COMPRENDRE ET CONCEVOIR LES BASES DE DONNÉES RELATIONNELLES

SERGE MIRANDA

éditests

COMPRENDRE ET CONCEVOIR LES BASES DE DONNÉES RELATIONNELLES

Connaissez-vous la collection « Langage » et les ouvrages sur l'IBM/PC chez P.S.I. et Éditests ?

Langage

— Programmer en C — Claude Nowakowski
— Clefs pour C — François Piette
— C et ses fichiers — Jacques Boisgontier et Jean-Pierre Lagrange
— Entrées et sorties en C — Jean de Brabandt
— Le langage C — Jean-Louis Fourtanier et Violaine Prince
— Turbo Pascal et ses fichiers — Jacques Boisgontier et Christophe Donay
— C sur Atari ST — Claude Nowakowski
— Turbo Pascal sur PC et compatibles — Frédéric Blanc et Pierre Brandeis
— Cobol sur PC et compatibles — Daniel-Jean David et Daniel Trécourt
— Programmer en Pascal — Daniel-Jean David et Jean-Luc Deschamps

IBM-PC

— Clefs pour PC et compatibles — Daniel Martin, Guy Herzet et Philippe Jadoul
— Clefs pour PC-AT et compatibles — Daniel Martin et François Piette
— Clefs pour MS-DOS, versions 2 à 4 — Yvon Dargery
— MS-DOS pas à pas, versions 2 à 4 — Alain Pinaud
— Le livre du MS/PC-DOS pour PC et AT — François Piette
— Systèmes PC-DOS MS-DOS, versions 2 à 4 — Jean-Pierre Lamoitier, Michel Treillet, Jacques Boyer
— Au cœur de l'IBM-PC — Jeanine Thièle

Pour connaître nos dernières nouveautés,
ou nous soumettre un problème technique,
nous mettons à votre disposition un service Minitel

Service Minitel

Sur le 3615 tapez OI puis * LIV

ISBN 2.86699.066.8

COMPRENDRE ET CONCEVOIR LES BASES DE DONNEES RELATIONNELLES

SERGE MIRANDA

édi tests

«COMPRENDRE et CONCEVOIR
les Bases de données relationnelles»

S. MIRANDA
Professeur Université de Nice
Conseiller scientifique au CERAM
et chez CODD and DATE INTERNATIONAL
Co-fondateur de SOPHIATEC

Du même auteur (avec José Busta comme co-auteur) aux Éditions Eyrolles

L'Art des Bases de données
Tome I: Introduction aux bases de données,
(1re édition en octobre 84, 3e édition en janvier 88)

Tome II: Les bases de données relationnelles
1re édition en mars 1986

Contribution au livre «Bases d'information généralisées», Édition Dunod, 1985, (C. Chrisment et al éditeurs)

Éditeur du livre «Des bases de données aux bases de connaissance» rassemblant les actes des journées AFCET, 28-29 septembre 1987 publié aux Editions P.S.I. en septembre 1987.

Remerciements
Je tiens à remercier tous les étudiants qui en prenant «leur baptême de l'air» base de données avec ce cours m'ont permis de... garder les pieds sur terre.

«A José qui lutte pour retrouver l'entrée du monde réel où il fait bon vivre…»

SOMMAIRE

AVANT-PROPOS

«Pourquoi faire simple quand c'est si facile de... tout compliquer»
J.L. Godard.

En informatique, aujourd'hui, les expressions «Base de Données», (BD) ou «Système de gestion de base de données» (SGBD), ressemblent à une auberge espagnole: tout le monde en parle, tout le monde y va, mais on y trouve n'importe quoi (du système de gestion de fichier avec une interface type BASIC, au SGBD relationnel avec une interface «non procédurale»).

La littérature scientifique actuelle sur les BD s'adresse à des informaticiens, à des spécialistes et peut donc apparaître hermétique aux utilisateurs potentiels des SGBD (très nombreux dans un contexte micro-ordinateur), à ceux qui doivent faire des choix de SGBD relationnels ou comprendre les concepts.

