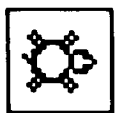


Estructura y organización de una base de datos

Guglielmo Trentin



Rx Ma

*Cualquier profesor inquieto que acumule datos para su trabajo sabe lo complicado de organizarlos y explotarlos bien. De hecho manejar bien la información es tan importante o más para la enseñanza y el aprendizaje como la propia información. Una de las herramientas del ordenador de uso generalizado, **las bases de datos**, constituye por tanto un recurso básico para el profesor, el alumno y el propio proceso educativo. Aquí se hace una presentación de posibilidades y problemas para emplearlo con eficacia.*

INTRODUCCION

Organizar la información en función de una rápida recuperación es indudablemente un factor determinante en muchos sectores de la actividad, sin olvidar las necesidades tanto en el ámbito de la decisión, como el consultar y analizar siempre mucha más cantidad de información; basta por ejemplo, con pensar en el problema cotidiano del que administra un almacén o una biblioteca. De ahí la utilidad de organizar la información en estructuras fácilmente administrables y que aseguren una elevada velocidad de acceso.

Existen factores muy interesantes en la fase de proyectar y realizar una base de datos, factores de notable interés conceptual sobre todo si se ve la planificación de la base de datos como el momento de ordenación y construcción de la realidad en términos de entidad y de su interrelación.

En este trabajo veremos la evidencia de estos aspectos conceptuales sin entrar en los detalles funcionales, y menos aún de los de implementación, de los sistemas específicos para la gestión de la base de datos.

QUE ES UNA BASE DE DATOS

Antes de entrar en consideraciones vamos a observar el análisis, el proyecto y la realización de una base de datos. Empezaremos por una definición.

Una base de datos es un conjunto de datos estructurados y permanentes agrupados por su homogeneidad y relacionados entre ellos, organizados con la mínima redundancia para ser usados en aplicaciones diversas, de modo controlado.

En la práctica una base de datos es un conjunto de archivos, convenientemente ordenados, que debe responder a las siguientes características:

- estar completamente integrado (no sólo juntar los datos sino ponerlos en relación entre ellos);
- asegurar la velocidad de acceso a la información;
- asegurar una completa gestión de los datos;
- permitir un acceso competente a la información;
- asegurar la privacidad de la información;
- asegurar la reconversión de los datos en caso de mal funcionamiento;

Seguidamente vamos a tomar en consideración sólo alguna de estas características, no queremos profundizar en los argumentos relativos al acceso competente, a la privacidad de la información y a la reconversión.

LA IMPORTANCIA DE SABER MODELAR LA REALIDAD

A menudo, cuando buscamos adquirir conocimientos más profundos de una realidad parcialmente conocida, se hace un modelo fijando la atención al aspecto que más interesa profundizar. Del estudio del modelo a menudo es posible obtener la información que nos permita tomar decisiones en función de una finalidad previamente fijada. Para comprender mejor este aspecto clave, escogeremos un ejemplo sencillo tomando como referencia el esquema expuesto en la fig. 1.

Supongamos que se debe averiguar la aerodinámica del prototipo de un coche; en este caso la realidad es el coche en sí mismo. Pero lo que interesa, al terminar el estudio de la aerodinámica, es el perfil del coche, eventualmente el material con que se construirá la carrocería, descuidando todos los aspectos que no estén asociados al objetivo que nos proponemos como peso total, el tipo de motor, el interior, etc.

Determinando los intereses particulares hemos hecho una *abstracción*, es decir hemos separado ciertos aspectos de la realidad cognoscitiva de los otros, que sin embargo conocemos. En efecto, en distintas circunstancias, cada uno de nosotros, frente a una cierta realidad, ejercemos una elección distinguiendo las características que nos interesan de las que no nos interesan y por consiguiente las ignoramos.

Una vez determinados los aspectos significativos de la realidad que se quiere moldear, nace la necesidad de *representarla*. Prosiguiendo con el ejemplo del coche, la representación consistirá en la realización de una maqueta en la que destacaremos únicamente las características relativas al perfil y posiblemente el material utilizado para construir el chasis.

Respecto a la representación en este momento ya podemos trabajar, o sea podemos ejecutar las oportunas *elaboraciones* en la maqueta para tratar de obtener la información que necesitamos (la maqueta del coche va a pasar por el túnel de viento y será expuesto a una serie de pruebas y medidas).

