

Com és una Base de Dades

per Rafael Camps i Paré

Avui dia el terme "Base de Dades" (BD) forma part del vocabulari usual de nombrosos camps de la ciència i de la tècnica. El terme sol designar un gran conjunt de dades emmagatzemades en les memòries auxiliars d'un ordinador i els mitjans necessaris per actualitzar-les i accedir-hi. En aquest article intentarem explicar molt informalment què són i com són les BD. Suposarem que el lector té una certa familiaritat, per petita que sigui, amb els ordinadors.

Rafael Camps i Paré (Tortosa, 1940) és llicenciat en informàtica i professor de la facultat d'informàtica de la Universitat de Barcelona; treballa com a tècnic en sistemes informàtics des de l'any 1962. Actualment ho fa a SPERRY, S.A.

Els fòssils i l'ordinador

No fa gaire temps, va aparèixer en aquesta mateixa revista un article ("Com fer el fitxer per a la nostra col·lecció de fòssils", de Joan Roca, març 1983) explicant com organitzar manualment unes fitxes (uns fitxers) que facilitin, al paleontòleg o a l'aficionat, la recerca d'informació sobre la seva col·lecció de fòssils.

Si les dades, la informació d'aquestes fitxes, estiguessin emmagatzemades en un disc d'ordinador, constituint la BD de fòssils, haurien estat enregistrades usant una codificació o representació binària (bits). Les dades estarien distribuïdes (organitzades físicament) en el disc de tal manera que facilitarien la seva ràpida localització.

Suposem que l'ordinador disposa de programes per accedir a les dades (*consultar-les*), així com per incloure'n de noves, modificar-les i suprimir-les (*actualitzar-les*). Aquests programes se solen conèixer amb el nom de *Sistemes de Gestió de BD*, SGBD.

L'usuari de la BD de fòssils desitja assegurar-se davant d'un teclat i una pantalla, per mantenir un diàleg de consulta o d'actualització de les dades de la BD, de la forma més lògica possible. Això vol dir que no li caldrà conèixer ni la representació ni l'organització físiques de les dades, ja que són coses que realment no li interessien.

En el cas de les fitxes manuals proposades per Joan Roca, l'usuari té una

visió directa de la representació física, veu (i toca) exactament allò que hi ha enregistrat (escrit a mà o a màquina) tal com està enregistrat físicament. I les dades les veu organitzades de la mateixa manera que ho estan físicament: unes quantes fitxes amb unes dades disposades en un cert ordre, repetides en diverses fitxes, etc. Treballa a nivell físic.

En cas de disposar de l'ordinador, l'usuari no veu directament la representació ni l'organització físiques, ja que entre ell i allò enregistrat sobre els discs (bits) s'interposa un "intermediari intel·ligent" (bàsicament programes d'ordinador, el SGBD), que permet que l'usuari vegi les dades de la forma més convenient per a ell. Treballa a un nivell superior al físic.

Fins i tot, hi haurà usuaris que les podran veure d'una forma i altres d'una altra. Seran diferents visions d'una mateixa BD. Com si s'usessin diferents ulleres per veure una mateixa cosa de formes diferents. Per exemple, a alguns usuaris de poca confiança se'ls podria "amagar" la informació del *calaix* on és el fòssil que els interessa i la informació del *jaciment* on es va trobar. Encara que la BD que fem servir com a exemple és molt simple, i són poques, per tant, les visions que podem donar-ne als usuaris, habitualment les BD reals tenen centenars o milers de dades amb *interrelacions* entre elles. Així, una BD per a la gestió del manteniment d'una gran planta petroquímica tindria dades sobre les característiques i l'estat dels diferents components de la planta, dades sobre el manteniment preventiu repetitiu, sobre les peces

disponibles al magatzem, sobre el personal de manteniment, sobre les ordres de treball en curs, sobre la documentació disponible, sobre els subministradors, etc. El nombre d'usuaris, de tipus d'usuari i de visions sol ser elevat.

Per facilitar l'exposició, però, seguirem amb l'exemple dels fòssils, i enumerarem tot seguit les dades que suposem tenir per a cada exemplar de la col·lecció (no hi incloem totes les descrites per Joan Roca):

- número de registre de l'exemplar
- espècie
- estrat
- coordenades del jaciment
- nom del jaciment
- localitat del jaciment
- calaix
- com es va obtenir

¿Què és una BD?

Ara ja estem en disposició de fer una definició més precisa de què s'entén actualment per BD.

Una BD és un conjunt de dades, que poden estar interrelacionades, emmagatzemades en un ordinador. Associat al conjunt de dades existeix un conjunt de programes (SGBD) que facilita als usuaris mitjans simples per incloure, consultar, modificar i suprimir dades, mitjans que són independents de la forma física de representació, tant de les dades com de les seves interrelacions.