Cet ouvrage, qui s'adresse à un très large public, a un triple but:
– premièrement définir d'une manière rigoureuse et simple les concepts du modèle relationnel (conçu par E.F. Codd, directeur de recherche à IBM San José), c'est-à-dire ses structures de base, ses opérateurs, et ses contraintes d'intégrité (chapitres 1, 2);

– deuxièmement, présenter pour la première fois en France, la méthode de conception d'un schéma relationnel proposée par Codd et Date, qui ne s'appuie pas sur le modèle entité/association (chapitre 3);

– enfin permettre au lecteur, à l'issue de ce livre, d'exercer son sens critique face aux SGBD «relationnels commercialisés» (présentés en annexe 1) et l'inciter à approfondir ses connaissances avec l'un des nombreux ouvrages spécialisés indiqués en bibliographie.

En conclusion, nous donnerons un aperçu des recherches en cours. Une étude de cas complète avec CAMPUS[1], un SGBD totalement relationnel, résultat d'un transfert université (de Nice) – industrie, est présentée en annexe 2 pour montrer la puissance d'un véritable SGBD relationnel comme noyau d'un système d'information et d'aide à la décision.

1. CAMPUS a été en 1987 retenu dans le cadre du nouveau plan informatique pour tous en France. Enfin, Codd lui-même le créateur du modèle relationnel, a décidé de le commercialiser sous le nom de «CODD and DATE CAMPUS», à partir de décembre 1987, sur le marché «formation» et «prototypage».

INTRODUCTION

«No existing data base management system that I know can honestly be claimed to be relational»
CF. Codd (CODD85; p.720)

L'appellation «relationnelle» d'un système de gestion de base de données (SGBD-R), relève très souvent d'une origine commerciale (A.O.C.). C'est un peu comme le vin ou le jus d'orange: quel est le lien entre une orangeade gazeuse et un véritable jus d'orange pressée malgré la même appellation... commerciale? La «relationnalité» d'un produit a souvent été inversement proportionnelle au budget marketing du produit!

Le marché des SGBD-R a représenté 648 millions de dollars en 1984; il devrait être multiplié par 4 d'ici 1990, sans compter 500 millions de dollars pour les SGBD-R répartis sur un réseau d'ordinateurs. En 1990, les SGBD-R devraient constituer alors le seul type de SGBD commercialisé (source IDC 1986). Chassez le relationnel... il revient au galop: en effet, tous les systèmes experts et systèmes d'information et d'aide à la décision (SIAD) qui sont en cours de réalisation, sont construits autour d'un SGBD-R. Quand on sait que le marché des SIAD va être multiplié par 8 de 1985 à 1990, on voit l'attrait supplémentaire d'une A.O.C. pour un SGBD.

Mais qu'est-ce qu'un SGBD-«relationnel»? Il se trouve que le modèle relationnel n'est pas un concept flou que chacun peut redéfinir, mais une référence formelle à des propositions établies par E.F. Codd en 1970 (CODD70). En 1987, aucun SGBD relationnel commercialisé n'implantait complètement le modèle relationnel défini par Codd.

Par exemple, très peu de SGBD relationnels commercialisés sur méga et mini-ordinateur[2] ne comprennent le concept fondamental de «Domaine (sémantique)» qui permet d'une part la mise en œuvre d'interfaces conviviales pour non-informaticiens (interfaces relation universelle et langue naturelle), et d'autre part la définition automatique d'index systèmes.

Les arguments concernant les faiblesses pratiques du modèle relationnel sont systématiquement émises par des concepteurs de SGBD-R du type AOC qui n'ont jamais étudié le modèle relationnel de Codd (certains n'en ont même jamais entendu parlé!): Les principales critiques du modèle relationnel éma-

2. Seul CAMPUS, opérationnel sur IBM-PC-XT intègre ce concept dans sa double dimension:
- définition comme un «type de données sémantiques»,
- manipulation avec les opérateurs relationnels liés aux domaines (domain constrained),

Supra, par exemple ne le prend en compte que dans sa partie définition. Nous reviendrons sur ce concept sémantique crucial en p. 17.

nent de personnes qui ne connaissent pas ou mal le modèle ou qui l'interprètent à leur façon, soit trop «théorique», soit trop «académique»!

