

INTRODUCCIÓN A LOS SISTEMAS DE BASES DE DATOS.

Evolución de los sistemas de bases de datos.

Uno de los usos más frecuentes de los ordenadores es el de almacenamiento masivo de datos. Aquí es importante hacer distinción entre datos e información. Cuando nos referimos a datos, estamos hablando de un texto, un número o un conjunto arbitrariamente grande de ambos. En el momento en que hablamos de información, nos referimos a datos que pueden ser relacionados entre sí, de forma más o menos inteligente, con el objetivo de sacar alguna conclusión que nos permita tomar decisiones importantes en nuestra empresa u organización.

El ordenador en sí no es inteligente, lo que pasa es que es capaz de hacer millones de operaciones simples en muy poco tiempo, con lo que es posible procesar un enorme número de datos para extraer información. ¿Cómo extraer información de los datos? Hay muchas formas: obtener estadísticas sobre destinos hoteleros preferidos, obtener movimientos turísticos según la procedencia, emitir facturas en una agencia de viajes, enviar cartas a nuestros mejores clientes, etc.

Para poder realizar todas estas operaciones es necesario almacenar de alguna forma datos sobre los clientes, los paquetes de viajes, las reservas, información estadística, etc. según la información que se desee obtener.

Tradicionalmente, antes del uso de ordenadores, la información se extraía tras un minucioso examen de los datos almacenados en fichas de cartón que se guardaban en los cajones de grandes ficheros en una oficina, lo cual era muy costoso debido al tiempo requerido, o bien a la poca exactitud de los resultados obtenidos, que hacía tomar decisiones poco acertadas.

Con la llegada de los ordenadores, y los dispositivos de almacenamiento electrónico, discos duros, tambores magnéticos, fichas perforadas, cintas magnéticas, etc., aparece el concepto de fichero electrónico. Un fichero electrónico es igual que un fichero con cajones y fichas de cartón en su interior, sólo que están almacenados sobre una superficie magnética (con el consiguiente ahorro de espacio), y son gestionados a través de un ordenador (con el consiguiente ahorro en velocidad a la hora de acceder), lo que posibilita su examen de forma veloz a través de la máquina.

Normalmente, los datos necesarios no se almacenan en un sólo fichero, sino que para extraer una determinada información, es necesario consultar ficheros de distinto tipo: ficheros de clientes, de pedidos, de proveedores, de ofertas, de facturas, de albaranes, etc. P.ej., para realizar un pedido de comestibles en un hotel, sería necesario consultar un fichero de existencias o stock, y el fichero de proveedores, realizar un pedido, y dejar constancia del mismo en el fichero de pedidos, para que, una vez recibido el pedido, nos cercioremos de que lo que se ha recibido es aquello que se pidió.

Este fue un gran paso en lo que respecta al almacenamiento y gestión de los datos. No

obstante, a medida que pasó el tiempo, se observó que la información extraída de estos ficheros era a menudo contradictoria y propensa a errores de coordinación (incoherentes): aparecen albaranes con números de clientes que no existen en el fichero de clientes, cada vez que es necesario indicar un proveedor en la ficha de pedidos hay que poner su nombre y dirección, con lo cual un mismo proveedor puede aparecer con varias direcciones distintas por error, etc.

Para solucionar este problema, interviene de nuevo la capacidad de los ordenadores para hacer rápidamente muchas cosas simples. Se optó por hacer que el ordenador controlase la concordancia entre los datos, y para ello se dio a los datos una estructura distinta, ya no basada en ficheros como tales, sino basada en listas o tablas de fichas (llamadas tuplas o líneas), y en cada una de las cuales no se encontraba toda la información, sino que era necesario relacionar datos de varias tablas para obtener la información deseada. Aparecen así las bases de datos.

La evolución de la informática ha pasado por numerosas formas de abordar los problemas mediante programas, en función de la potencia de las máquinas en cada época, y de la experiencia de los informáticos que ha ido aumentando a lo largo de los años. De esta forma, aparecieron las bases de datos jerárquicas y en red, cuyas características veremos más adelante. Posteriormente las bases de datos relacionales, que son las que nosotros estudiaremos, sustituyeron rápidamente a estas dos, ya que las relacionales se basan en estudios matemáticos que aseguran la eficiencia de las operaciones que permite realizar, así como su correctitud y completitud, y definen claramente la forma en que deben abordarse los problemas.

Actualmente, las bases de datos relacionales son las más usadas a nivel profesional, aunque para aplicaciones avanzadas y en los ambientes académicos también tienen bastante importancia las bases de datos orientadas a objetos (especialmente para trabajos sobre Internet), y las bases de datos deductivas (para trabajos en Inteligencia Artificial).

Ver el punto 2.1 de OBJETOS: Conceptos, métodos y herramientas. (Valduriez).

Objetivos a cubrir por una base de datos.

Los objetivos que debe cubrir una base de datos son:

- Evitar la redundancia e inconsistencia en los datos.

No sólo se debe aprovechar la enorme capacidad de los ordenadores para almacenar grandes cúmulos de información. Éstos pueden emplearse en controlar que los datos existentes sean consistentes entre sí, y que no haya información contradictoria entre los datos almacenados.