El resultado de la elaboración (turbulencias, inestabilidad,...) debe ser llevado al contexto real del cual se ha extraído el modelo (el coche que se debe construir); este proceso es en algún sentido la operación inversa de la abstracción, que llamaremos *interpretación*. En efecto, la interpretación de los resultados de la elaboración da la pauta para decidir cómo actuar en la realidad y por consiguiente, en definitiva, dar sentido al mismo modelo.

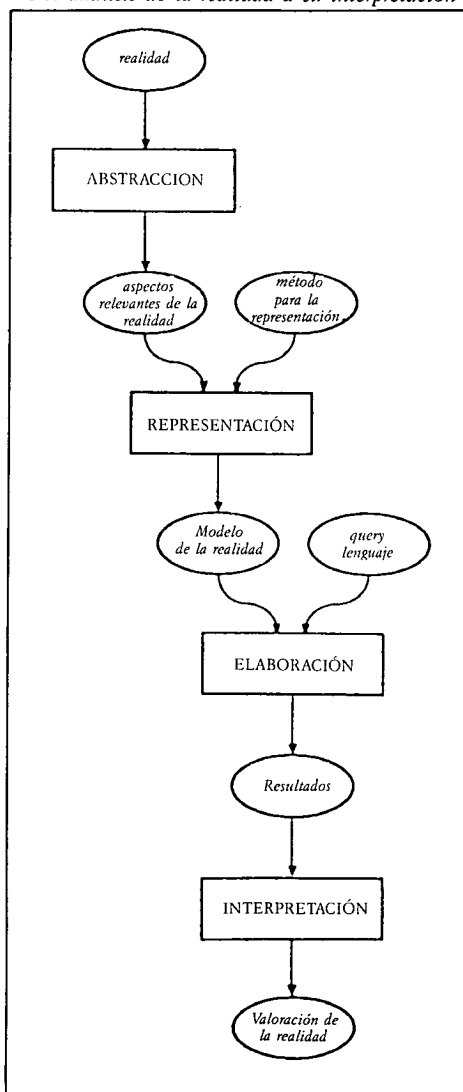
LA PROYECCION CONCEPTUAL

En la fase de proyección conceptual de una base de datos, el objetivo principal concierne al análisis de aquella porción de la realidad y de los acontecimientos que se verifican a ella, que sean de interés para el entorno de aplicación en el cual la base de datos debe poder ser utilizada.

El análisis de la realidad puede ser visto en dos direcciones: la primera, que llamaremos horizontal, atañe a la realidad considerada en su extensión, por consiguiente a todos los aspectos que la componen. La segunda, que llamaremos vertical, concierne en cambio a nuestra relación con la realidad considerada: lo que efectivamente de ella nos

FIGURA 1

Del análisis de la realidad a su interpretación



interesa. La actividad de análisis viene normalmente transferida a través de los procesos de unificación, de clasificación y de generalización.

La *unificación* consiste en reducir varios objetos en una sola unidad: una cabeza, dos brazos, dos piernas y un tronco forman la entidad hombre.

El segundo proceso, la clasificación, permite dividir los individuos en clases y por consiguiente dar una cierta estructura a la realidad. La clasificación además se puede hacer por extensión y por designación: en el primer caso se da explícitamente la lista de los individuos que pertenecen a la clase (perro, gato, topo...), en el segundo se determina la clase a través de la cual se especifica una propiedad que caracteriza a sus elementos (cuadrúpedos). Se observa que en la base de datos la clase (o identidad) siempre se da por extensión, en cuanto un archivo contiene exactamente los datos de varios individuos que componen la misma clase; al revés, la búsqueda de la información está hecha habitualmente por designación, o sea especificando una propiedad que caracteriza a los individuos que buscamos.

Finalmente, en lo que concierne al proceso de generalización, consideramos el siguiente ejemplo: una escuela X se la puede ver como un conjunto de clases y cada clase compuesta por distintos alumnos. Generalizando se puede decir que cada alumno, pertenece a una cierta clase y también es alumno de la escuela X. En el ejemplo resulta evidente que la propiedad más importante de la generalización es la capacidad de transmisión de las características.

