modelo RM/T.
- Modelos semánticos.

Ejemplos de modelos lógicos son:

- Modelo relacional.
- Modelo Codasyl.
- Modelo Jerárquico.

A continuación se comentarán los modelos lógicos más importantes.

3.3.2 Modelo jerárquico

Era utilizado por los primeros SGBD, desde que IBM lo definió para su IMS (Information Management System, Sistema Administrador de Información) en 1970. Se le llama también modelo en árbol debido a que utiliza una estructura en árbol para organizar los datos.

La información se organiza con un jerarquía en la que la relación entre las entidades de este modelo siempre es del tipo padre / hijo. De esta forma hay una serie de nodos que contendrán atributos y que se relacionarán con nodos hijos de forma que puede haber más de un hijo para el mismo padre (pero un hijo sólo tiene un padre).

Los datos de este modelo se almacenan en estructuras lógicas llamadas segmentos.

Los segmentos se relacionan entre sí utilizando arcos.

La forma visual de este modelo es de árbol invertido, en la parte superior están los padres y en la inferior los hijos.

3.3.3 Modelo en red (Codasyl)

Es un modelo que ha tenido una gran aceptación (aunque apenas se utiliza actualmente). En especial se hizo popular la forma definida por Codasyl a principios de los 70 que se ha convertido en el modelo en red más utilizado.

El modelo en red organiza la información en registros (también llamados nodos) y enlaces. En los registros se almacenan los datos, mientras que los enlaces permiten relacionar estos datos. Las bases de datos en red son parecidas a las jerárquicas sólo que en ellas puede haber más de un padre.

En este modelo se pueden representar perfectamente cualquier tipo de relación entre los datos (aunque el Codasyl restringía un poco las relaciones posibles), pero hace muy complicado su manejo.

3.3.4 Modelo relacional

En este modelo los datos se organizan en tablas cuyos datos se relacionan.

Edgar Frank Codd definió las bases del modelo relacional a finales de los 60.

En 1970 publica el documento “A Relational Model of data for Large Shared Data Banks” (Un modelo relacional de datos para grandes bancos de datos compartidos).

Es el modelo de bases de datos más popular. Hoy en día casi todas las bases de datos siguen este modelo.

Objetivos

Codd perseguía estos objetivos con su modelo:

- Independencia física. La forma de almacenar los datos, no debe influir en su manipulación lógica. Si el almacenamiento físico cambia, los usuarios no tienen ni siquiera por qué enterarse, seguirán funcionando sus aplicaciones.
- Independencia lógica. Las aplicaciones que utilizan la base de datos no deben ser modificadas por que se modifiquen elementos de la base de datos. Es decir, añadir, borrar y suprimir datos, no influye en las vistas de los usuarios.
- Flexibilidad. La base de datos ofrece fácilmente distintas vistas en función de los usuarios y aplicaciones.
- Uniformidad. Las estructuras lógicas siempre tienen una única forma conceptual (las tablas).
- Sencillez.

3.3.5 Modelo de bases de datos orientadas a objetos

Desde la aparición de la programación orientada a objetos (POO u OOP) se empezó a pensar en bases de datos adaptadas a estos lenguajes. La programación orientada a objetos permite cohesionar datos y procedimientos, haciendo que se diseñen estructuras que poseen datos (atributos) en las que se definen los procedimientos (operaciones) que se pueden realizar con los datos. En las bases orientadas a objetos se utiliza esta misma idea.

A través de este concepto se intenta que estas bases de datos consigan arreglar las limitaciones de las relacionales. Por ejemplo el problema de la herencia (el hecho de que no se puedan realizar relaciones de herencia entre las tablas), tipos definidos por el usuario, disparadores (triggers) almacenables en la base de datos, soporte multimedia...

Se supone que son las bases de datos de tercera generación (la primera fue las bases de datos en red y la segunda las relacionales), lo que significa que el futuro parece estar a favor de estas bases de datos. Pero siguen sin reemplazar a las relacionales, aunque son el tipo de base de datos que más está creciendo en los últimos años.

Su modelo conceptual se suele diseñar en UML y el lógico actualmente en ODMG (Object Data Management Group, grupo de administración de objetos de datos, organismo que intenta crear estándares para este modelo).

3.3.6 Bases de datos objeto relacionales

Tratan de ser un híbrido entre el modelo relacional y el orientado a objetos. El problema de las bases de datos orientadas a objetos es que requieren reinvertir capital y esfuerzos de nuevo para convertir las bases de datos relacionales en bases de datos orientadas a objetos. En las bases de datos objeto relacionales se intenta conseguir una compatibilidad relacional dando la posibilidad de integrar mejoras de la orientación a objetos.

Estas bases de datos se basan en el estándar SQL 99. En ese estándar se añade a las bases relacionales la posibilidad de almacenar procedimientos de usuario, triggers, tipos definidos por el usuario, consultas recursivas, bases de datos OLAP, tipos LOB.

Las últimas versiones de la mayoría de las clásicas grandes bases de datos relacionales (Oracle, SQL Server, Informix, ...) son objeto relacionales.

Capítulo 4

EL Futuro de las Bases de Datos

4.1 Evolución y Nuevos Retos

Evolución de la gestión de datos Desde que se empezaron a introducir los ordenadores para automatizar la gestión de las empresas en la década de los sesenta, empleando programas escritos en COBOL, la evolución de los sistemas de información ha tenido una considerable repercusión en la gestión de los datos, al exigirse cada vez mayores prestaciones de la información almacenada en el sistema.

Poco a poco, el centro de gravedad de la informática, que estaba situado en el proceso, se desplazó hacia la estructuración de los datos, siendo actualmente los aspectos relacionados con este tema un eje fundamental alrededor del cual gira una gran parte del conjunto de problemas con los que se enfrenta todo diseñador de un sistema de información.

Se cambia, por tanto, de sistemas orientados hacia el proceso a sistemas orientados hacia los datos, donde estos adquieren el protagonismo, pasando desde el plano más bien oscuro y difuso en el que estaban situados a ocupar un lugar privilegiado en el interés de todo informático.