Por otro lado, es el usuario quien tiene que decidir qué se almacena en la base de datos y qué no. Evidentemente sólo deben almacenarse aquellos datos que se necesitan para algo, aquello que nos será útil en nuestro trabajo. El usuario es quien debe decidir además en que forma se almacenan los datos: si un número de teléfono se almacenará como un número, o como una tira de letras (lo cual permitiría guardar el guión que separa el prefijo del número en sí), etc. Como veremos, existen numerosas reglas que el diseñador

de bases de datos debe tener en cuenta a la hora de realizar un diseño eficiente y depurado. Evitar redundancias (datos o información repetida) en los datos almacenados es un objetivo prioritario ya que con ello se consigue:

- a) No desperdiciar capacidad de almacenamiento (cinta, disco, etc.).
- b) Evitar la extracción de una misma información por dos caminos distintos, con lo que la información extraída puede ser contradictoria.

- Facilitar el acceso a los datos.

En los sistemas antiguos, los datos almacenados en ficheros electrónicos sólo podían ser accedidos a través de programas de ordenador que realizaban alguna tarea específica muy concreta. Para cada tipo de información que se necesitaba, era necesario crear un programa que la extrajera de los datos de los ficheros; de esta manera, si un empleado de contabilidad deseaba saber los clientes de Ronda que habían comprado más de un millón de pesetas en el año 1.996, debía presentar dicha necesidad al departamento de proceso de datos, que debía crear un programa para atender tal requerimiento.

Evidentemente, esta forma de trabajo es lenta, y tiene un cuello de botella en el centro de proceso de datos, con el agravante de que da lugar a una miríada de programas distintos, cada uno para atender una necesidad particular. Además, si por cualquier causa, cambia la estructura de los datos (p.ej., se decide que en lugar de almacenar al completo el nombre y apellidos de un cliente, se guardará por un lado el nombre, y por otro los apellidos), algunos de los programas ya hechos no servirán, porque fueron creados en base a una estructura determinada, que ahora se ha visto alterada.

Para evitar todos estos problemas, se inserta entre el nivel a que se hallan los datos, y las aplicaciones, un programa auxiliar de máxima importancia: es el Sistema Gestor de Bases de Datos (S.G.B.D.), cuyas múltiples funciones iremos viendo a lo largo de la asignatura. Una de estas funciones es la de suministrar un lenguaje que permita consultar fácilmente la base de datos sin necesidad de hacer un programa para cada consulta; bastará con formular la consulta según este lenguaje simple, para obtener un resultado más o menos inmediato fruto de las gestiones realizadas por el S.G.B.D. sobre los datos.

- Facilitar la creación de programas que trabajen sobre los datos.

Como se ha indicado anteriormente, los datos pueden estar repartidos en múltiples tablas o en múltiples ficheros; incluso puede haber varias bases de datos de cuyo conjunto sea necesario extraer información. Es más, cada base de datos puede estar creada con un sistema distinto (es algo así como si en cada base de datos la ficha de cartón que guarda los datos de un cliente tuviera los datos cambiados de sitio, o incluso distintos datos).

De esta forma, es difícil crear aplicaciones que trabajen sobre los datos. De nuevo, esto se soluciona con un S.G.B.D. que aísla al creador de aplicaciones de este caos de datos, y le da una visión uniforme de los datos, facilitando de esta manera la programación de aplicaciones complejas.

- Permitir el acceso concurrente.

Un ordenador es una máquina que sólo es capaz de hacer una cosa cada vez: leer de memoria, escribir en ella, realizar un cálculo, etc.; en definitiva cosas muy simples. Sin embargo, en la mayoría de los trabajos se requiere que varias personas puedan trabajar a la vez, quizás aparentemente no con la misma máquina (cada persona tendrá un teclado y un monitor distinto, por supuesto; es lo que se llama un terminal), pero sí sobre los mismos datos; piénsese si no en un programa de reserva de billetes de avión de una compañía determinada. En este caso los datos sobre reservas se almacenan una sola vez en un ordenador central, que gestiona todas las operaciones efectuadas sobre cada terminal de usuario.

Así, cada usuario podrá efectuar peticiones sobre el ordenador central en el momento en que lo necesite, y la central deberá responderle en un tiempo prudencial (no debe hacer esperar innecesariamente). Dado que varias personas trabajan a la vez, es posible que el ordenador central reciba peticiones de acceso sobre los datos de forma simultánea: es lo que se llama acceso concurrente.

Para observar los problemas que puede producir el acceso concurrente, supongamos que el ordenador central es un escribiente al que cada usuario le dicta la ficha que debe tomar y lo que debe hacer sobre ella: consultar un dato de la misma, modificar algún campo, destruir la ficha o tomar una ficha en blanco e insertar datos en ella.

Si no hubiese requerimientos de tiempo por parte de cada usuario, no habría problema, pues al escribiente le bastaría con atender las peticiones por riguroso orden de llegada. Sin embargo, para dar la sensación de que el escribiente atiende a todo el mundo por igual, este hace uso de una técnica llamada tiempo compartido, o sea, reparte cada minuto de su tiempo en partes iguales, una para cada usuario que le ha hecho una petición.