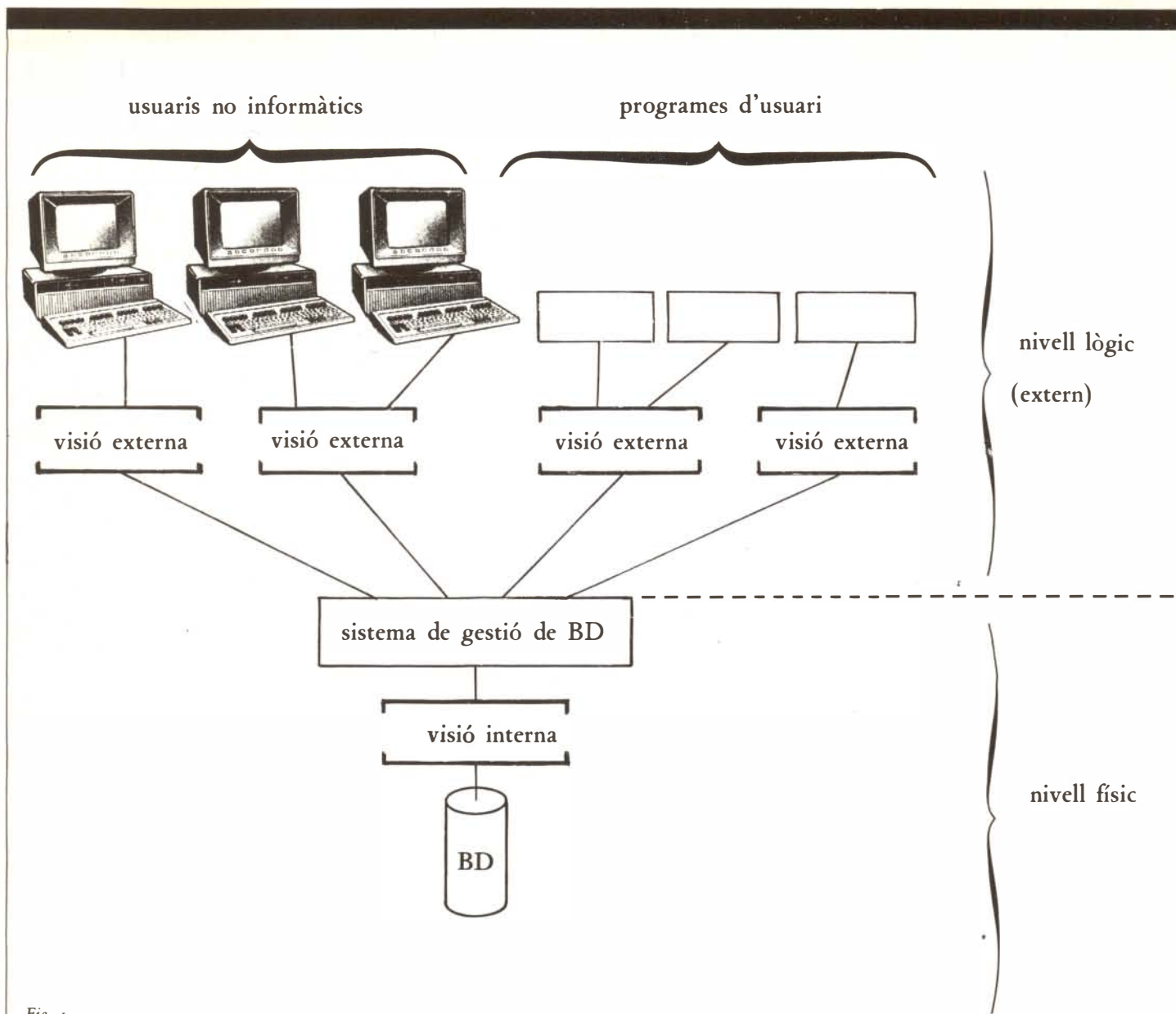


Fig. 1
Estructura a dos nivells d'una Base de Dades

Tres nivells

Hem vist fa un moment l'existència de dos nivells de representació o de visió de la BD. Es podrien anomenar *nivell físic* (discs, bits,...) i *nivell lògic* (usuaris), o també, *intern* i *extern*, respectivament.

Hem de pensar que, a més dels usuaris externs, no informàtics, hi ha també usuaris informàtics que escriuen programes per fer llistes de jaciments o d'espècies per àrea geogràfica... o, potser també, programes que dialoguen amb els usuaris no informàtics. Aquests usuaris informàtics escriuen programes usant llenguatges de programació d'alt nivell (com el FORTRAN, COBOL, BASIC, etc.), és a dir, llenguatges bastant independents de les característiques físiques. Per tant, convé que ells també vegin

la BD de forma lògica, independentment de les seves característiques físiques, tal com la veuen els no informàtics. Però interessarà que la seva visió segueixi l'estil del llenguatge de programació que estan usant, i que a certs programes se'ls puguin amagar determinades informacions, etc.

En resum, per a una mateixa BD poden haver-hi moltes *visions* externes diferents i cada una d'elles pot ser emprada per diferents usuaris, siguin o no informàtics. Però, és clar, de visió o de representació interna (física), només n'hi haurà una. Aquesta representació serà la usada pel SGBD. És el SGBD que es comunica, dona i rep dades, amb els programes usuaris. Vegeu la figura 1.

Les visions o representacions es descriuen al SGBD amb un llenguatge adequat (ja en parlarem més endavant).

Molts dels sistemes de BD que actualment funcionen, tenen una estruc-

tura amb dos nivells de *descripció* de la BD (intern i extern), com la que acabem de veure; però hi ha una tendència cap als sistemes amb tres nivells. Quin és aquest tercer nivell?

Sembla raonable disposar d'una descripció única de tipus lògic, de la mateixa manera que en tenim una de física. Les descripcions externes es faran tot prenent com a referència una descripció única, la dita descripció de *nivell conceptual*. Aquest és el tercer nivell. Vegeu la figura 2. Ara tindrem dos nivells de tipus lògic, l'extern i el conceptual.