Por lo tanto en cada análisis, la realidad se presenta como una colección de entidades conectadas entre ellas por una red de asociación. Utilizando una simbología muy sencilla (fig. 2) es posible obtener una formalización de la realidad llegando al denominado esquema conceptual; en particular, en este trabajo utilizaremos el modelo entidad-relación. En el esquema conceptual, ya sea por la entidad como por la asociación, vienen especificados los atributos, es decir, la propiedad y las características que nos interesan para el análisis de la porción de realidad a la cual se refiere (fig. 3). El atributo o los atributos que determinan unívocamente a una entidad se llaman «claves primarias»

LA ASOCIACION

Un aspecto relevante del esquema conceptual atañe a la tipología de la asociación entre entidades. Las asociaciones son la red de conexión de una base de datos y sólo tomaremos en consideración, aquellas que nos pueden aportar una contribución efectiva en la estructuración. En general las asociaciones pueden ser de dos tipos: uno-a-muchos (1:n) y muchos-a-muchos (n:m). Un ejemplo de asociación uno-a-muchos es aquel que asocia una clase de alumnos con los mismos alumnos

FIGURA 2
Simbología usada en el modelo entidad/relación

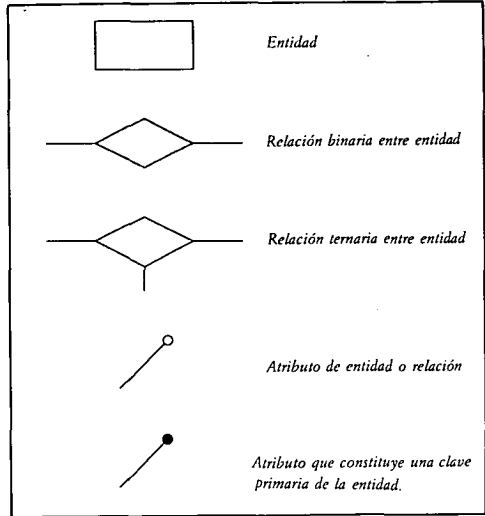
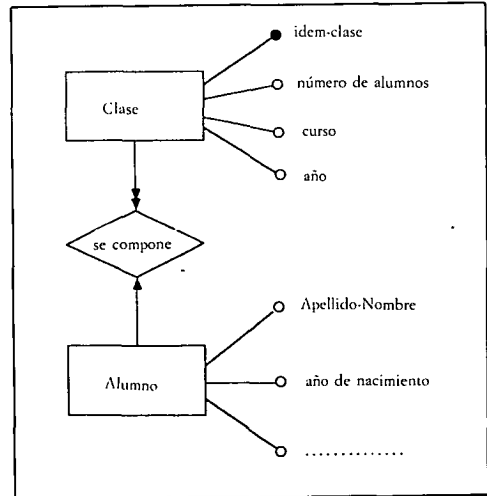
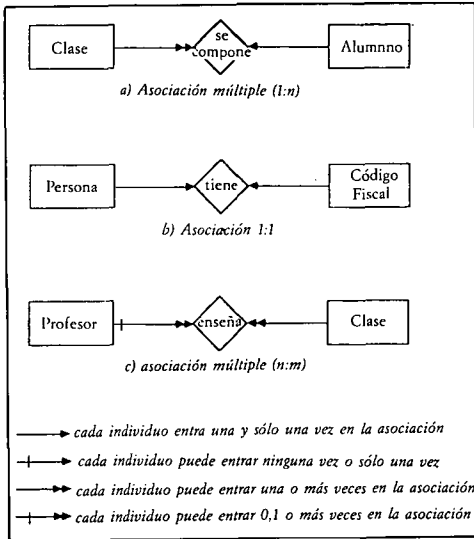


FIGURA 3
Los atributos de la entidad



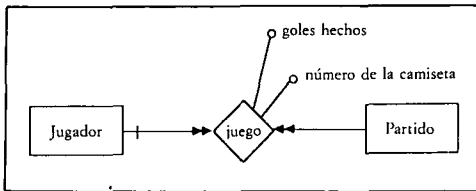
(fig. 4,a): en efecto muchos estudiantes están asociados a la misma clase, mientras un estudiante puede formar parte de una sola clase. Se observa que la asociación uno-a-uno (1:1) es una particularidad de la una-a-muchos; un ejemplo es la asociación entre la entidad persona y la entidad código fiscal (fig. 4,b): cada persona se la puede asociar a un solo código fiscal y cada código fiscal se refiere a una sola persona. Al revés, un ejemplo de asociación muchos-a-muchos es aquel que asocia la entidad maestro de una escuela a la entidad clase (fig.4,c): un maestro puede enseñar a varias clases y cada clase puede tener muchos maestros.