De esta forma, pueden aparecer problemas en los accesos concurrentes a los datos. P.ej., ¿qué pasa si un usuario pide que se elimine la ficha del cliente Hotel Miramar, mientras que otro dice que se le modifique el apartado de correos?. Si hacemos primero la eliminación, ¿cómo modificar una ficha que no existe?!

Este problema se ve agravado enormemente en el momento en que se introduce el concepto de transacción. Una transacción es un conjunto de operaciones sobre los datos que debe ser tratada como una única operación: o se hace entera, o no se hace. Un ejemplo de transacción puede ser: Incrementar en un 2'8% el crédito máximo dado a cada cliente. Si tenemos varios miles de clientes, ésta es una operación que puede durar bastante tiempo, y mientras se produce puede que sea necesario atender otras peticiones. Por la estructura propia de la transacción (o se ejecuta al completo, o no se hace nada), es necesario hacer uso de copias de cada ficha de cliente afectado (de manera que si se produce algún problema, se destruyen todas las copias, pero los originales quedan intactos). ¿Qué ocurre si mientras se gestionan las copias de las fichas aparece una petición de cambiar el crédito máximo del cliente Bar Juamig? ¿Se le aumenta también el 2'8%, se pone el nuevo valor, o se deja como estaba? Todo esto debe ser gestionado también por el S.G.B.D., que debe asegurar que el resultado final debe ser consistente, y

debe informar al usuario afectado si se ha producido algún error en la transacción que intentó realizar.

- Controlar la seguridad en los accesos a los datos.

En los antiguos sistemas de fichas de cartón, los ficheros más importantes se hallaban fuera del alcance de la mayoría de los usuarios, y sólo podían tener acceso algunos privilegiados. Este método de seguridad tomaba muchas formas, según el grado de seguridad con que se quería dotar a esos datos: puertas con cerrojo, tarjetas magnéticas, guardas de seguridad con porras disuasorias, e incluso encriptación de los datos más importantes, al más puro estilo del agente 007.

Cuando se trabaja con ficheros electrónicos, el ordenador que los almacena, como se indicó en el apartado anterior, suele recibir peticiones de otros terminales, con lo que el acceso a los datos no está restringido sólo a los que tocan el ordenador directamente, pudiendo ser utilizado desde un terminal remoto situado, posiblemente, en cualquier parte del mundo. Por tanto los métodos de seguridad tradicionales no son suficientes (aunque siguen siendo necesarios con objeto de, p.ej., evitar la destrucción malintencionada del ordenador central por parte de algún saboteador industrial, o por algún hecho fortuito), y es necesario recurrir a medios electrónicos de protección que impidan acceder a los datos más importantes, excepto a aquellas personas que estén autorizadas.

P.ej., el departamento de pedidos no tiene por qué tener acceso a las nóminas del personal de la empresa, datos que estarán restringidos al personal contable de la misma.

Para conseguir esto, se asignan una serie de prioridades a cada usuario, así como un nivel de seguridad a cada tipo de dato, de manera que cada usuario sólo puede acceder a los datos de nivel inferior o igual a su prioridad. Además, existe un registro electrónico que toma nota de los accesos a los datos de mayor nivel, dejando así constancia de la fecha, hora y usuario que efectuó el acceso, o intentó un acceso frustrado.

Todo esto es llevado a cabo por el S.G.B.D. y debe ser inspeccionado regularmente por el departamento de proceso de datos, en busca de intentos de violación del sistema.

- Asegurar la integridad de la base de datos.

Este es uno de los aspectos más importantes, y que hacen uso intensivo de la capacidad de los ordenadores para procesar la información.

Los datos deben poseer unas características determinadas en función de lo que se pretende hacer con ellos, y del área de trabajo en que se desarrolla la empresa. Nos referimos a características que hacen que los datos almacenados representen una información coherente para el usuario que los maneja. Es lo que se llama la Integridad de la base de datos.

P.ej., se debe impedir que en el NIF de un cliente se coloque su Nombre, o que su Nº de teléfono aparezca vacío, o incluso cosas más complejas que relacionen varios ficheros, como p.ej. que haya un pedido pendiente a proveedor, y dicho proveedor no exista en el fichero de proveedores, o que haya una reserva de viaje en un vuelo que no

existe, o incluso que el número de reservas en un vuelo a Bangkok supere el overbooking máximo permitido (el 10%).

Todas estas son restricciones de integridad que el S.G.B.D. debe mantener en la base de datos, permitiendo, eso sí, en los casos en que sea necesario, una situación momentánea de falta de integridad, como p.ej. insertar una reserva por parte de un cliente inexistente, siempre y cuando el usuario asegure que va a dar de alta en breve a dicho cliente.

Ver Silberschatz, capítulo 1.

Definición de base de datos.

Existen numerosas definiciones de lo que es una base de datos, y de lo que es un S.G.B.D. De hecho, a menudo se confunden ambos términos dada la relación biunívoca existente entre ambos.

La definición más aceptada dice algo así como:

«En esencia, una base de datos no es más que una colección de información que existe a lo largo de un período de tiempo, a menudo de varios años. Más claramente, el término base de datos se refiere a una colección de datos gestionada por un Sistema Gestor de Bases de Datos, S.G.B.D., o simplemente Sistema de Bases de Datos». Ver Ullman: A first course in database systems.