La descripció conceptual serà, doncs, una descripció lògica global, independent de la descripció física o interna i de les diferents visions externes, independent també dels usos que es facin de la BD. Això facilita, entre altres coses, la independència entre el nivell extern i l'intern (més endavant direm alguna cosa més sobre aquesta independència).

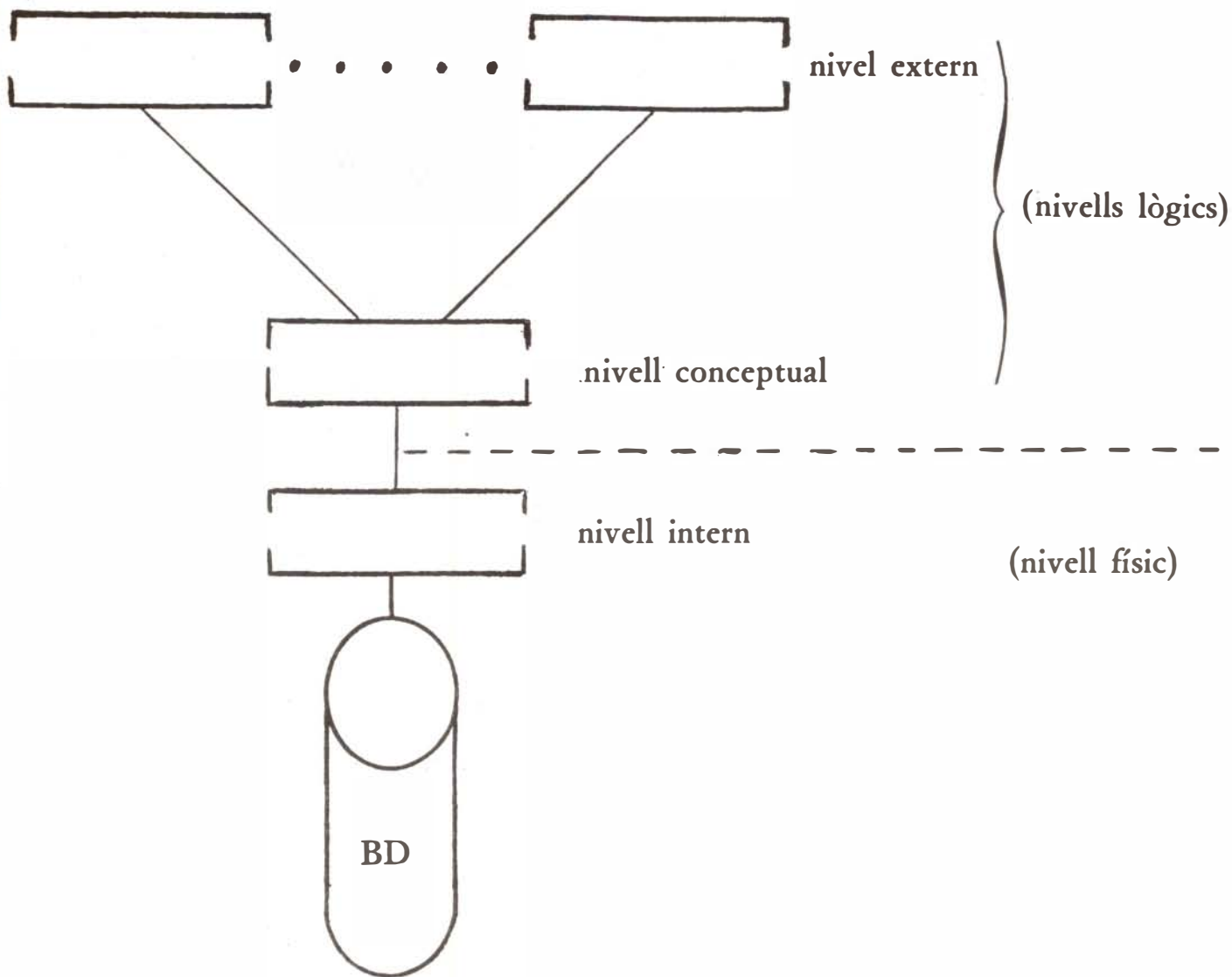


Fig. 2
Estructura a tres nivells d'una Base de Dades

Model d'una BD: L'enfocament relacional

Una representació de les dades, una visió d'una BD, és un model (així com una BD és un model de la realitat que representa). Hi ha moltíssimes formes de representar una BD, de fer-ne un model. Representem ara, sobre aquest paper, la BD de fòssils, fent-ne un model, sense pensar en qüestions informàtiques.

Podríem disposar la informació en una *taula* de vuit *columnes*, una per dada, i amb una *línia*, per cada exemplar de fòssil existent en la nostra col·lecció (en la realitat, no són raïses les BD amb milions de línies). Cada línia contindrà tota la informació d'un exemplar. Vegeu la figura 3.

Cada línia es pot identificar pel número d'exemplar, que és l'única co-

lumna que en aquesta taula tindrà obligatòriament valors diferents en totes les línies. D'aquesta columna, en direm que és *identificadora*.