FIGURA 4
Tipo de asociación



Tanto para el valor, como para la asociación es posible definir atributos. Existen, en efecto, situaciones en las cuales ciertos atributos no son propiedad separada de una de las dos entidades en relación, pero son características de la asociación en cuanto a tal. Por ejemplo, si nos referimos a la situación de la fig. 5, en la cual viene ilustrada la asociación entre partidos del campeonato de fútbol y jugadores de los diferentes clubs, es evidente que los atributos «goles hechos» y «número de la camiseta» no son propios de la entidad, dato que puede variar según el encuentro que se disputa.

FIGURA 5
Los atributos de la asociación



Hemos visto hasta ahora cómo el proceso de formalización conduce al esquema conceptual. Ya sea el esquema conceptual o en índice de funcionalidad que está realizado sobre él, son el productos de la actividad de abstracción, y por esto la base de partida de la actividad de representación (fig. 1).

Como en un modelo físico (véase el ejemplo del coche) la representación de la realidad está asociada a los materiales disponibles y al modo de como éstos se pueden plasmar, así en un modelo

informático debemos tener en cuenta la estructura de la elaboración, su capacidad y la tecnología informática de que disponemos hoy. También la base de datos, en cuanto modelo informático, no escapa de este tipo de dependencia y su representación está íntimamente conectada a problemas de tipo tecnológico.

En síntesis se puede decir que la representación de una base de datos está en función de la tecnología, es decir el modelo adoptado para estructurar y administrar la información contenida en la misma base de datos, en una palabra, de la Base de Datos Management System (BDMS). En efecto, se dice que el esquema lógico de una base de datos, concepto que revisaremos en seguida, tiene en cuenta el modelo de BDMS adoptado.

Para comprender mejor las diferentes tipologías del BDMS, creemos oportuno abrir un paréntesis referente al modo de cómo los datos pueden estar interrelacionados entre sí en una estructura. Efectivamente, es la naturaleza de esta relación la que determina una estructura física distinta de la base de datos.

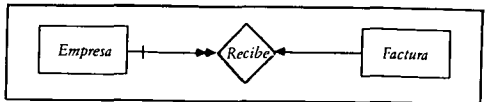
EL CONECTOR

Para relacionar la entidad de una misma base de datos, normalmente se acostumbra usar dos tipos de referencia: el conector gráfico y el conector lógico.

Vamos a analizar estas dos técnicas distintas con la ayuda de un simple ejercicio. Supongamos que una firma X decide organizar una base de datos propia articulándola en dos archivos (entidad): el archivo EMPRESA en donde se registra la información relativa a la empresa a quien la firma X vende ciertos productos y el archivo FACTURA, en donde se registra las facturas emitidas por la firma X. A esta situación corresponde el esquema conceptual de la fig. 6.

FIGURA 6

Relación entre factura y empresa a la cual se le ha emitido



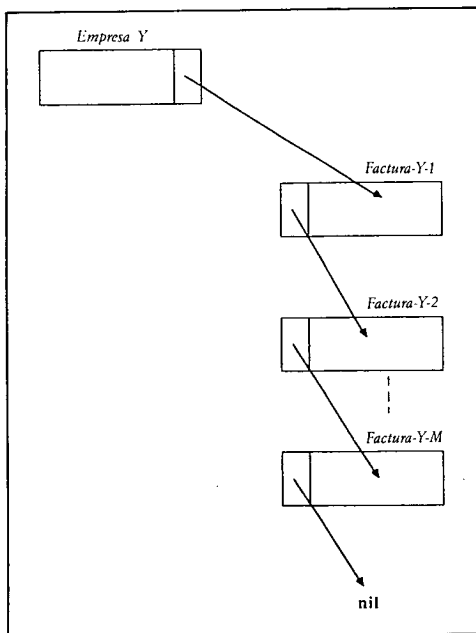
Vamos a ver ahora dos maneras distintas de relacionar entre sí esta información.