Surge así, a finales de los sesenta y principios de los setenta, la primera generación de productos de bases de datos en red (basados en lo que posteriormente se conocería por modelos jerárquicos y Codasyl), entre los que destaca-

ron por su impacto el IMS de IBM y el IOMS de Cullinet. Estos productos, si bien resultaban bastante eficientes, presentaban lenguajes procedimentales, que obligaban al programador a navegar (registro a registro) por la base de datos, y que no disponían de la suficiente independencia físico/lógica, lo que conllevaba una escasa flexibilidad.

Cuando, en 1970, el Dr. Codd propuso el modelo relacional, no podía pensar que lo que se consideraba más bien una elegante teoría matemática (un “juego”, según algunos) sin posibilidad de implementación eficiente en productos comerciales, iba a convertirse en los años ochenta, en un señuelo obligado en la promoción de todo SGBD. Aparece así la segunda generación de productos de bases de datos: ORACLE, DB2, INGRES, INFORMIX, SYBASE, etc. que presentan una mayor independencia físico/lógica, mayor flexibilidad y lenguajes de especificación (que actúan sobre conjuntos de registros). Este tipo de productos se ha ido imponiendo en el mercado y ha sido uno de los principales focos de investigación de los últimos tiempos.

Los trabajos de investigación que se realizaron durante la década de los ochenta se centraron en la optimización de consultas, lenguajes de alto nivel, teoría de la normalización, organizaciones físicas para el almacenamiento de las relaciones, algoritmos para la gestión de memorias intermedias (buffers), técnicas de indexación para un acceso asociativo más rápido (distintas variaciones de los árboles B), sistemas distribuidos, diccionarios de datos, gestión de transacciones, etc. Estas investigaciones han tenido como consecuencia la elevada tasa de transacciones de muchos de los productos actuales que permiten asegurar entornos transaccionales en línea (OLTP) muy eficientes y seguros. También cabe recordar que durante la primera mitad de los ochenta se estandariza el lenguaje SQL (el SQL/ANS se aprueba en 1986), ofreciendo, al cabo de poco tiempo, prácticamente todos los productos una interfaz SQL, aún los no relacionales (sistemas “renacidos”).

En los últimos años venimos asistiendo a un avance espectacular en la tecnología de bases de datos. Temas que hasta hace poco parecían exclusivos de laboratorios y centros de investigación, comienzan a aparecer en las últimas versiones de algunos SGBD y en nuevos productos: bases de datos multimedia, activas, deductivas, orientadas a objetos, seguras, temporales, móviles, paralelas, difusas, etc.

Esta nueva generación de bases de datos (La “tercera”), se caracteriza por “proporcionar capacidades de gestión de datos al igual que sus predecesores, permitiendo que grandes cantidades de datos persistentes sean compartidos

por muchos usuarios. También proporcionan gestión de objetos, permitiendo tipos de datos mucho más complejos, objetos multimedia, datos derivados, encapsulamiento de la semántica de los datos, así como otras nuevas capacidades. Algunos proporcionan incluso gestión de conocimiento, soportando un gran número de reglas complejas para inferencia automática de información y también para mantener las restricciones de integridad entre datos”, [1].

Esta misma idea es la introducida por [5], que denomina a este nuevo tipo de SGBD “inteligente”, definiéndolo como aquel que “Gestiona Información de una manera natural, haciendo esta información fácil de almacenar, acceder y utilizar”. Para ello destaca la necesidad de:

- Un fuerte soporte para texto, imagen, voz, animación y vídeo.
- Un modelo de datos orientado a objetos, y
- Un soporte de reglas declarativas para expresar las interrelaciones semánticas entre objetos.

Todos estos avances, tanto en el campo teórico como en el práctico, se apoyan en dos elementos clave de los SGBD, la arquitectura y los modelos de datos, [2]. Respecto a la arquitectura, las propuestas de ANSI e ISO en sus modelos de referencia han influido positivamente, no sólo en las investigaciones teóricas, sino también en las aplicaciones prácticas. En gran parte de las normas de ISO, y de los productos, subyacen dos principios de los modelos de referencia, la arquitectura a tres niveles y la descripción recursiva de los datos.

En lo que se refiere a los modelos de datos, ha sido el modelo relacional el que ha marcado las líneas de investigación en los últimos veinticinco años y el que, como acabamos de señalar, se ha asentado en la mayoría de los productos actuales. Aunque han surgido otros modelos, como los orientados al objeto, que presentan grandes ventajas para determinado tipo de aplicaciones, y cuyos conceptos también se empiezan a incorporar a los sistemas relacionales (extendidos).

Nuevos Retos Aun cuando se podría pensar que la tecnología de las bases de datos ha alcanzado ya su madurez, la tercera generación de bases de datos pone de manifiesto que no se conocen todavía las soluciones a los problemas de las bases de datos de los años 2.000.

Nuevos tipos de aplicaciones Las áreas de aplicación de las bases de datos se han ido extendiendo a medida que los avances de los SGBD permitían atender las exigencias de nuevos dominios culturales, científicos, industriales y comerciales. Los productos de la primera generación de bases de datos proporcionaban soluciones a los problemas de tipo administrativo (gestión de personal, reserva de plazas, etc.), pero resultaban inadecuados para responder a consultas no planificadas como las que lleva consigo la toma de decisiones, debido a la falta de independencia ya mencionada y a sus interfaces de bajo nivel (que impedían su utilización directa por los usuarios finales).

La llegada de los productos relacionales cambia esta situación e incrementa los campos de aplicación de las bases de datos. Sin embargo, en la actualidad todavía existen importantes áreas culturales, científicas o industriales donde la tecnología de las bases de datos apenas ha comenzado a introducirse, debido a las especiales exigencias de este tipo de aplicaciones:

- **CASE.** La ingeniería del software asistida por ordenador exige gestionar conjuntos de información asociada a todo el desarrollo de un sistema informático, como son la Planificación y control, diseño de la base de datos y de los programas, código fuente, documentación, etc. Este tipo de aplicaciones, similares a las de otras áreas de la ingeniería tiene exigencias especiales respecto a los SGBD que las soporten, como es el control de versiones, facilidades en cuanto a disparadores, almacenamiento de diagramas, matrices, etc.
- **CAD/CAM/CIM.** La fabricación y el diseño asistido por ordenador es otra área que va a obligar a introducir nuevas facilidades en los SGBD, como alertas y disparadores, para poder gestionar los datos relativos a todas las fases de operación de una planta de producción (código para máquinas herramienta, resultados de test, planificación de la producción, etc.) que estarían almacenados en la base de datos.
- **SIG.** Los sistemas de información geográfica, más conocidos por sus siglas inglesas GIS (Geographical Information Systems), son aplicaciones relativas a datos geográficos, con información codificada de mapas, que se utiliza en investigaciones ambientales y militares, prestación de servicios urbanísticos, etc. Este tipo de tecnología de bases de datos se engloba dentro de la categoría de “bases de datos espaciales”.
- **Información Textual.** El tratamiento y recuperación de información textual ha sido realizada hasta ahora por medio de un software específico.