La BD manual que Joan Roca proposava, consta de quatre fitxers. L'un és una llibreta de registre d'entrada (amb molt poques dades), un altre consta d'una fitxa per exemplar; un altre conté una fitxa per espècie, on consten tots els exemplars que en tenim, i, un últim fitxer, amb una fitxa per jaciment. La major part de dades estan repetides a diversos fitxers, és a dir, hi ha una gran redundància. L'objectiu d'aquesta divisió en quatre fitxers i del disseny de cada un d'ells, és facilitar-ne l'ús. Per exemple, fer ràpides (és una qüestió física) i simples (és una qüestió lògica) les consultes més habituals: Però, en la representació que hem fet a la figura 3 no ens hem preocupat gens de l'eficàcia de l'ús que se'n pugui fer. (L'ús és una cosa que pot canviar amb el temps.)

Així, la nostra representació és independent de l'ús.

La representació de la figura 3 (taula o fitxer únic) presenta també molta redundància. Per exemple, es repeteix més d'una vegada que el jaciment de coordenades 41° 30'N/01° 30' E es denomina "Els Moions" i que es troba a la Poble de Claramunt, o bé que al calaix A5 hi ha fòssils de l'espècie *Heteraster Oblongus*.

Fixem-nos ara en les interrelacions de les dades. Es veu immediatament en la figura 3 que, un cop donat un número d'exemplar, totes les altres columnes queden determinades, és a dir, que l'espècie, l'estrat, el jaciment, etc., venen determinats pel número d'exemplar; podríem dir-ne *atributs* de l'exemplar.

Aquest fet és una interrelació o *dependència funcional* entre dades, i l'hem expressat posant totes les dades en una sola taula i subratllant la columna "número d'exemplar". Però, analit-

N.º de l'exemplar	espècie	estrat	coordenades jaciment	nom del jaciment	localitat del jaciment	calaix	com es va obtenir
121	Heteraster Oblongus	Aptià	40°54'N/ 00°30'E	Les Sitges	Xerta	A5	in situ
122	Heteraster Oblongus	Aptià	40°50'N/ 00°31'E	Sant Roc	Tortosa	A5	in situ
123	Retepora Cellulosa	Bartonià	41°30'N/ 01°30'E	Els Moions	Pobla de Clar- munt	B6	in situ
129	Heteraster Oblongus	Barrenià	41°13'N/ 01°40'E		Olèrdola	A6	in situ
130	Operculina Alpina	Bartonià	41°30'N/ 01°30'E	Els Moions	Pobla de Clar- munt	A9	in situ
131	Heteraster Oblongus	Aptià	40°34'N/ 00°60'W		Morella	A6	J. Creixell
133	Heteraster Oblongus	Aptià	40°38'N/ 00°14'E	Sant Roc	La Sénia	A6	in situ
134	Operculina Alpina	Bartonià	41°30'N/ 01°30'E	Els Moions	Pobla de Clar- munt	A9	no se sap

Fig. 3
Tota la BD de fòssils en una sola taula

zant el significat la semàntica, de les dades, trobem més dependències. El nom i la localitat del jaciment venen determinats per les seves coordenades. El nom de jaciment no determina la localitat ni tampoc al revés, ja que poden existir jaciments diferents en una mateixa localitat. Suposarem, a més, que els fòssils d'una espècie estan dintre d'un o de més calaixos, però que mai no es barrejaran en un calaix fòssils de diferents espècies. Llavors, el calaix determina l'espècie (no al revés). Suposem que ja no hi hagi més dependències.

Ara que ja coneixem les dependències entre columnes, les podem fer explícites dividint l'única taula de la figura 3 en les tres taules de la figura 4. Una per l'exemplar, una pel calaix i una altra pel jaciment. Les tres taules estan *interrelacionades* gràcies als noms de les columnes que es repeteixen a altres taules. Concretament, la columna coordenades i la columna calaix. Aquesta representació amb tres taules és una altra modelització de la mateixa BD que a la figura 3 hem representat amb una sola taula. La divisió en tres taules té com a finalitat fer explícites les dependències.

Les dependències no es poden determinar pel contingut de la BD en un moment donat, sinó que s'han de determinar pensant en tots els possibles continguts. Així, en l'exemple de la figura 3, s'hi veuen totes les dependències citades abans, però si eliminéssim alguns exemplars, no es veurien. L'anàlisi de dependències exigeix

el coneixement del problema, és una qüestió de *semàntica* de dades.

Cada taula té, com a mínim, una columna identificadora. De les columnes identificadores se'n solen dir *claus*. En altres BD podria passar que la identificació de les línies necessités més d'una columna. Per exemple, si la BD de fòssils inclogués diverses col·leccions, la clau de l'exemplar es compondria de dues columnes, la col·lecció i el número d'exemplar.

Dèiem abans que la taula de la figura 3 presenta redundàncies. En descompondre-la en les tres taules de la figura 4, hem hagut de repetir les columnes calaix i coordenades per permetre la interrelació de les taules. Això provoca una redundància. Però veiem també que ens n'hem estalviat molta en eliminar bona part de les repeticions de l'espècie, de la localitat i del jaciment.

Fixem-nos que en cap moment no ens hem preocupat de l'ús que es farà de la BD, ni de si és una BD informàtica o bé és de fitxes de cartó. La veritat és que si organitzéssim les nostres fitxes manuals com indica la figura 4, la majoria de consultes habituals serien molt feixugues. Així, per conèixer tots els jaciments d'on tenim exemplars d'una espècie determinada, hauríem de buscar, primer, a les fitxes de calaix, tots aquells calaixos que les contenen; després, per a cada un d'aquests calaixos, buscar a les fitxes d'exemplar per trobar les coordenades, i, amb aquestes coordenades, buscar a les fitxes de jaciments. Realment

costós! Però si el model de les tres taules fos una representació de nivell lògic (extern o conceptual) en un sistema informàtic, l'usuari podria fer una consulta com l'anterior, directament, i el sistema se les compondria per tal de respondre correctament. Per una altra part, la resposta no seria necessàriament lenta, ja que les qüestions de rendiment depenen bàsicament de la representació física, de la qual no hem parlat.