Uso del conector gráfico. La situación prevé que cada elemento (récord) del archivo EMPRESA, además de contener información de una empresa determinada Y (atributo), contenga también información acerca de la posición física de la primera factura, emitida de X a Y, registrada en el archivo FACTURA. A su vez los elementos del archivo FACTURA poseen, más información de la que contiene una factura normal, la referencia a la po-

sición física de las sucesivas facturas emitidas de X en comparación con Y (fig. 7).

En esta manera de referenciar los elementos de la base de datos, el archivo EMPRESA se denomina owner, mientras el archivo FACTURA member: esto da lugar a que se defina una información primaria (aquella contenida en EMPRESA) relacionada de algún modo a información secundaria (la que está contenida en FACTURA). La decisión de definir un valor owner o member depende del tipo de operaciones ejecutadas más frecuentemente en la base de datos.

FIGURA 7
Conector gráfico



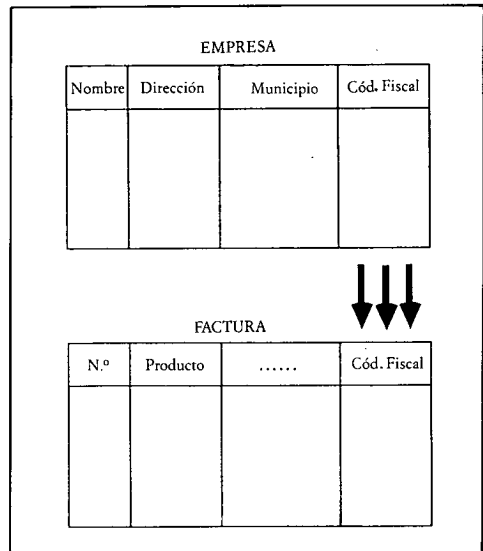
En general, para acelerar el acceso a la información, el owner está indicado: en nuestro ejemplo la indicación podría hacerse en el atributo Nombre Empresa. En el ejemplo se parte de la hipótesis de que la pregunta típica sea: dada una empresa Y extraer todas las facturas que se le han estado emitiendo. Obviamente si las operaciones más frecuentes fueran la extracción de la empresa a quien corresponde una cierta factura, podríamos haber definido como owner la entidad FACTURA y member la entidad EMPRESA.

Última consideración: el ejemplo se refiere a la versión más sencilla de enlace mediante el conector gráfico. Son, en efecto, soluciones en las cuales el vínculo entre los elementos de la información secundaria es de tipo bidireccional. De todos modos, ya sea la gestión de los vínculos unidireccional o bidireccional, y la gestión del conector a la posición gráfica de los elementos de

un archivo, son actividades totalmente transparentes al usuario y se gravan únicamente en la implementación del modelo de BDMS que se escogerá.

Uso del conector lógico. En este caso no tenemos creados vínculos físicos entre los elementos de la entidad pero se juega con sus atributos comunes. En el ejemplo se ha utilizado el atributo Código Fiscal de la empresa repetido en la información que caracteriza la entidad factura (fig. 8). En el momento de interrogar será la BDMS la que pondrá en relación las dos entidades a través de la clave de lectura Código Fiscal.

FIGURA 8
Conector gráfico



Modelo de base de datos

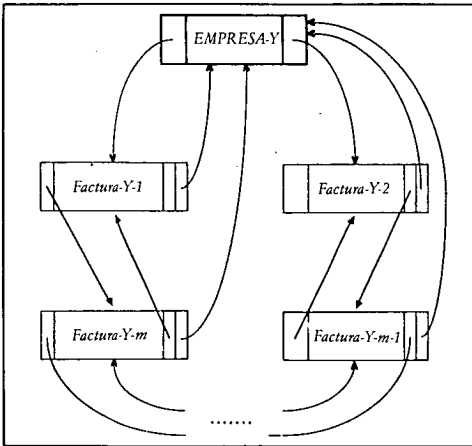
Al llegar a este punto ya se habrá intuido que uno de los elementos que caracteriza el modelo de una base de datos es la técnica usada para realizar el conector. Normalmente los modelos están clasificados en sentido jerárquico, reticular y relacional. Seguidamente vamos a examinar sólo los dos últimos modelos dado que el jerárquico no es más que una particularización del modelo reticular.