La tendencia en estos momentos es a integrar los datos estructurados, propios de la mayoría de bases de datos actuales, con distintos tipos de datos no estructurados, entre ellos información textual.

- **Aplicaciones científicas.** Uno de los sistemas que más requisitos imponen a los SGBD es el Earth Observing System, una colección de satélites que la NASA ha lanzado en 1998, que ayudó a los científicos a estudiar las tendencias de la atmósfera, los océanos y la tierra. Estos satélites envía a la tierra 1 petabyte de información al año. Toda esta información se almacenara en el EOSDIS (EOS Data and Information System).
- **Sistemas médicos.** El personal sanitario necesita información de diverso tipo sobre los pacientes, que además se puede encontrar dispersa por muchos centros. También presentan necesidades (como otros tipos de aplicaciones) de reconocimiento de patrones asociados a imágenes. El almacenamiento de millones de bits necesarios en las aplicaciones de imágenes plantean problemas especiales a los SGBD que han de manejar eficientemente estos objetos no estructurados. Dentro de los sistemas médicos también cobran gran importancia los temas relativos a la seguridad de las bases de datos.
- **Publicación digital.** El sector editorial también está experimentando grandes cambios al poderse archivar libros de forma electrónica, así como combinar texto con audio, vídeo, imágenes gráficas, anotaciones, etc.
- **Educación y formación.** También sufrirá cambios este sector, ya que en lugar de las clases asistenciales se potenciará la utilización de paquetes de enseñanza multimedia a distancia así como “aulas y conferencias virtuales”. En estos dos últimos tipos de aplicaciones se necesita entregar grandes cantidades de datos en tiempo real, y cobra gran importancia la protección de la propiedad intelectual.

4.1.1 Tratamiento de datos multimedia. Del tiempo, de la seguridad de la incertidumbre

Introducción Los sistemas de bases de datos más conocidos en la actualidad, los relacionales, se caracterizan por gestionar de manera eficiente datos formateados (tipo numérico, carácter, fecha, etc.) con un moderado grado de seguridad (confidencialidad, integridad y disponibilidad). Sin embargo,

las aplicaciones que están surgiendo para atender a nuevos tipos de negocio requieren:

- Soportar tipos de datos más sofisticados (voz, vídeo, imagen, texto, etc.),
- Tratar la dimensión temporal,
- Garantizar una mayor seguridad, y
- Manejar datos imprecisos.

Por todo ello, diversas instituciones académicas y laboratorios de fabricantes de SGBD están trabajando con el fin de ampliar las capacidades de los sistemas de bases de datos y adecuarlos así a estas necesidades.

A continuación presentaremos, de manera muy resumida, las características de los SGBD multimedia, temporales, seguros y difusos, remitiendo al lector interesado en profundizar en algunos de estos temas a la bibliografía referenciada.

Bases de Datos Multimedia En la actualidad se está desarrollando toda una serie de aplicaciones que incorporan el tratamiento de datos multimedia (televisión interactiva, sistemas de información geográficos, enciclopedias electrónicas, aplicaciones musicales, etc.). Si las bases de datos no quieren “quedarse fuera” de este tipo de aplicaciones deben soportar el tratamiento de los datos multimedia de una manera eficiente.

Hay que tener en cuenta que este tipo de datos presenta algunas características especiales:

- Los datos multimedia son muy grandes y voluminosos, por lo que a pesar del avance del hardware, no parece probable que se mantengan en discos magnéticos. Se necesita un nuevo nivel de memoria, conocido como “memoria terciaria”, por ejemplo juke boxes de discos compactos.
- Estos nuevos tipos de datos, llevan consigo operaciones que requieren una implementación muy eficiente.
- Los datos multimedia presentan restricciones en la velocidad de entrega; por ejemplo, los objetos de un vídeo se deben recuperar a una velocidad constante.

- Para cada tipo de dato multimedia, debe definirse la calidad de servicio deseada, cómo se degrada, que se hace ante una degradación, etc.
- Se necesita un nuevo tipo de interfaces, para visualizar de forma gráfica, espacial, y poder realizar consultas a la base de datos basándose en la forma, color u otras características de los objetos.

Los SGBD actuales no están concebidos, sin embargo, para manejar grandes cantidades de datos en dispositivos como CD-ROM o videodiscos. En general, podemos afirmar que el modelo relacional no es el más adecuado para soportar los datos multimedia, aunque en la actualidad la mayor parte los productos ofrecen la posibilidad de definir BLOBs (Binary Large Objects) para almacenar texto, vídeo, sonido, etc. Con este mecanismo no es posible expresar la semántica asociada a los objetos multimedia ni realizar accesos por determinados componentes de éstos.

Aunque no existen propuestas universalmente aceptadas sobre qué componentes o características debe poseer un SGBD multimedia, en [4] se propone un interesante modelo de referencia, que se muestra en la figura 4.1 de la página 34. Según esta arquitectura, un SGBD multimedia consta de tres niveles:

- Nivel de SGBD monomedia, que proporcionan las funcionalidades esenciales para gestionar un medio particular.
- Nivel de gestión/composición multimedia, que permite integrar los monomedia para componer documentos multimedia, así como coordinar los diferentes SGBD monomedia en caso de que estuviesen distribuidos.
- Nivel de interfaz de usuario, que ofrece diversas facilidades para presentación y visualización de imágenes, vídeo, etc. así como varios lenguajes de consulta.