Bé. Hem vist una manera molt simple de modelitzar una BD a base de taules, columnes i línies. Aquesta és la forma més usual de com els sistemes informàtics mostren les BD als usuaris (visió externa), almenys als usuaris no informàtics. Aquesta manera de modelitzar les BD, aquest formalisme, és conegut com enfocament *relacional*. L'hem presentat aquí molt informalment, però està molt formalment definit. Hi ha desenvolupada tota una teoria basada en el concepte matemàtic de relació (concepte molt anterior a les BD). Aquí hem usat una terminologia diferent; la nostra taula és la *relació* de la teoria, la columna és aproximadament l'*atribut*, la línia és la *tupla*, etc.

Les principals virtuts de l'enfocament relacional són: primera, la seva simplicitat. Tothom entén les idees de taula, columna, línia i clau. Segona, la seva independència de característiques informàtiques físiques. Tercera, el fet de poder aplicar els resultats de la teoria relacional.

La teoria relacional facilita una al-

N.º de l'exemplar	estrat	coordenades del jaciment	calaix	com es va obtenir	calaix	espècie
121	Aptià	40°54'N/00°30'E	A5	in situ	A5	Heteraster Oblongus
122	Aptià	40°50'N/00°31'E	A5	in situ	A6	Heteraster Oblongus
123	Bartonià	41°30'N/01°30'E	B6	in situ	A9	Operculina Alpina
129	Barrenià	41°13'N/01°40'E	A6	in situ	B6	Retepora Cellulosa
130	Bartonià	41°30'N/00°30'E	A9	in situ		
131	Aptià	40°34'N/00°06'W	A6	J. Creixell		
133	Aptià	40°38'N/00°14'E	A6	in situ		
134	Bartonià	41°30'N/01°30'E	A9	no se sap		

Taula 1

Taula 2

gebra relacional i un *càlcul relacional* que permeten fer operacions sobre les relacions (les taules), per ajuntar o seleccionar línies, columnes, etc. Evidentment, els sistemes informàtics basats en l'enfocament relacional, no exigeixen a l'usuari de conèixer l'àlgebra o el càlcul relacionals, sinó que posen a la seva disposició una versió "traduïda", simple, menys formal, més natural.

Normalització

En el moment d'escollir una representació d'una BD es poden seguir molts criteris. Un podria ser l'estalvi d'espai i de temps, per exemple, minimitzant el nombre total de signes a emmagatzemar, o bé fent ràpides consultes habituals. Aquestes qüestions se solen deixar per a la descripció de nivell físic (intern). En l'exemple dels fòssils hem volgut veure la BD sense preocupar-nos del rendiment. Així, per a la representació de la figura 4, hem adoptat el criteri de reflectir la màxima claredat semàntica, posant de manifest les dependències. Es tracta d'un criteri lògic, no físic. L'objectiu es podria enunciar també així "representar cada fet en un sol lloc"; per exemple, el fet que al calaix A6 hi hagi només l'espècie *Heteraster Oblongus*, es representa en un sol lloc (segona línia de la taula de calaixos), i, en canvi, en la figura 3 apareixia a

tants llocs, tres, com exemplars hi havia al calaix.

Al resultat que s'obté usant aquest criteri se'l coneix amb el nom de *tercera forma normal*, i el procés de descomposició de taules que ens hi ha portat rep el nom de *normalització*. Allò que ha fet aquest procés de normalització és aconseguir que cada taula tingui una columna clau, que les altres columnes siguin independents entre elles i depenguin només de la clau. (Això no passa a la figura 3.)

Convé subratllar que una de les conseqüències, més aviat de tipus físic, que sol tenir la representació en tercera forma normal, és la disminució de la redundància, com ja hem vist abans.

Existeixen moltes altres formes normals. La tercera és la més coneguda. En la pràctica, podem considerar que les formes normals són regles que ens poden ajudar a decidir la representació a adoptar.

Llenguatges de descripció de la BD. Esquemes

Perquè un SGBD, amb una arquitectura a tres nivells (extern, conceptual, intern), pugui fer la seva feina, caldrà donar-li, per a cada un d'aquests nivells, una descripció adequada de la BD. Per al nivell extern, se li'n donaran tantes com visions diferents es vulguin donar als usuaris.

La descripció de nivell intern sol rebre el nom d'*esquema intern* o *esquema físic*. La descripció de nivell conceptual sol rebre el nom d'*esquema conceptual*, o *esquema a seques*. Les descripcions de nivell extern solen rebre el nom de *subesquemes* o *esquemes externs*, o també *visions*.

De la mateixa manera que per usar la BD (per dialogar amb el SGBD o escriure programes) s'hauran d'usar llenguatges especials, per escriure les descripcions de la BD o esquemes s'usaran llenguatges especialitzats en la descripció de dades.