Silberschatz extiende un poco más el concepto de base de datos, incluyendo en el término a los programas para acceder a estos datos, aunque no deja claro si son sólo los programas componentes del S.G.B.D., o incluye también a los programas de aplicación, hecho que, otros autores, como Julio Carmona, dan por sentado, aunque con mucha menos rigurosidad.

Nosotros consideraremos como base de datos a la colección de datos junto con el S.G.B.D.

Idealmente, el S.G.B.D. debe poseer una serie de características indispensables para satisfacer a los usuarios:

- Debe poseer un lenguaje de definición de datos que permita fácilmente la creación de nuevas bases de datos, así como la modificación de su estructura.
- Debe poseer un lenguaje de manipulación de datos, que permita la inserción, eliminación, modificación y consulta de los datos de la base, de la forma más eficiente y conveniente posible.
- Debe permitir el almacenamiento de enormes cantidades de datos (miles de millones de caracteres), sin que el usuario perciba una degradación en cuanto al rendimiento global del sistema.
- Debe permitir la gestión segura de los datos, con respecto a accesos no autorizados, y a accidentes producidos por los dispositivos mecánicos o electrónicos que soportan los datos almacenados.
- Debe permitir el acceso simultáneo por parte de varios usuarios, impidiendo además, que dichos accesos concurrentes den lugar a datos corruptos o inconsistentes.

- Debe suministrar independencia física de los datos, que asegure que, sea cual sea la estructura de los datos en los dispositivos electromecánicos de almacenamiento, el usuario y las aplicaciones los percibirán siempre de manera uniforme y útil.

Estas características dan lugar a otra definición de base de datos, quizás más orientada a la moda actual en informática: el diseño orientado a objetos, que consiste en hacer programas viendo el mundo como un conjunto de objetos que se relacionan entre sí para conseguir un objetivo común. Según esta visión, podemos considerar una base de datos como «una organización coherente de datos permanentes, y accesibles para usuarios concurrentes». Esta definición lleva implícito el concepto de S.G.B.D., ya que es éste quien se debe encargar de almacenar los datos (y por tanto de hacerlos permanentes y no volátiles), de gestionar su integridad (y por tanto de que sean coherentes), y de controlar a los múltiples usuarios (permitiendo así la concurrencia).

La forma en que las distintas bases de datos comerciales y académicas abordan estas características difieren enormemente, no sólo por las técnicas utilizadas sino también por las aproximaciones o paradigmas con que se han desarrollado. Aunque en el tema 2 haremos un breve estudio de los tipos de bases de datos, nos centraremos exclusivamente en el tipo más extendido: las bases de datos relacionales, ya que tienen un formalismo subyacente que las hace muy potentes. Además, fueron desarrolladas hace ya bastantes años, y han evolucionado lo suficiente como para suministrar poderosas herramientas que hacen fácil su gestión. De hecho, todas las características que hemos visto que debe poseer un S.G.B.D., son suministradas a través de entornos e interfaces amigables y comprensibles que permiten un rápido aprendizaje de todas las funciones propias de una base de datos.

En contraposición, otro tipo de bases de datos, como las Orientadas a objetos, requieren que casi todas las funciones de creación de base de datos, manipulación, etc., se efectúen a través de programas, lo cual requiere un profundo conocimiento de las técnicas de programación, impropio de una carrera como la de Turismo.

Por otro lado, sistemas igual de evolucionados, como el jerárquico o en red, han caído en desuso, y su aprendizaje supone un esfuerzo que aporta más bien poco al estudiante que debe enfrentarse de inmediato ante un mundo de datos básicamente relacional.

Componentes de un sistema de base de datos.

Hasta ahora hemos visto una idea general de lo que es una base de datos, y la necesidad de que exista un gestor central que separe los datos en sí de los usuarios, y que facilite una serie de tareas, así como se asegure de que se cumplen determinados requisitos indispensables para el buen funcionamiento del sistema.

Ahora vamos a desmenuzar un poco más el concepto de base de datos, y vamos describir sus componentes básicos, justificando además la necesidad de cada uno de ellos.

Datos.

Efectivamente, una base de datos no tiene sentido sino está compuesta por datos. Lo que no está tan claro es la forma en que estos datos se deben disponer, qué datos se deben almacenar, y cómo los debe entender la máquina.

La disposición de los datos depende del ámbito de aplicación concreto en que se enmarque la base de datos. No es lo mismo una base de datos que almacene un dibujo vectorial de monumentos históricos que una base de datos para almacenar las reservas de clientes en un hotel. Lo que varía fundamentalmente en uno y otro caso, es la forma en que los datos se relacionan entre sí, y el tipo de accesos a la información que va a realizar el usuario.