4.1.2 Bases de Datos Temporales

De manera general, en las bases de datos temporales se suelen distinguir dos aspectos importantes: la gestión de la historia y la gestión de versiones. En estos últimos años se ha logrado un consenso en cuanto a la semántica de la historia, pero no a la gestión de versiones, para la que existen muchas

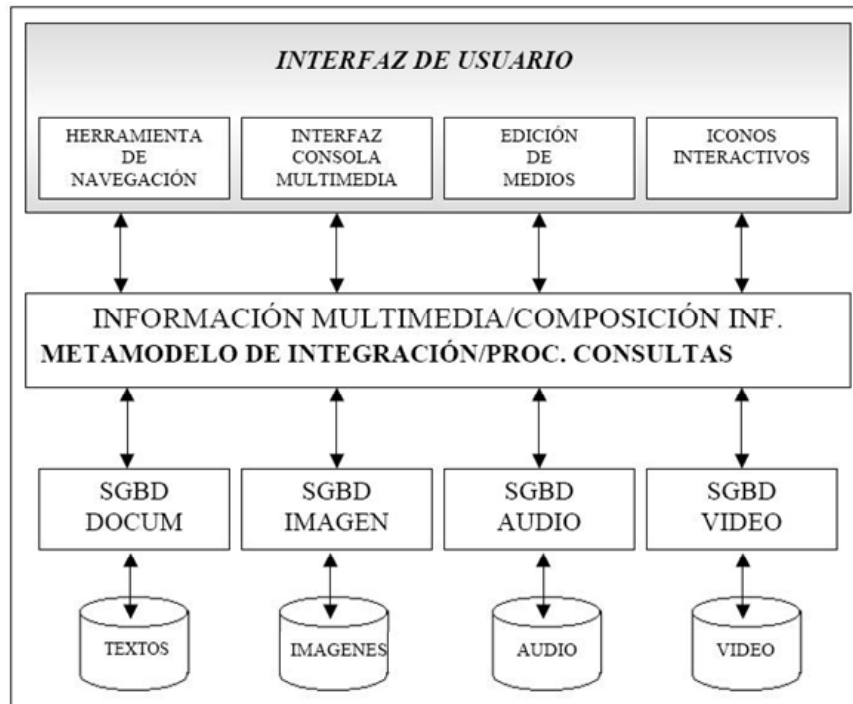


Figura 4.1: Modelo de referencia para SGBD multimedia.

propuestas en el contexto de los sistemas de diseño asistido por ordenador e ingeniería de software.

En un sentido más estricto, se conocen como bases de datos temporales aquellas que gestionan la historia, pudiendo contemplar dos dimensiones del “tiempo”:

- Tiempo válido, en el que un hecho es verdadero en el mundo real (con independencia de su registro en la base de datos).
- Tiempo de transacción, durante el cual el hecho estuvo presente en la base de datos.

Las dos dimensiones son ortogonales, y permiten distinguir de esta manera cuatro tipos de SGBD, según soporten:

- Ninguna dimensión: SGBD “instantáneos” (snapshots), como son los productos relacionales más difundidos en la actualidad,
- Sólo tiempo válido,
- Sólo tiempo de transacción,
- Ambas dimensiones: sistemas bitemporales.

4.1.3 SGBD Orientados a objetos multinivel

En [8] se presenta una taxonomía para este tipo de sistemas, basada en distintos parámetros a tener en cuenta a la hora de diseñar un modelo de seguridad para los SGBDOO:

A) Semántica de la clasificación:

Apartado en el que se incluyen dos aspectos importantes:

A.1.- Modelo subyacente, que es el modelo sobre el que se basa la clasificación. Puede ser de tres tipos:

- a) Niveles explícitos, si se asignan -explícitamente- niveles de confidencialidad a las entidades y de autorización a los sujetos.

- b) Listas de control de acceso, asociadas a las entidades que contienen los identificadores de los sujetos que están autorizados a accederlas.
- c) Capacidades, esto es, identificadores no falsificables que poseen los sujetos, que podrán acceder a una determinada entidad en función de su capacidad.

A.2.- Interpretación de la protección, indica qué es lo que se protege:

- a) El dato, en cuyo caso hablamos de “protección de acceso”.
- b) El hecho de que el dato existe, “protección de existencia”.

- B).- Clasificación estructural:

Este apartado considera la influencia de la estructura de los datos en la clasificación de las entidades:

B.1. -Entidades protegibles, que pueden ser objetos, servicios y atributos de objetos, clases, servicios y atributos de clases, etc. Existen dos opciones a destacar:

- a) Todas las características del objeto tienen el mismo nivel de sensibilidad: “objeto mononivel”.
- b) Los atributos y servicios de un objeto pueden clasificarse de forma individual, en cuyo caso hablamos de “objeto multinivel”.

B.2. -Instanciación de la clasificación, parámetro que especifica cómo y cuando se clasifican las entidades; existiendo, por lo que respecta a los objetos, al menos, tres posibilidades:

- a) Clasificar la clase y aplicar esta clasificación a todos los objetos (ejemplares de la misma).
- b) Clasificar todas las características del objeto en el momento en que se instancia.
- c) Especificar reglas que determinen la sensibilidad del objeto.

B.2. -Restricciones de interrelación, que afectan a las entidades relacionadas por interrelaciones de agregación, instanciación, herencia, niveladas (asociaciones), etc. Las restricciones pueden ser:

- a) Obligatorias, si se tienen que cumplir como resultado de elecciones tomadas con anterioridad o por ser inherentes a la estructura.
- b) Adicionales, otras que el modelo prescribe, por ejemplo, a fines de simplificación.

C . - Clasificación dinámica:

Esta sección abarca tres aspectos importantes:

- C 1. - Flujo de autorización, que trata de los casos y los modos en los que la autoridad de un sujeto se ve influida por activaciones del servicio que actúa en su lugar.
- C2. -Flujo de sensibilidad, aborda la sensibilidad de los mensajes en el sistema.
- C3. -Restricciones del flujo de información, que se aplican dinámicamente con el fin de asegurar que la información, una vez “desencapsulada” del objeto, sigue sin estar expuesta a accesos no autorizados.

4.1.4 Metodologías para diseño de bases de datos seguras

Aunque existen algunas propuestas para extender el modelo E/R con el fin de poder especificar el nivel de seguridad de las entidades y atributos de una base de datos, véase [7]; la aportación en estos momentos más importante, a nuestro juicio, la constituye la metodología. Esta metodología, que como su nombre indica esta basada en OMT de Rumbaugh, consta también de tres fases: análisis, diseño del sistema y diseño del objeto.