Per als subesquemes, o visions externes, que els usuaris tindran de la BD, és freqüent disposar de llenguatges d'estil relacional, o sigui, llenguatges que permetin descriure la BD com un conjunt de taules i, per a cada taula, descriure les columnes (nom de la columna, és clau o no, etc.), l'ordre en què apareixeran, etc. Habitualment, els SGBD tenen un llenguatge de descripció externa per a cada llenguatge d'ús, adaptat al seu estil i necessitats.

Per a l'esquema conceptual, únic, se sol estar d'acord que no es disposa, de moment, d'un llenguatge prou adequat, però els SGBD basats en l'enfocament relacional solen adoptar-ne un del tipus de l'usat per a les visions externes, o sigui, un llenguatge que descriu la BD en termes de taules, columnes i claus.

La BD de fòssils podria descriure's al nivell conceptual com a composta de les tres taules de la figura 4, i l'esquema conceptual podria tenir un aspecte semblant al de la figura 5. Es podria disposar de tants subesquemes o visions externes d'aquesta BD com

	coordenades del jaciment	nom del jaciment	localitat del jaciment
	40°34'N/00°06'W 40°38'N/00°14'E 40°50'N/00°31'E 40°50'N/00°30'E 41°13'N/01°40'E 41°30'N/01°30'E	— Sant Roc Sant Roc Les Sitges — Els Moions	Morella La Sénia Tortosa Xerta Olèrdola Pobla de Claramunt

Fig. 4
La BD de fòssils amb tres taules

Taula 3

es vulgués. Per exemple, a un usuari que s'interessés només per l'espècie i l'estrat dels diferents exemplars, se li podria fer veure la BD com si estigués formada únicament per aquesta taula:

TAULA FÒSSILS (NÚM-EXEMPLAR, ESPÈCIE, ESTRAT)

L'enfocament relacional no entra en qüestions físiques, de manera que els SGBD basats en aquest enfocament tenen, cada un, el seu propi tipus d'esquema intern (els que el tenen...). Què es descriu en l'esquema intern? Bàsicament, la representació física de les dades i els mètodes que s'empraran per fer eficient l'accés. Es descriu la correspondència entre les taules de l'esquema conceptual i les taules físiques (aquí seguim usant els termes taula i columna, encara que no són els més usats a nivell intern). Així, el nombre de taules i la distribució de columnes no tindria per què ser igual en els dos esquemes. També es defineix el sistema de codificació interna utilitzat per a cada columna (això determina el nombre de bits necessaris). Per a les columnes a través de les quals se suposa que es faran consultes amb més freqüència (per exemple les claus i altres columnes, com l'espècie a la nostra BD de fòssils), es defineixen mecanismes que facilitin un accés ràpid (per exemple, índexs).

La persona (o grup) que dissenya i escriu els esquemes se sol conèixer amb el nom d'*administrador de la BD*.

La principal raó per la qual les BD se solen descriure a tres nivells, és la d'obtenir *independència*. Independència del nivell intern respecte a l'extern, i al revés. D'aquesta manera es poden (si el SGBD és prou intel·li-

gent) afegir, suprimir o modificar les visions externes, el món dels usuaris, sense necessitat de canviar la BD física, real. D'altra banda, i potser més important encara, es poden modificar les característiques de la BD física (normalment per millorar el temps d'accés o l'espai ocupat), sense que això afecti la visió dels usuaris ni els programes. Si el lector té alguna experiència en fitxers informàtics clàssics, sabrà apreciar aquesta independència. La veritat és que la independència en els SGBD reals es més o menys gran, però mai no és absoluta. I així, per exemple, hi ha SGBD que no permeten, a les visions externes (subesquemes), una distribució en taules que no sigui un subconjunt de la distribució en taules física (la de l'esquema intern).

L'enfocament relacional no és l'únic que existeix. N'hi ha molts altres, però sembla que el relacional s'està extenent ràpidament. Un altre enfocament molt estès (i bastant diferent del relacional) és el conegut com *CODASYL*. (*CODASYL* és el nom d'un organisme, nord-americà, especialitzat en la definició de llenguatges per a sistemes informàtics de gestió i del qual formen part la majoria de constructors d'ordinadors i certs usuaris "importants".) Més que un enfocament, es tracta d'una definició concreta de llenguatges de descripció de la BD per als tres nivells (a més de llenguatges d'ús). Aquest enfocament no es queda només al nivell lògic, sinó que (en la versió 1978) arriba a concretar un llenguatge per al nivell intern. Però, fins i tot en els nivells lògics és més "informàtic", menys "natural" que el relacional, i així, per

exemple, no té definit cap llenguatge d'ús per a usuaris no informàtics.

Els diferents enfocaments o models no són necessàriament excloents. Poden coexistir-ne més d'un, en un mateix SGBD. Per exemple, hi ha SGBD que permeten visions externes (i llenguatges d'ús) relacionals de BD *CODASYL* (definides com a *CODASYL* en els nivells conceptual i intern).

Llenguatges d'ús

Per usar les BD es fan servir uns llenguatges especialitzats en la consulta i actualització de dades. N'hi ha per a usuaris no informàtics i per a informàtics.

Els llenguatges per a no informàtics són llenguatges per dialogar amb el SGBD des d'un terminal. N'hi ha d'estil "llenguatge natural" (anglès, és clar!), però la major part són molt poc "naturals", encara que molt senzills.