Además, los datos deben disponerse de manera que las consultas sean lo más eficiente posible, evitando a la vez la existencia de datos duplicados que pueden dar al traste con la coherencia de la base de datos. De esta forma, no sólo se consideran datos aquellos que el usuario desea almacenar, sino toda estructura de apoyo que el sistema necesite para hacer más eficiente una consulta. P.ej., supongamos que trabajamos en un hotel cuya base de datos posee un fichero con todas las agencias de viajes que contratan nuestros servicios de alojamiento en alguno de sus paquetes de viajes. Supongamos que para facilitar los accesos más comunes tenemos las fichas de tales agencias ordenadas por nombre. ¿Qué ocurre si queremos acceder a todas las agencias de Egipto para mandarles una carta felicitándolas por el año nuevo musulmán?. Resulta inaceptable que miremos una por una todas las agencias de nuestro fichero, ya que esto nos llevaría un tiempo desorbitado.

La solución informática a este problema consiste en crear un fichero aparte cuyas fichas contienen sólo dos campos: País de la agencia, y Nombre de la agencia. Las fichas de este fichero estarán ordenadas por País, y dentro de cada país, por orden alfabético del Nombre de la agencia. Para acceder a las agencias de Egipto, no accedemos al fichero principal, sino que nos dirigimos a este fichero auxiliar ordenado por país, y de ahí podemos averiguar directamente cuáles son los nombres de las agencias de Egipto. A continuación nos vamos al fichero principal, accediendo sólo a las fichas cuyos nombres hemos extraído anteriormente del auxiliar y de estas extraemos la dirección, lo que permite, finalmente, el envío de las cartas. El fichero auxiliar es lo que se llama un índice del fichero principal.

La gran ventaja de este método, es que el mantenimiento del fichero auxiliar, cuando trabajamos con un ordenador, lo realiza automáticamente el sistema, lo que nos ahorra la tarea de tener que escribir explícitamente una ficha auxiliar cada vez que se inserta una nueva agencia en el fichero principal. De esta forma se obtiene una gran eficiencia en las consultas, sin que el usuario tenga que trabajar más: el ordenador trabaja por nosotros.

Pues bien, este fichero aparte, creado con el único objetivo de facilitar el acceso a la información, también lo consideramos datos, aunque deban ser gestionados automáticamente, e incluso algunos usuarios no tengan ni la menor idea de su existencia.

Por otro lado, es muy importante decidir qué datos son los que se van a almacenar. A

menudo ocurre que en una base de datos se almacenan las facturas con una antigüedad inferior a x años. El número de años que se desea almacenar una factura electrónicamente puede ser una decisión crucial dependiendo de la capacidad de almacenamiento de nuestra máquina y del volumen de facturas que se expidan anualmente.

Además, si casi no se necesita acceder a dicha información histórica, puede ser más interesante destruir electrónicamente esa información, y almacenarla exclusivamente en papel, con lo que el ahorro de espacio en disco puede contrarrestar el esfuerzo de hacer las consultas manualmente.

Contrariamente, cuando se crea la base de datos, pueden parecer innecesarios ciertos datos de un cliente, como p.ej. Fecha de nacimiento. Posteriormente, si trabajamos en una agencia de viajes, y creamos un paquete de viajes de aventura por la selva del Amazonas, puede interesarnos promocionarlo sólo a aquellos clientes menores de 35 años, con lo que nos damos cuenta de que habría sido buena idea almacenar la fecha de nacimiento. Aunque el S.G.B.D. nos permite modificar la estructura de la base de datos para almacenar este nuevo campo, introducirlo para todos los clientes es una tarea ardua y fatigosa.

Por último, también es interesante, por cuestiones de seguridad y aprovechamiento de los recursos del ordenador, decidir en qué formato se van a almacenar los datos: si se van a almacenar encriptados para que nadie los pueda copiar, o si se van a codificar o a comprimir de alguna forma para hacer que ocupen menos espacio.

Existe toda una teoría sobre compresión y codificación de datos, que puede hacer que, dependiendo de las características de los datos a almacenar, se requiera mucha menos memoria de la necesaria siguiendo los métodos convencionales de almacenamiento.

Metadatos.

La evolución de las bases de datos, ha hecho que los informáticos se den cuenta de determinados problemas, y que los hayan ido subsanando a medida que perfeccionaban sus productos. Uno de los problemas que se producían consiste en la propia evolución de la base de datos.

Desde el momento en que se crea una base de datos, hasta el momento en que se desecha porque se compra un sistema mejor, o se instala una nueva base de datos, la estructura de la base de datos (o sea, los datos que se deciden almacenar, y la estructura con que se almacenan) cambia a medida que cambian las necesidades sobre la información a obtener de la base de datos, de manera que puede ser necesario introducir nuevas fichas, nuevos campos en fichas ya existentes, eliminar campos, modificar lo que se almacena en determinados campos, e incluso reestructurar la base de datos entera, creando ficheros de apoyo, etc.

Dado que la base de datos que solucione unas necesidades concretas puede adoptar muchas formas posibles, es muy interesante el poseer algún lugar que indique al personal encargado de mantener la base de datos, cuál es el objetivo de cada dato particular almacenado en

la base, así como en qué aplicaciones es utilizado, y con qué propósito, si es un dato fundamental, o si puede ser omitido por el que introduce los datos, etc.

De esta forma, antes de modificar el esquema o estructura de la base de datos, el departamento de proceso de datos debe consultar esta información sobre los datos de la base, con cuidado para no cometer errores graves que repercutan sobre el buen funcionamiento de todo el sistema.