En la fase de análisis se describe el sistema desde tres puntos de vista:

- El modelo de objetos que representa los aspectos estructurales y que pretende controlar ciertos tipos de inferencias no autorizadas.
- El modelo dinámico que representa los aspectos de control de las aplicaciones.
- El modelo funcional que representa los aspectos transformacionales.
- Durante la fase de diseño del sistema se diseña la base de datos multinivel y, durante la fase de diseño de objetos se determinan los detalles del sistema.

4.1.5 Bases de Datos Difusas

El problema que plantea es la información desconocida/incompleta y su representación por medio de valores nulos en las bases de datos. Dentro de la misma línea que pretende tratar con datos y consultas imprecisas, también se han introducido los denominados Sistemas de Gestión de Bases de Datos Relacionales Difusos (FRDBS), basados en la teoría de conjuntos difusos.

Un conjunto difuso es un conjunto de elementos en el que cada uno tiene un valor (entre 0 y 1) que indica el grado de pertenencia al conjunto. Así, se puede considerar que los valores de los atributos en un dominio o las tuplas en una relación tienen asociado un grado de pertenencia.

Aunque a alguno le suene a ciencia-ficción, este tipo de bases de datos resultan muy útiles ya que, “casi toda la información que manejamos acerca del mundo real es incompleta, imprecisa, incierta o vaga”.

Otra aplicación de estas teorías, puede ser la integración de consultas difusas en SGBD tradicionales (precisas), proporcionando una gran flexibilidad y superando el carácter restrictivo de los lenguajes de consultas actuales, [3].

Este tipo de sistemas necesita, como es lógico, ampliar los fundamentos teóricos de modelos como el relacional, a la vez que hace imprescindible extender los lenguajes, basados en cálculo o en álgebra relacional.

Capítulo 5

El Camino Hacia la Distribución, Integración e Interoperabilidad

5.1 Introducción

En los últimos años asistimos a un avance espectacular de las comunicaciones. Así, por ejemplo, se van reemplazando las redes Ethernet de 10Mbps por redes FDDI de 100 Mbps, a la vez que se despliegan redes ATM. Cuando esta tecnología se combina con la de bases de datos, aparecen nuevos retos y dificultades que deben ser superados si se pretende construir sistemas de información eficaces y eficientes de cara al futuro.

Por ejemplo, un “fenómeno” que está suponiendo una auténtica revolución incluso en la forma de hacer negocios y de concebir los sistemas informáticos lo constituye la difusión de las “infopistas” (Internet y el WWW, World Wide Web). Varias empresas ya diseñan sus servidores de información teniendo en cuenta el contexto de Internet/Web.

Por otro lado, en [6] se destaca la necesidad de investigar en tecnologías de integración, especialmente:

- Arquitecturas cliente/servidor.

- Adquisición de la semántica de los datos, que permita reconciliar las diferencias entre definiciones de distintas bases de datos.
- Calidad de datos, se ha hecho más énfasis en la calidad de los programas de aplicación, o en la construcción que en la calidad de los datos.
- Semántica de evolución, las bases de datos autónomas evolucionan de forma independiente no sólo su contenido sino también su semántica.

Existen también entornos distribuidos que demandan cada día mayores prestaciones, existiendo ya sistemas que necesitan replicación en varias decenas de miles de nodos. Ello lleva consigo que sea necesaria la administración y la operación remota de bases de datos dispersas, y que se deba lograr la integración con entornos de administración y redes (SNMP).

Por último, cabe destacar la aparición de un nuevo paradigma denominado “informática móvil” o “computación nómada” (piénsese la difusión que han experimentado los teléfonos móviles) que obliga a considerar a usuarios que no están permanentemente conectados a la base de datos, sino que se conectan de forma ocasional y que, por tanto, no son tan “controlables” como en bases de datos tradicionales.

5.2 Bases de datos distribuidas

A finales de la década pasada ya empezaron a aparecer los primeros productos de bases de datos que soportaban capacidades de distribución, entre sistemas del mismo fabricante en un principio, y que luego se han ido extendiendo hasta abarcar a diversos tipos de SGBD.

Temas como la optimización de consultas que tiene como objetivo, en un entorno de bases de datos distribuidas, maximizar el paralelismo y minimizar el tráfico de la red, fueron ampliamente investigados a nivel teórico en los ochenta y principios de los noventa. Otro aspecto importante fue la implementación de la gestión de transacciones mediante el protocolo 2PC (Two -Phase Commit), y la aparición de algoritmos avanzados de gestión de datos replicados.

Otro avance para los SGBD que funcionan en arquitecturas cliente/servidor (la mayoría según las últimas estadísticas) fue la incorporación de facilidades para crear procedimientos almacenados (que residen en la base de datos) y

que permiten disminuir el número de datos y mensajes intercambiados entre cliente y servidor.

En un primer momento el concepto de base de datos distribuida se centró en sistemas que permitían a los usuarios acceso transparente e integrado a una colección de bases de datos. Así se les exigía transparencia:

- De red (distribución).
- De replicación.
- De fragmentación.

Como una extensión natural al concepto de independencia entre datos y aplicaciones que marcó la arquitectura ANSI/SPARC.

Actualmente parece que la transparencia e integración (que nunca se han conseguido en su totalidad) pueden ser requisitos incompatibles con la autonomía y heterogeneidad, que van ganando un mayor peso. En efecto, podemos clasificar los SGBD de acuerdo a tres dimensiones, véase figura 5.1 de la página 42.

Como se puede observar en la figura 5.1 de la página 42 esas tres dimensiones son ortogonales, y permiten clasificar los distintos tipos de SGBD distribuidos.

Distribución: en la que se considera que los datos pueden estar distribuidos físicamente entre múltiples nodos, o bien almacenados en uno sólo (BD centralizadas):

- Autonomía: que se refiere al control de la distribución e indica el grado en el que un SGBD puede operar de forma independiente. Así, se puede hablar de un sistema altamente integrado (en el que los usuarios disponen de una sola imagen de la base de datos), o de un sistema semiautónomo, (en el que los SGBD operan de forma independiente pero que han sido diseñados para participar en una federación), o, por último, de un sistema completamente autónomo (multibases de datos).
- Heterogeneidad: a distintos niveles: plataforma (hardware, sistema operativo, protocolos de comunicación), SGBD (modelos y lenguajes), semántica de la base de datos (conflictos a nivel extensional e intencional), etc.