A més d'aquests llenguatges, els usuaris informàtics en disposen d'altres més potents, encara que més complexos (a vegades són simplement ampliacions dels primers) i sobretot disposen de llenguatges per escriure programes. Aquests, més que llenguatges complets, solen ser conjunts de verbs (instruccions) que s'afegeixen als llenguatges de programació habituals (*COBOL*, *FORTRAN*, etc.).

L'àlgebra i el càlcul relacionals són la base dels llenguatges (de diàleg o de programació) dits relacionals.

TAULA EXEMPLARS (NÚM-EXEMPLAR, ESTRAT, COOR, CALAIX, OBTENCIÓ)

NÚM-EXEMPLAR:	5)	CARÀCTERS	ÉS CLAU
ESTRAT:	12)	CARÀCTERS	
COOR:	10)	CARÀCTERS	
CALAIX:	2)	CARÀCTERS	
OBTENCIÓ:	15)	CARÀCTERS	

TAULA CALAIXOS (CALAIX, ESPÈCIE)

CALAIX:	2	CARÀCTERS	ÉS CLAU
ESPÈCIE:	25	CARÀCTERS	

TAULA JACIMENTS (GOOR., NOM-JAC., LOCALITAT)

COOR:	10	CARÀCTERS	ÉS CLAU
NOM-JA:	25	CARÀCTERS	
LOCALITAT:	25	CARÀCTERS	

Fig. 5
Un possible esquema conceptual de la BD de fòssils

CODASYL té només definit un llenguatge per escriure programes. Més ben dit, verbs per afegir al COBOL i també al FORTRAN.

Aquest control se sol montar sobre el *codi d'usuari*. Cada usuari s'identifica per un nom o codi (acompanyat de contrasenyes secretes, per no ser suplantat). Cada codi, cada usuari, tindrà associats uns atributs. L'un pot ser, posem per cas, la llista de subesquemes amb què pot treballar. Així, per exemple, certs usuaris poden treballar només amb un subesquema X, o sigui, només poden tenir la visió X, i aquesta visió segurament no inclou totes les dades. Un altre atribut podria ser: "només pot fer operacions de consulta, però no d'actualització". Alguns sistemes associen a cada usuari una categoria. Quanta menys categoria, més prohibicions tindrà.

i s'ha escrit ja sobre la BD la línia de la taula EXEMPLARS, però falta escriure la de la taula JACIMENTS. Suposem aquí que la BD física consta de les tres taules de la figura 5). En aquest cas, allò que solen fer els SGBD automàticament quan torna l'alimentació, és desfer les operacions ja fetes pels processos interromputs i avisar als usuaris.

Un exemple d'un altre tipus de situació podria ser: es detecta l'existència d'un programa nou no prou ben pensat, que des de fa dos dies modifica erròniament la BD. Per arreglar situacions d'aquest estil, es fan de tant en tant còpies (totals o parcials) de la BD, sovint sobre cinta magnètica. Quan es detecta l'anomalia, es reconstrueix la BD amb l'última còpia de contingut correcte. Llavors se li demana al SGBD que repeteixi automàticament totes les actualitzacions fetes després d'aquella còpia, però saltant-se les fetes pel programa erroni.

Tant per desfer el fet com per repetir-lo, si es vol que sigui automàtic, caldrà disposar d'uns fitxers, usualment anomenats *diaris*, en els quals, durant els processos d'actualització de la BD, s'enregistrin també les actualitzacions. Els processos d'obtenció

Qüestions auxiliars, però importants

Fins aquí hem parlat de l'estructura de les BD, dels llenguatges de descripció i dels d'ús. Però no donaríem una visió prou global de "com són" les BD si no parléssim de certs problemes auxiliars que planteja l'ús d'una BD, com són la seguretat i la integritat.

Sovint, el contingut d'una BD és *confidencial* (totalment o parcialment). No es pot permetre que sigui vist per qualsevol. I potser encara, amb més raó, no es pot permetre que qualsevol pugui actualitzar les dades. Per això, els SGBD faciliten un *control d'accés* a les dades.

D'altra banda, a l'esquema (segurament al conceptual) se sol poder especificar proteccions a les dades, com per exemple: per suprimir línies de la taula A s'haurà de donar la contrasenya B.

Durant el funcionament, l'ús, d'un SGBD poden presentar-se molts incidents capaços d'afectar la *integritat* de la BD. Així, per exemple, se'n pot anar l'alimentació elèctrica al mig d'un procés d'actualització i quedar la BD en un estat incoherent (s'estava introduint un fòssil d'un jaciment nou