Esta información que el sistema guarda sobre los datos almacenados, es lo que se llaman metadatos. Es más, estos metadatos se almacenan como otra base de datos propiamente dicha, y puede ser gestionada y consultada como tal.

Estos metadatos suelen conformar lo que se da en llamar diccionario de datos.

El sistema gestor de bases de datos.

El sistema gestor de bases de datos (S.G.B.D.) ha sido ya introducido, y su importancia destacada en todas las características que debe poseer una base de datos.

Según Silberschatz, podemos considerar un S.G.B.D. como «un programa que proporciona la interfaz entre los datos de bajo nivel almacenados en la base de datos, y los programas de aplicación y consultas hechos al sistema».

Ya hemos indicado las características propias de una base de datos, la mayoría de las cuales recae precisamente sobre el S.G.B.D. Para dejarlas más patentes, las reunimos y ampliamos aquí:

- Interactuar con el Sistema Operativo. Como se ha indicado, el S.G.B.D. no es más que un programa. El Sistema Operativo es el programa principal que se encarga de controlar que el ordenador funciona bien, entre otras cosas permitiendo el acceso a los dispositivos de entrada y de salida, como el teclado, el ratón, el monitor, y los dispositivos de almacenamiento: el disco duro, las disqueteras, el CD-ROM, las cintas magnéticas, etc.

Así, para asegurar que no pasan cosas raras, el único que puede tocar estos dispositivos es el Sistema Operativo (S.O.). Dado que el S.G.B.D. necesita almacenar datos en el disco duro, p.ej., debe interactuar con el S.O. para poder acceder al disco duro y que allí se almacenen los datos que se quieren. Igualmente debe contactar con el S.O. siempre que desee recuperar algún dato de estos dispositivos.

- Mantener la integridad. Como ya se ha dicho, debe mantener las restricciones de integridad propias de la aplicación concreta que sea. P.ej., evitar que la edad de un cliente supere los 90 años.

- Mantener la seguridad. Evitar accesos fraudulentos a los datos, así como la extracción de información codificada.

- Permitir las copias de seguridad. Dado que un ordenador no es un sistema infalible, y puede romperse por causas propias (fallo de un circuito), o ajenas (aumento de la tensión en la red eléctrica), es posible que los datos almacenados por él lleguen a corromperse con la consiguiente pérdida de información y los problemas que ello puede acarrear a la

empresa.

Para evitar estos desagradables resultados, es buena idea el efectuar una copia de los datos a un dispositivo auxiliar de almacenamiento, pensado precisamente para guardar fiel copia del contenido de la base de datos en un momento determinado. Si los datos originales se destruyen, bastará volcar la copia sobre el disco duro del ordenador central, con lo que los datos volverán a tomar la misma forma que cuando se efectuó la copia. De esta manera, para que la base de datos recupere la forma que tenía en el momento en que quedó destruida, bastará con efectuar los cambios que se hicieron en el tiempo transcurrido desde la copia de seguridad que se acaba de volcar.

- Controlar la concurrencia. Como ya se explicó anteriormente, debe permitirse el acceso simultáneo a los datos por parte de varios usuarios, lo que conlleva numerosos problemas de coherencia y coordinación. El S.G.B.D. debe controlar que la información representada por los datos al final de cada acceso de usuario siga siendo consistente.

- Suministrar mecanismos que faciliten la interacción con la base de datos. Estos mecanismos suelen venir dados en forma de lenguajes de manipulación y definición de datos. Además, suministran independencia de los datos, en el sentido de que, a pesar de la evolución del esquema de los datos, las aplicaciones deben sufrir las mínimas modificaciones imprescindibles. P.ej., si las aplicaciones antiguas están pensadas para trabajar sobre números de teléfono expresados mediante dígitos, ¿qué ocurre si se decide cambiar todos los números a formato textual? En este caso, debe haber un mecanismo que oculte a las aplicaciones antiguas el nuevo formato de los números de teléfono, y le haga vez el formato antiguo; en definitiva, debe haber algo que suministre a las aplicaciones antiguas una visión ligeramente distinta de lo que hay realmente almacenado en la base de datos.

Usuarios de la base de datos. El administrador.

Hasta ahora, hemos entendido como usuarios a aquellos individuos que esperan obtener algún resultado de la base de datos. Sin embargo, también podemos considerar usuario a toda persona encargada del mantenimiento y puesta a punto de los datos. De esta forma, podemos dividir a los usuarios en dos grandes grupos:

- a) Administrador. Es aquella persona o conjunto de personas encargada (exclusivamente, en la mayoría de los casos) del control global de la base de datos. Sus tareas son:

- Definición del esquema: Como se ha indicado, el esquema de una base de datos está compuesto por el tipo de fichas que almacena, el formato que toman cada una de ellas, así como sus características y reglas para ser rellenadas.

El administrador es, pues, quien decide qué se va a almacenar, y cómo. La información del esquema debe ser almacenada como metadatos, o sea, información sobre el formato de los datos almacenados.