Le modèle relationnel créé par E.F. Codd en 1970 constitue un des apports les plus remarquables à l'informatique dans son rapprochement vers le non-informaticien que l'on peut résumer en quatre points (DATE 76) (CODD 82):

- simplicité des concepts de base:

une relation par exemple, est un ensemble au sens mathématique qui va être visualisé par une table. Il s'en suit:
– une possibilité d'apprentissage par des concepteurs ou des utilisateurs non-programmeurs (dans les modèles hiérarchiques ou en réseau de la génération précédente, concevoir et manipuler une base de données était exclusivement une affaire de programmeurs; ce n'est plus le cas avec le modèle relationnel qui satisfait le principe «d'indépendance» des données[3]);
– une plus grande «communicabilité» du résultat (appelé SCHEMA) de l'application du modèle à un certain univers réel entre différents utilisateurs puisque ce schéma va être exempt (à la différence des schémas hiérarchiques ou en réseau) de toute référence à une implantation ou à une méthode d'accès;

- puissance des opérateurs de manipulation

Les opérateurs relationnels sont des opérateurs ensemblistes «set at a time»: l'application d'un opérateur génère une relation – et donc un ensemble – (d'où un caractère de «fermeture») qui peut à son tour être manipulée; l'utilisateur n'obtient pas un seul enregistrement mais tous les enregistrements qui satisfont le critère de recherche.

Un langage relationnel a donc un caractère dit «non procédural»: l'utilisateur décrit ce qu'il veut, le système écrit la procédure. Dernière conséquence: il n'est plus nécessaire dans un langage relationnel (à la différence des langages hiérarchiques en réseau) d'avoir des instructions d'iteration (pour faire des boucles de recherche): un langage relationnel peut donc être autonome!

- rigueur des concepts (aussi bien des structures que des opérateurs)

Le modèle relationnel, à la différence de tous les autres modèles dits «sémantiques» définis depuis 1970, est le seul à s'appuyer sur une base formelle qui est la logique des prédicats ou la théorie ensembliste. Les concepts de base (structures et opérateurs) ne peuvent donc pas faire l'objet d'interprétation ou de définition ambiguë comme c'est le cas par exemple avec «le modèle» entité/association ou avec certains langages dits de «4ème génération». La conséquence première de la mise en place d'un véritable SGBD relationnel c'est le gain de productivité aussi bien pour l'utilisateur final qui peut interroger une base de données relationnelle sans écrire (ou demander à ce qu'on lui écrive) une procédure, que pour le développeur d'application.

Le gain de développement d'une application de gestion a été estimé de 25 à 75% avec l'utilisation d'un SGBD relationnel, ce qui modifie la gestion des ressources de programmation («turnover» des programmeurs...). Un SGBD relationnel offre donc une réponse cohérente au problème de la croissance du

3. Notons qu'il a fallu 20 ans pour passer d'un modèle à l'universalité de SGBD qui s'en réclament!

coût (humain) de développement de logiciels et à la pénurie annoncée de programmeurs.

Il y aura besoin de moins en moins de programmeurs d'applications (ce n'est pas vrai pour les programmeurs système) et de plus en plus de «méthodologistes», c'est-à-dire des personnes capables d'appliquer le modèle relationnel.

- contribution à la cohérence du système d'information global de l'entreprise: un SGBD-R peut être le support uniforme, s'appuyant donc sur une méthodologie unique, de toutes les applications de gestion ou de production «verticales», d'info-centre «horizontales» ou de serveur. Le «tout relationnel» est en marche avec l'assurance de la pérennité de l'outil qui défie les modes: on en prend pour... 20 ans».

1

MODÈLE RELATIONNEL DE CODD ET LANGAGES RELATIONNELS