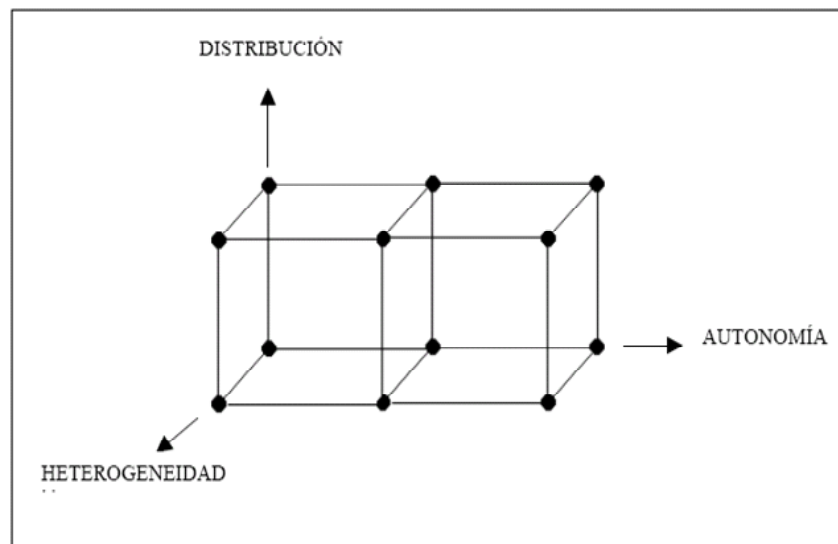


Figura 5.1: Distintas posibilidades de SGBD distribuidas.

Los productos comerciales existentes soportan principalmente SGBD distribuidos homogéneos y, en algunos casos, heterogéneos dentro del mismo modelo (relacional, por ejemplo, Oracle con DB2 y SYBASE) o entre diferentes modelos (por ejemplo, UniSQL, que permite integrar relacional y orientado a objetos). Los sistemas federados y de multibase de datos, se encuentran todavía en una fase de investigación, aunque existen ya diversos prototipos.

Ahora bien, dentro de los SGBD distribuidos, todavía quedan para los fabricantes varios retos por superar, que frenan en algunas ocasiones la difusión de este tipo de sistemas:

- Problemas de escalamiento de red. No existen demasiados estudios sobre el rendimiento de las bases de datos distribuidas sobre grandes redes. Falta modelos complejos que permitan simular el rendimiento teniendo en cuenta todos los parámetros necesarios.
- Diseño de bases de datos distribuidas. Aunque ya en 1983 Ceri y Pelagatti publicaron un libro sobre bases de datos distribuidas en el que se abordaba su diseño, todavía sigue siendo necesaria una mayor investigación, por ejemplo, en lo relativo al particionamiento mixto (vertical

y horizontal) y su relación con la ubicación de los fragmentos. También es necesario que estas técnicas se encuentren soportadas por las herramientas CASE.

- Procesamiento de consultas distribuidas, se necesita mejorar los modelos de coste para la optimización global de consultas.
- Procesamiento de transacciones distribuidas, por ejemplo, como señalan [10], en cuanto al escalado de los algoritmos de gestión de transacciones.
- Integración con sistemas operativos distribuidos ya que muchas veces el SGBD distribuido no sabe aprovechar los recursos que le brinda el SO distribuido; y, en el peor de los casos, estos dos sistemas se solapan e interfieren.

5.3 Bases de datos federadas y multibases de datos

Existen varias razones que han incrementado la necesidad de los SGBD federados:

- La sustitución de las prácticas tradicionales de procesamiento de datos por bases de datos.
- La proliferación de bases de datos heterogéneas en las empresas, en las que a veces cada departamento tiene su propio gestor de bases de datos.
- La compartición de datos de diversas bases de datos en la empresa.
- La consolidación de recursos software, hardware y de personal.
- La necesidad de mantener la autonomía de las bases de datos locales.

Este último requisito es clave en los sistemas de gestión de multibases de datos, y puede concretarse en:

- Autonomía de diseño: no realizar ningún cambio a los SGBD locales.
- Autonomía de ejecución: cada SGBD local mantiene un control total sobre la ejecución de transacciones en su localidad.

- Autonomía de comunicación: los SGBD locales no son capaces de coordinar las acciones de las transacciones globales que se ejecutan en varias localidades.

La solución más empleada hasta el momento para cubrir las necesidades mencionadas anteriormente consiste en implementar pasarelas (gateways) entre pares de SGBD, que se encargan de traducir las consultas en el lenguaje de un sistema al otro. Pero no se soporta la gestión de transacciones, control de concurrencia y recuperación de manera conjunta.

Cuando hay varios sistemas de bases de datos en uso, con diferentes modelos y lenguajes que tienen que coexistir, esta solución no suele ser la más adecuada. Además, hay que tener en cuenta la dificultad que supone para un usuario acostumbrado a un sistema centralizado, acceder a datos almacenados en distintos sistemas. Para ello existen dos soluciones ampliamente admitidas:

- Construir un “frontal” (front-end) sobre los sistemas existentes, que soporta un único modelo de datos y sólo un lenguaje de consulta.
- Crear una vista temporal con los datos pertinentes a la consulta del usuario.

La taxonomía de las diferentes soluciones para la compartición de información global, puede verse en la figura 5.2 de la página 45, que además de las características que se definió en el apartado anterior, tiene en cuenta el enfoque de diseño escogido, que puede ser:

- Esquema global. Esta opción sigue la filosofía de bases de datos distribuidas, y coloca un nivel global por encima de los esquemas externos locales, aunque a diferencia de las bases de datos distribuidas, este nivel global puede integrar esquemas de diferentes modelos de datos.
- Lenguaje de multibases de datos. En esta aproximación se transfiere parte de la responsabilidad de la integración a los usuarios, ofreciéndole nuevas funciones y otras extensiones a lenguajes que permiten manipular la representación de los datos.