- Definición de la estructura de almacenamiento y de los métodos de acceso. Indicamos

anteriormente que el S.G.B.D. interactuaba con el S.O. para poder acceder a los datos guardados sobre los dispositivos de almacenamiento, o sea en los ficheros. Pues bien, estos ficheros están a su vez estructurados en base a unidades más pequeñas con el objetivo de facilitar al S.O. el acceso a los datos. Es tarea del administrador el decidir las características de estas unidades mínimas de almacenamiento a fin de agilizar la recuperación de los datos, y de mejorar los accesos en función de las características propias de las consultas más frecuentes por parte de los usuarios.

También en base a las consultas más frecuentes, el administrador debe decidir sobre la necesidad de mantener estructuras de apoyo que hagan más fácil la recuperación de determinados datos. Cuando hablamos de estructuras de apoyo, nos referimos a los índices que veíamos en un punto anterior. Creando los índices apropiados, la velocidad de determinadas consultas puede aumentar vertiginosamente, incrementando así la satisfacción de los usuarios en el manejo de la base de datos. Por otro lado, veíamos que los índices son mantenidos automáticamente por el sistema, lo que puede producir ralentizaciones ante algunas operaciones de inserción, eliminación o modificación de datos, sin que aparentemente haya motivo alguno. Además, un índice ocupa un espacio cuyo tamaño depende del del fichero original, por lo que la creación de excesivos índices puede requerir una cantidad de memoria nada proporcional a los beneficios obtenidos de su utilización.

- Modificación del esquema y de la organización física. A medida que la empresa que utiliza la base de datos evoluciona para adecuarse a las condiciones cambiantes del mercado, es necesario adecuar dicha base de datos a las necesidades cambiantes que se producen. De esta forma, debe ser posible alterar el esquema de los datos, así como la forma de almacenamiento para hacer que las nuevas necesidades sean satisfechas de forma eficiente.

Asimismo, es interesante que las aplicaciones antiguas sigan funcionando con un mínimo de modificaciones, por lo que es de desear que, a pesar de dichas modificaciones, las aplicaciones antiguas tengan una visión alterada de los datos, pero adecuada a su forma de funcionamiento. Por supuesto, estas visiones, o vistas, deben ser coherentes con los datos almacenados en el esquema nuevo, y no tergiversar el contenido de los datos.

- Concesión de permisos y privilegios para el acceso a los datos. Asignación de prioridades a los datos. Para gestionar la seguridad en la base de datos, hemos comentado la necesidad de que los datos dispongan de una serie de niveles de importancia, de manera que sólo puedan ser accedidos por los usuarios con un privilegio superior o igual a dicho nivel. Pues bien, es el administrador quien se encarga de asignar dichos niveles y privilegios, en función de las características particulares de cada usuario. Asimismo, debe asignar inicialmente las claves de acceso a los usuarios, ya que cada uno de ellos sólo puede acceder al sistema mediante una contraseña particular.

- Especificación de restricciones de integridad. Los datos almacenados en la base de datos guardan una relación entre sí, en función de las necesidades que pretenden cubrir, de manera que la información que se pueda extraer de ellos sea útil y coherente. Esta coherencia viene determinada por un conjunto de reglas o restricciones que los datos deben poseer; p.ej. ningún cliente debe tener pendientes de pago facturas cuyo montante supere el crédito máximo permitido. El

administrador debe indicar estas restricciones de integridad, así como modificarlas o flexibilizarlas a medida que evolucionan las necesidades de la empresa.

A menudo, estas restricciones se suelen considerar incluso como parte integrante del esquema, con lo que también quedan reflejadas en el diccionario de datos.

b) Otros usuarios. El resto del personal que interactúa con la base de datos puede considerarse simplemente como usuarios, aunque también podemos distinguir distintos tipos entre ellos:

b.1) Programadores de aplicaciones. suelen formar parte del departamento de proceso de datos. Reciben peticiones de otros usuarios, en forma de necesidades de acceso a los datos, y se encargan de escribir los programas que satisfacen dichas necesidades. Normalmente estos programas están escritos en un lenguaje de programación convencional (pascal, Basic, C, etc.), en el que se insertan órdenes especiales que es capaz de comprender el S.G.B.D. De esta forma el S.G.B.D. suministra los datos, y el lenguaje convencional (también llamado anfitrión porque alberga las sentencias reconocidas por el S.G.B.D.) los procesa, presenta al usuario, modifica, etc.

b.2) Usuarios directos. Son aquellos que interactúan con la base de datos haciendo uso directamente del lenguaje que proporciona el S.G.B.D. En posteriores capítulos estudiaremos el SQL como principal lenguaje de manipulación y definición de datos. El administrador debe tener cuidado con la prioridad que asigna a estos usuarios, ya que al tocar directamente sobre la base de datos, pueden alterar su estructura si disponen de contraseñas de alta prioridad, o si por error se les ha suministrado una prioridad demasiado grande.

b.3) Usuarios indirectos. Son aquellos que no saben nada de la base de datos, excepto que existe. Interaccionan con ella a través de las aplicaciones desarrolladas por los programadores, y son incapaces de acceder a los datos directamente a través del lenguaje del S.G.B.D. El administrador debe prestar especial atención a que este tipo de usuarios no acceda directamente a los datos por los problemas que su inexperiencia pueda acarrear.