Modelo reticular

El modelo reticular adopta la técnica del conector gráfico para relacionar los elementos de la base de datos. En él la información está estructurada según una red de archivos conectados entre sí en relación 1:n. En efecto la estructura es muy parecida al esquema conceptual: la entidad y la asociación representada en el esquema conceptual, que se convierte en el record type y el ret respectivamente.

En general el modelo se basa en una estructura de dos archivos de los cuales uno, llamado *master*, sirve para el acceso a la información primaria, y el otro, llamado *variable*, para el acceso a la información secundaria. Utilizando siempre el ejemplo de la empresa y de la factura, la fig. 9 muestra la interrelación entre los elementos de las diferentes entidades.

FIGURA 9
El modelo reticular



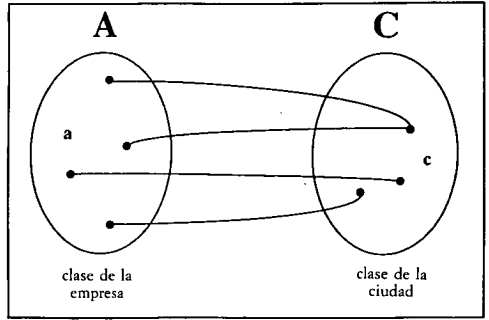
El hecho de utilizar enlaces bidireccionales ya sea a nivel de elementos de una misma entidad (Factura-Y-1, Factura-Y-2) que a nivel de elementos de entidad diferente (Empresa-Y, Factura-Y-1) garantiza un acceso rápido a los datos a pesar de que partiendo de una información secundaria (Factura-i-j), se quiere remontar a la correspondiente información primaria (Empresa-i a quien se le ha facturado).

Modelo relacional

El modelo relacional adopta la técnica del punteador lógico y tiene un enfoque rigurosamente matemático. Para comprender mejor el concepto de relación vamos a hacer un ejercicio muy simple. Si observáis la fig. 10, en donde A identifica la clase de Empresa italiana y C la clase de la Ciudad italiana.

Para indicar que la empresa FIAT tiene su sede en Turín, se puede escribir usando una anotación muy parecida a la matemática *tiene-sede-en*(FIAT, TURIN)
 Si consideramos la clase entera de la empresa $A = (a_1, a_2, \dots, a_n)$
 Y si a cada una de ellas asociamos la ciudad relativa, obtenemos una clase de pareja.
 $C = 8(a_1, c_1), (a_2, c_2), \dots, (a_n, c_n)$
 que concreta la relación «tiene-sede-en».

FIGURA 10
La dependencia funcional



Ahora, si en lugar de utilizar una fórmula matemática utilizamos una lista de dos columnas (fig. 11.a) una relativa a la empresa y la otra se refiere a la ciudad, habremos caracterizado igualmente una relación entre la clase Empresa y la clase Ciudad. Luego la relación se podría extender a la información sobre la dirección, el código fiscal, etc. (fig. 11,b).

FIGURA 11
Tablas y relación

Nombre Empresa		Ciudad	Nombre Empresa		Ciudad	Dirección	Cod.Fiscal
FIAT	Torino		FIAT	TO	Via...	

(a) (b)

Decimos que el conjunto de «tuple» (llamamos así al renglón del cuadro en la terminología del modelo relacional) que caracteriza una relación, constituyen también un archivo, en el cual el récord es la tuple en singular. En este punto es necesario hacer algunas precisiones sobre la terminología usada en el modelo relacional, o mejor, ver qué correspondencia hay entre los conceptos de tipo general vistos hasta ahora (y válidos también para el modelo reticular) y los análogos del mundo relacional. Con este objeto hemos confeccionado el cuadro de la fig. 12.

Lo más patente es que mientras en el modelo reticular una tabla, por analogía está asociada al concepto récord, en el modelo relacional el cuadro define las verdaderas y propias relaciones entre clases.

Finalmente, si se vuelve a tomar en consideración la fig. 8, es fácil comprender cómo en el modelo relacional es posible utilizar la técnica del conector lógico para asociar las tablas entre ellas.

FIGURA 12
Nota sobre la terminología usada

Modelo relacional	Terminología general
relación	archivo/entidad
tupla	récord
posición en la tupla	campo
(nombre de la) claset	atributo
dominio	conjunto de los valores de un atributo...