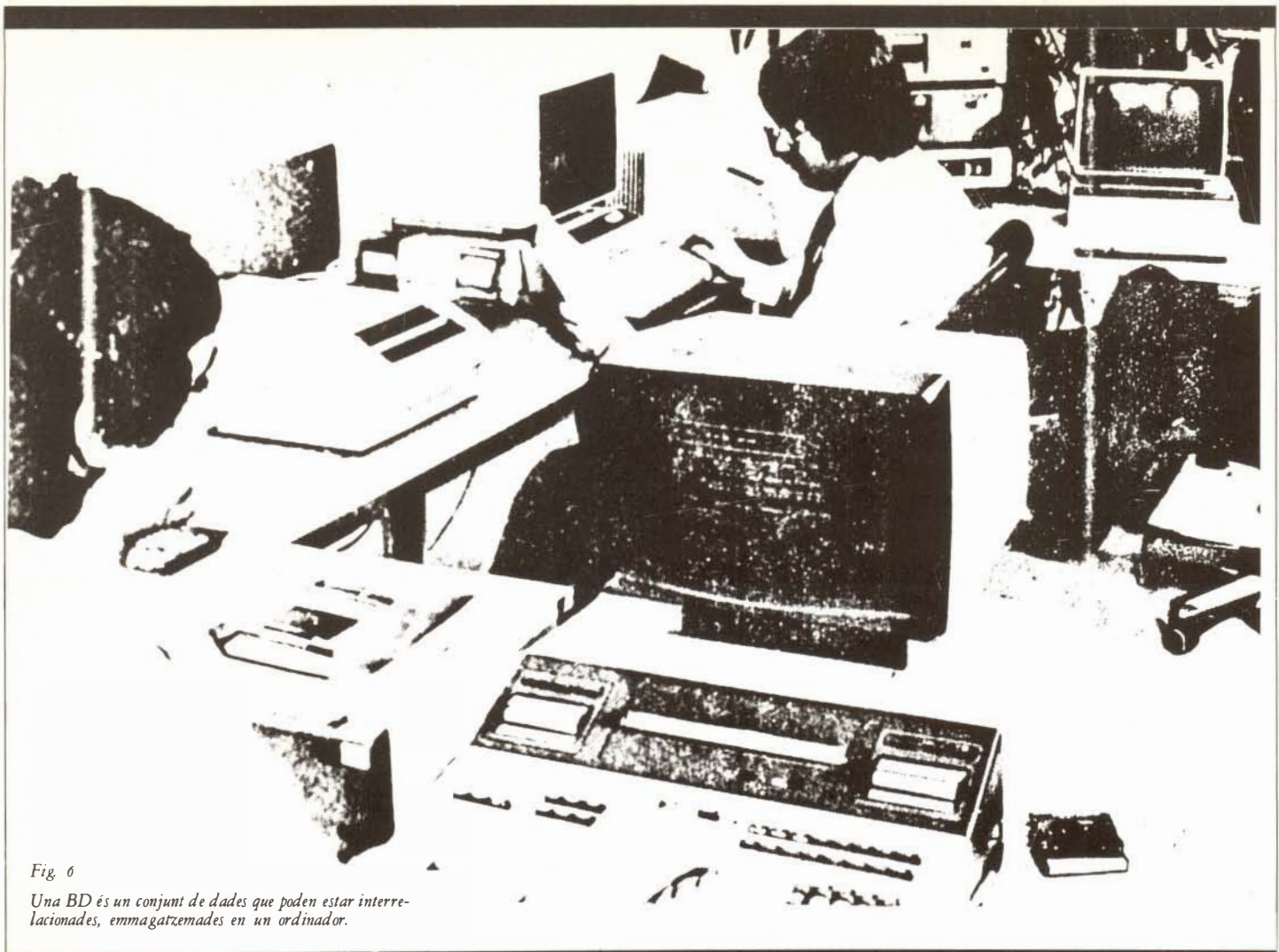


Fig. 6

Una BD és un conjunt de dades que poden estar interrelacionades, emmagatzemades en un ordinador.

d'una BD correcta, utilitzant aquests diaris (fent-desfent), s'anomenen processos de recuperació.

Normalment, els SGBD accepten simultàniament diversos usuaris. Si tots ells fan només consultes, no es planteja cap problema especial d'accés simultani. Però si algun usuari està actualitzant dades, podria passar que, un altre usuari que estigués consultant-les, obtingués resultats incoherents en obtenir-ne algunes ja actualitzades i d'altres encara no. O bé, un programa vol actualitzar unes dades *y* en funció d'unes altres *x*, però, després de llegir *x* i abans d'actualitzar *y*, un altre programa podria modificar *y*. Per evitar problemes d'aquests tipus, els SGBD solen tenir uns mecanismes de cues d'espera (que inevitablement baixaran el rendiment). Amb aquestes cues es pot presentar una situació, anomenada *abraçada mortal*, en la qual un procés *A* fa cua per dades que està usant el procés *B*, i aquest fa cua per dades que està usant l'*A*... Els SGBD detecten aquestes situacions i les resolen tallant alguns dels processos. A més a més, alguns SGBD intenten evitar que s'arribin a produir les abraçades mortals, però no sempre ho aconsegueixen.

Recerca actual

La tecnologia de les BD és un tema en plena efervescència i s'hi dedica actualment un gran esforç de recerca. Citarem breument els principals temes de les recerques teòriques actuals.

Diferents tipus de normalització i de dependències entre dades. Representació de la semàntica de les dades. Llenguatges de descripció al nivell conceptual. Possibilitat d'accedir a la BD amb llenguatge natural. Un dels temes més de moda és el de les BD distribuïdes. Es tracta de BD repartides entre diferents ordinadors interconnectats per una xarxa de comunicacions. Ja hi ha al mercat alguns sistemes que s'autodenominen així, però és un terreny una mica verd encara. Els principals problemes que estan a nivell de recerca en el context de les BD distribuïdes són: Accés simultani i abraçada mortal, recuperació, distribució òptima.

Rafael Camps i Paré

Nota bibliogràfica

Al lector que vulgui tenir una idea general del tema, més àmplia que la que aquí donem, li recomanem el text de James Martin, *Principles of data-base management*, Ed. Prentice-Hall. A qui vulgui arribar a un coneixement més seriós i profund del tema li suggerim el text de C.J. Date, *An Introduction to Database Systems*, Ed. Addison Wesley.