Elementos de seguridad.

El administrador debe conocer en profundidad los elementos de seguridad que suministra el S.G.B.D. de que dispone, y sacar el máximo partido posible de ellos. En general, se pueden tener niveles de acceso clasificados por:

- Las fichas o tuplas a que se tiene acceso. En una empresa grande y ampliamente informatizada, cada usuario debe poder acceder exclusivamente a los datos que competen a su tarea. El departamento de contabilidad no tiene porqué acceder a la información sobre reservas de alojamiento, aunque sí a las peticiones de tintorería, para hacer los cargos correspondientes a cada cliente. A su vez, el recepcionista, sólo accederá a las reservas efectuadas, y podrá abrir una factura de alojamiento en la que se irán insertando por parte de cada sección del hotel los cargos de los

servicios prestados al cliente, pero no podrá tener en ningún momento acceso a las nóminas, ya que ello es responsabilidad de la habilitación.

De esta forma vemos que cada usuario sólo debe tener acceso a un tipo de información, o lo que es lo mismo, a un tipo de fichas determinado. Por tanto, debe existir un mecanismo de seguridad que restrinja el ámbito de acceso de cada usuario en función de sus competencias.

- Las operaciones que se pueden realizar sobre las fichas o tuplas. Siguiendo con el caso anterior, la sección de contabilidad tendrá acceso a las facturas a clientes, pero no podrá modificarlas, sino tan sólo consultarlas a efectos contables para generar el balance de cada mes. Por otro lado, las diferentes secciones de cafetería, restaurante, tintorería, sauna, joyería, etc. sí que deben poder acceder a las facturas de los clientes para realizar en ella los cargos pertinentes. Sin embargo, no podrán crear facturas nuevas, si sólo pueden atender a clientes que se alojen en el hotel.

Por tanto, no sólo es importante el acceder o no a los datos, sino también la forma en que este acceso se produce, en función de las características propias de la sección: la habilitación sólo puede consultar facturas de clientes; la recepción puede crear y modificar, así como cerrar una factura; la tintorería sólo puede añadir cargos; la joyería puede crear y añadir cargos, etc.

En general, las operaciones que se pueden efectuar sobre un tipo de fichas se agrupan en cuatro grandes bloques: Altas, Bajas, Modificaciones, y Consultas.

- El acceso al diccionario de datos y a la estructura de la base de datos. Como se ha comentado anteriormente, los metadatos almacenados en el diccionario de datos y que almacenan información sobre la estructura de la base de datos, son gestionados, a su vez, como si de una base de datos especial se tratase. Sin embargo, dada su primordial importancia, su acceso debe estar muy restringido, ya que cualquier modificación puede dar lugar a resultados desastrosos en la base de datos: pérdida de información, corrupción en los datos, falta de integridad, etc. Por ello, es necesaria la existencia de prioridades o privilegios especiales que sólo permitan el acceso al personal que compone la administración de la base de datos, que es el único capacitado para modificar estos metadatos.

Lenguajes de bases de datos.

La interacción del usuario con la base de datos debe efectuarse a través de alguna técnica que haga fácil la comunicación, y que permita al usuario centrarse en el problema que desea solucionar, más que en la forma de expresarlo con las técnicas que se le suministran. La mejor forma de alcanzar este objetivo, es darle un lenguaje parecido al lenguaje natural, que le permita expresar de forma sencilla los requerimientos. En función de estos requerimientos, podemos tener, fundamentalmente dos tipos de lenguajes para comunicarnos con el S.G.B.D.:

- Lenguaje de definición de datos (LDD). Este lenguaje es utilizado en exclusiva por el administrador de la base de datos, ya que permite la construcción de frases o sentencias que le indican al S.G.B.D. las características particulares de la base de datos sobre la que se está trabajando, así como la creación de nuevas bases de datos. La creación de esquemas y su modificación, la creación y supresión de índices, la especificación de unidades de almacenamiento

en los ficheros, así como la asignación y privación de prioridades se realizan a través de este lenguaje. Además permite la creación y recuperación de copias de seguridad, así como la importación de datos desde sistemas antiguos, y la exportación a nuevos sistemas que vayan a sustituir al actual.

- Lenguaje de manipulación de datos (LMD). El lenguaje de manipulación de datos es el que usan los usuarios directos para efectuar sus operaciones sobre la base de datos. Como se indicó, estas operaciones son básicamente de inserción, eliminación, modificación y consulta de datos, aunque también se pueden introducir capacidades para crear visiones de los datos que faciliten otros accesos. Los usuarios directos interacciones con el S.G.B.D. a través de este lenguaje, mediante una interfaz agradable y fácil de usar.

Los programadores de aplicaciones emplean el LMD dentro de un lenguaje de programación que les da potencia expresiva, o sea permite sacar cosas por pantalla de una forma agradable, facilitar las operaciones de los usuarios indirectos, etc. Para ello, el LMD que emplean se extiende de diferentes formas para poderse integrar fácilmente en el lenguaje anfitrión, ya que ambos, LMD y lenguaje anfitrión deben poderse comunicar adecuadamente para que las aplicaciones resultantes sean simples de programar y de utilizar.