En general, esta área todavía requiere una gran investigación, y podemos destacar los siguientes importantes desafíos:

TIPO	NIVEL DE INTERFAZ GLOBAL AL SGBD LOCAL	TIPOS DE NODO LOCAL	¿BD GLOBAL?	MÉTODO DE INTEGRACIÓN GLOBAL
MULTIBD GLOBAL	INT.USUARIO	BD HETEROGENEAS	SI	ESQUEMA GLOBAL
BD FEDERADA	INT.USUARIO	BD HETEROGENEAS	SI	GLOBAL PARCIAL
LENG. MULTIBD	INT.USUARIO	BD HETEROGENEAS	SI	FUNC. DE LENG.
LENG. MULTIBD HOMOGENEO	INT.USUARIO Y FUNCIONES	BD HETEROGENEAS	NO	FUNC. DE LENG.
SISTEMAS INTEROPERATIVOS	APLICACION SOBRE SGBD	CUALQUIER FUENTE DE DATOS	NO	NO INTEGRACIÓN GLOBAL

Figura 5.2: Taxonomía para la compartición de información global.

- Integrar sistemas que no soportan el concepto de transacción tal cual se entiende en los SGBD relacionales, como los sistemas CAD, e incluso los sistemas de recuperación de la información.
- Diseñar e implementar nuevos modelos de transacciones, como los de flujos de trabajo, en los que las dependencias entre tareas puede ser más compleja que la soportada por un modelo de transacción tradicional.
- Soportar la definición y comprobación de restricciones globales .
- Incluir diferentes tipos de datos multimedia, [9].

5.4 Bases de datos móviles

Estamos asistiendo a cambios tan revolucionarios en las comunicaciones como la expansión de las comunicaciones celulares, LAN (redes de área local) inalámbricas, servicios de satélites, extensión de los ordenadores portables (como los PDA, Personal Digital Assistant, palmtop, laptop, etc.) que ofrecen a los usuarios “móviles” la posibilidad de acceder a la información en cualquier momento y desde cualquier lugar. Se crea así un nuevo paradigma denominado por algunos “computación nómada”.

Algunos aspectos a tener en cuenta a la hora de diseñar e implementar SGBD móviles, son los siguientes:

- Desconexión. No hay que olvidar que los terminales móviles están a menudo desconectados y que esta desconexión no se considera un fallo como en los sistemas tradicionales, sino que, en todo caso, se podrían ver como “fallos planificados”.
- Pequeño tamaño y peso de los terminales, que entre otras cosas hace necesario buscar protocolos y algoritmos eficientes en “energía”, debido a las restricciones de baterías que presentan este tipo de equipos. Es imprescindible también llegar a conseguir un equilibrio entre memoria y disco, por ejemplo, las técnicas de compresión permiten ahorrar disco pero al descomprimir la información se consume CPU y, por tanto, energía.

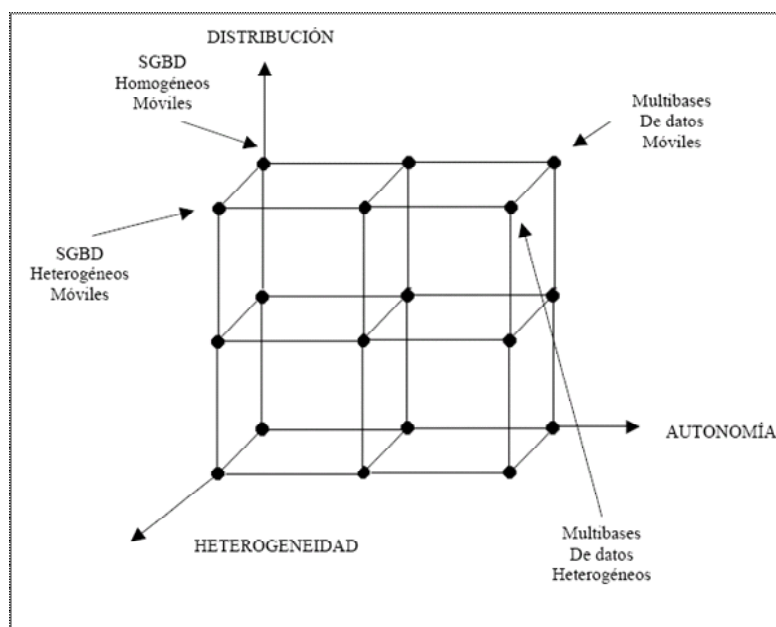


Figura 5.3: SGBD Móviles como extensión de los SGBD Distribuidos.

5.5 Bases de datos e infopistas

En la actualidad prácticamente todos los fabricantes anuncian sus productos como “La mejor bases de datos para Web”, a la vez que han aumentado sus

entornos de desarrollo siguiendo esta línea.

Sin embargo, todavía existen numerosos problemas asociados a la comunicación entre las bases de datos y la red, por ejemplo a la hora de pasar variables de formas HTML a procedimientos SQL. E incluso la Web ofrece un mayor alcance al usuario a través de un entorno heterogéneo, lo cual no siempre resulta fácil de gestionar.

La Web añade una nueva dimensión a tres bases de datos, aportando tres elementos:

- Nueva tecnología de IGU (Interfaces Gráficas de Usuario), mediante los visualizadores (browsers).
- Nuevo modelo cliente/servidor (protocolo HTTP).
- Un mecanismo de hiperenlaces entre bases de datos.

En principio, resulta evidente que los SGBD (Sistemas de Gestión de Bases de Objetos) soportan mejor los datos multimedia y la programación orientada a objetos, por lo que son más adecuados para trabajar en entornos Web, pero no siempre se muestran capaces de escalar y soportar las transacciones manteniendo un buen rendimiento.

Otra ventaja de los SGBD orientados a objetos es la posibilidad de que los clientes mantengan en memoria caché los objetos almacenados en la base de datos, lo que mejora considerablemente el rendimiento al pasar los datos directamente de la base de datos al servidor Web.

Uno de los temas que más atención recibe en los últimos años es el diseño de sistemas hipermedia con bases de datos. A este respecto, el profesor Tomas Isakowitz de la Universidad de Nueva York señalaba, en la conferencia BIWIT'97 celebrada en Biarritz , que existen tres posibilidades de diseño:

- Construir el entorno Web sobre la base de datos.
- Diseñar el Web para que interactúe con la base de datos.
- Diseñar ambos (Web y base de datos) de forma conjunta.