DE LA REPRESENTACION A LA INTERPRETACION

La adopción de un modelo particular de BDMS caracteriza el denominado esquema lógico de la base de datos. Anteriormente, hemos visto cómo partiendo de un mismo esquema conceptual (formalización de la base de datos), es posible obtener distintos esquemas lógicos en donde cada uno tiene en cuenta el modelo particular adoptado.

Un aspecto muy importante concierne al modo con el cual «interpretamos» los distintos modelos de la base de datos del sistema para la gestión de los datos, es decir, por ejemplo, cómo un sistema comercializado O-DS/II, ORACLE, etc.) implementa la filosofía reticular o la relacional. La adaptación del esquema de la base de datos a éste o a otro sistema existente en el mercado (o realizado expresamente) nos lleva a la definición del esquema físico. El esquema físico es pues en definitiva la representación informática de la base de datos.

Una vez la estructura representada y memorizada por el elaborador, utilizando el llamado (DDL) Definición del Lenguaje de los Datos, se pasa a la introducción de los datos y por consiguiente a su elaboración. Esta última consiste principalmente en la interrogación de la base de datos haciendo uso del «query» lenguaje, específico de cada BDMS. El resultado que de aquí se deriva son los elementos esenciales para poder efectuar cualquier tipo de interpretación acerca de la reali-

dad que se va investigando. Por ejemplo: «¿a cuánto asciende la facturación de una cierta empresa?», «¿qué empresa está al día con los pagos?», «¿qué empresa ha superado una fecha de facturación?», etc.

En la práctica, en la fase de interpretación, debemos poder evaluar, con un margen de sencillez, todos los aspectos que han orientado los modelos de la porción de realidad que nos interesa.

Puede ser que la fase de interpretación no satisfaga plenamente al usuario; esto llega por ejemplo, cuando se encuentran dificultades en la extracción o en la valoración de los resultados. En la mayoría de los casos, tales problemas son imputables a una estructuración de la base de datos poco adecuada a las exigencias reales del usuario. En estas circunstancias, las dificultades que surgen deben servir como sugerencias para las sucesivas modificaciones que se deben aportar a la base de datos.

CONCLUSIONES

Ultimamente, gracias a la evolución de la tecnología y la consiguiente reducción de los costes, hemos asistido a la creciente difusión de los elaboradores y de su uso para facilitar y potenciar la posibilidad de tratamiento de la información. Paralelamente las nuevas tecnologías informáticas proponen nuevos modelos de sistema para la gestión de los datos.

De todos modos no se debe olvidar que este instrumento nos reporta un beneficio en relación de cómo lo utilizemos. Un sistema informático en general sustituye, mejorándolo, a un sistema informático preexistente; si nos falta esto o está mal organizado, el esfuerzo que empleamos para construir el sistema informático es desproporcionado con respecto a los beneficios que podemos obtener. Es pura utopía creer que se puede sustituir el elaborador cuando falta orden y organización.

Resulta esencial, en primer lugar, saber conceptualizar la organización de los datos y de la información utilizada para modelar el trozo de realidad que queremos tratar. Por este motivo han sido creadas y dotadas distintas metodologías, algunas de las cuales, al menos en principio, calca la esquematización utilizada en este trabajo.

Referencias

- ALBANO A., ORSINI R. (1985) *Basi di Dati*, Serie de informática, Boringhieri, Torino.
- SPRUNGOLI R. (1987) *Le Basi di Dati*, Editori Riuniti, Roma.
- CERI S., MISSIKOFF M., y SCHMIDT J. W. (1988) *Advances in Data base Technology*. EDBT'88, International Conference on Extending Data base Technology. Venecia '88, Ediciones G. Goos y J. Hartmanis.

Estructura y organización de una
base de datos

Guglielmo Trentin

CL&E, 1992, 13, pp. 81-88

Datos del autor: Trabaja en el Instituto para la Tecnología.

Dirección: Istituto per le Tecnologie Drolettiche CNR. Genova- via all'Opera Pora, n.º 11, 16 145 Genova.

Artículo original: Struttura e organizzazione di un data base. *Golem* n.º 2, 1990, 8-14. Reproducido con autorización. Traducción de Lola Abelló.

© De todos los artículos deberá solicitarse por escrito autorización de CL&E y de los autores para el uso en forma de facsímil, fotocopia o cualquier otro medio de reproducción impresa. CL&E se reserva el derecho de interponer las acciones legales necesarias en aquellos casos en que se contravenga la ley de derechos de autor.