En estos últimos dos casos hace falta utilizar metodologías específicas de las que hay varias propuestas, y entre las que destaca la de este experto,

denominada RMM. En esta metodología, véase figura 5.4 de la página 49, además de las fases típicas (de estudio de factibilidad y modelado conceptual utilizando el ME/R) presentes en todo método de diseño de bases de datos, se modifican algunas fases y se añaden otras para cubrir específicamente la problemática hipermedial. Así, en el análisis de requisitos no se abordan sólo los relativos a la información, sino también los de navegación. Además, después del diseño conceptual en el modelo E/R se procede a estudiar de qué manera la información de las entidades se presentará a los usuarios, y cómo será el acceso a las mismas. Posteriormente se diseñan los caminos que permiten la navegación hipertextual, y se procede al diseño de protocolo de conversión, consistente en transformar cada elemento de los diagramas RMDM3 en objetos de la plataforma final. También se procede al diseño de la interfaz de usuario y al del comportamiento en tiempo de ejecución.

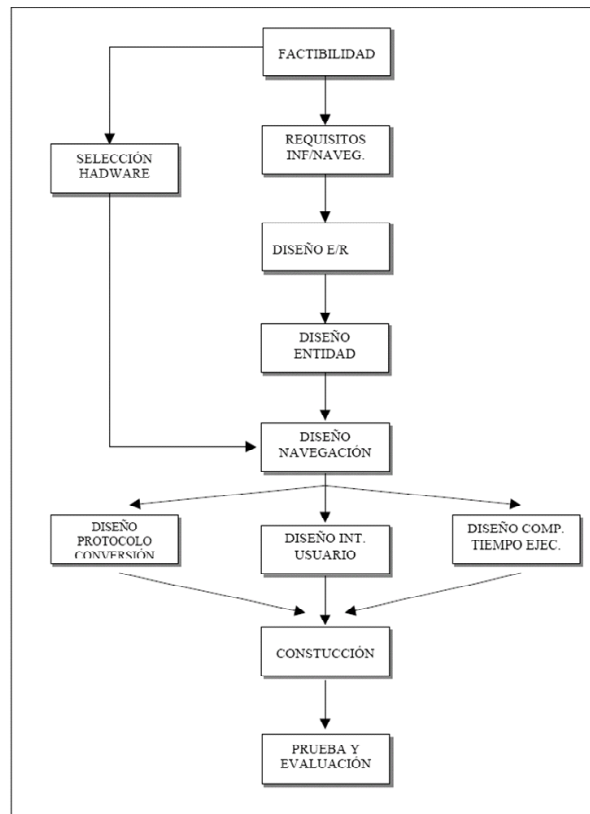


Figura 5.4: Metodología RMM.

Capítulo 6

Conclusiones

Las bases de datos se enfrentan por un lado a un entorno tecnológico cada vez más heterogéneo, y, por otro, a unas demandas empresariales más exigentes que determinan sus líneas de evolución.

Para afrontar estos nuevos retos, se requiere un mejor aprovechamiento de las facilidades que brindan las comunicaciones, así como la gestión de un mayor número de tipos de datos y una mayor flexibilidad en los modelos.

Los sistemas de bases de datos se encuentran difundidos actualmente por multitud de sistemas, existiendo claramente dos perspectivas diferenciadas,:

- SGBD abiertos, fáciles de usar y administrar y que constituyen la base adecuada para paquetes estandar.
- SGBD optimizados para diferentes plataformas, que proporcionan la ventaja competitiva y que forman la base para aplicaciones críticas.

Tanto en uno como en otro papel, la tecnología de bases de datos se ve sometida a numerosos cambios tanto desde el punto de vista empresarial como tecnológico.

Las nuevas aplicaciones están llevando hasta el límite a los sistemas de bases de datos disponibles, al incorporar documentos multimedia, imágenes, series temporales, datos activos, grandes cantidades de información (no olvidemos que los datos, al igual que los gases, se expanden hasta llenar el espacio disponible), etc.

Por otro lado la mejora espectacular en el número de instrucciones de máquina ejecutables en un segundo, coste de procesador, coste de la unidad de memoria secundaria y de memoria principal, número de bits transmitidos por unidad de coste y por segundo, obligan a los SGBD a evolucionar para aprovechar estos avances en el hardware y las comunicaciones.

Los sistemas de bases de datos han evolucionado hacia un mayor grado de funcionalidad, respondiendo a las necesidades de un nuevo tipo de aplicaciones que debe ser capaz de gestionar no sólo datos formateados, sino también imagen, vídeo, gráficos, estructuras espaciales, etc.

Otra dimensión que resulta fundamental incorporar en las bases de datos es la temporal, debiendo ofrecer el SGBD nuevos mecanismos que puedan soportar grandes volúmenes de datos sobre los que se puedan efectuar operaciones de tipo temporal.

También hay que tener en cuenta que existen ciertas aplicaciones, como en los entornos militares, médicos, etc. que presentan unos requisitos de seguridad muy fuertes, para los que los fabricantes ofrecen SGBD seguros.

Por otra parte, las bases de datos difusas ponen de manifiesto la necesidad de integrar técnicas de Inteligencia Artificial con bases de datos; parece que el camino indica que puede llegarse en un futuro no muy lejano a una teoría unificada, que se percibe como una necesidad en determinadas áreas.

Bibliografía

- [1] R. Cattell. *Object Database Management Group*. Diciembre 1991.
- [2] de Miguel Castaño; Piattini Velthuis. *Concepción y Diseño de Bases de Datos : del Modelo ER al Modelo Relacional*. Octubre 1993.
- [3] P. B. O. P. K. Farquhar. *Integrating Fuzzy Queries into an Existing Database Management System: An Example*. 1994.
- [4] Ghafoor. *Multimedia Database Management: Perspectives and Challenges*. Diciembre 1995.
- [5] K.Carey. *Storage management for Persistent Complex objects*. Octubre 1991.
- [6] S. Madnick. *Integration Technology : The Reinvention of the Linkage Between Information System*. Febrero 1993.
- [7] Smith. *From Operational to Denotational Semantics*. 1991.
- [8] M. S. O. S. H. V. Solms. *A Taxonomy for Secure Object-Oriented Databases*. ACM, Octubre 1994.
- [9] M. T. Özsu. . *An Object-Oriented SGML/HyTime Compliant Multimedia Database Management System*. ACM, 1997.
- [10] M. T. Özsu;Patrick Valduriez. *Distributed computing Systems: Separate Resources Acting as One*. . IEEE Computer Society Press Los Alamitos, CA, USA, Madrid-España, 1991.

