

**L'ART
DES BASES
DE DONNEES**

1

**INTRODUCTION
AUX
BASES DE DONNEES**

Serge MIRANDA
José-Maria BUSTA



EYROLLES

26613

L'ART DES BASES DE DONNÉES

**INTRODUCTION
AUX BASES DE DONNÉES**

*... à José qui lutte pour retrouver
l'entrée du monde réel où il fait
bon vivre...*

CHEZ LE MEME EDITEUR

- Des mêmes auteurs :** L'ART DES BASES DE DONNEES
2 - Les bases de données relationnelles
- Autres ouvrages :**
- G. Gardarin BASES DE DONNEES
Les systèmes et leurs langages
- G. Gardarin et P. Valduriez BASES DE DONNEES RELATIONNELLES
Analyse et comparaison des systèmes
- M. Adiba et divers auteurs NOUVELLES PERSPECTIVES DES BASES DE DONNEES
- J. Akoka LES SYSTEMES DE GESTION DE BASES DE DONNEES
Théorie et pratique
- C. Chrisment et F. Bascans MISE EN OEUVRE DES BASES DE DONNEES
Principes méthodologiques
- C. Rolland, O. Foucaut et G. Benci CONCEPTION DES SYSTEMES D'INFORMATION
La méthode REMORA
- J.P. Matheron COMPRENDRE MERISE
Outils conceptuels et organisationnels
- J.B. Crampes CONCEPTION DES SYSTEMES BUREAUTIQUES
- C. Delannoy INITIATION A DBASE III/DBASE III PLUS
- F. Ghoche et J.C. Guillemot DBASE III ET DBASE III PLUS
Le manuel pratique
- G. Grigorieff TOUTE LA PUISSANCE DE DBASE II/DBASE III
EN UN INSTANT

L'ART DES BASES DE DONNÉES

1

INTRODUCTION AUX BASES DE DONNÉES

par

S.M. MIRANDA

*Docteur d'État (UPS)
Master of Sciences (UCLA)
Professeur d'Informatique
Faculté des Sciences
Parc Valrose
06034 Nice Cedex (France)*

J.M. BUSTA

*Docteur
Directeur du Centre de Calcul
Université de Saint-Jacques-de-Compostelle
(La Coruña) Espagne*

TROISIEME EDITION
nouveau tirage


EYROLLES

61, boulevard Saint-Germain — 75005 Paris
1988

Si vous désirez être tenu au courant de nos publications, il vous suffit d'adresser votre carte de visite au :

Service « Presse », Editions EYROLLES,
61, Boulevard Saint-Germain
75240 PARIS CEDEX 05

en précisant les domaines qui vous intéressent.
Vous recevrez régulièrement un avis de parution des nouveautés en vente chez votre libraire habituel.

« La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite » (alinéa 1^{er} de l'article 40) ».

« Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal. ».

PRÉFACE

Cet « Art des Bases de Données » que S. MIRANDA et J.M. BUSTA présentent au fil des deux tomes de cet ouvrage, est particulièrement bienvenu au moment où l'on parle des systèmes de bases de données de 3^e génération.

Car dans ce domaine aussi, on a pris l'habitude de parler en termes de générations. La première, qui a vu le jour vers le milieu des années 60, était caractérisée par des systèmes hiérarchiques (IMS) ou réseau (IDS) relativement rigides et complexes à mettre en œuvre. L'introduction du modèle relationnel par E.F. CODD en 1970 a constitué une étape essentielle dans l'histoire des Bases de Données. En fournissant un support théorique solide aux chercheurs du domaine, le modèle relationnel allait donner naissance à une deuxième génération de systèmes. Après un démarrage prudent, on ne compte plus, aujourd'hui, le nombre de Systèmes de Gestion de Bases de Données relationnels commercialisés, aussi bien sur grands systèmes que sur micro-ordinateurs. Plus souples, plus évolutifs, plus simples à manipuler, ils procurent des avantages indéniables pour des utilisateurs non-informaticiens.

Mais déjà se profile la troisième génération de Systèmes de Bases de Données. Les systèmes existants, relationnels ou non, aussi sophistiqués soient-ils, ne savent prendre en compte que des informations très structurées, comme celles manipulées dans les applications de gestion. Les systèmes de CAO, d'EAO, de bureautique, les systèmes experts, les systèmes documentaires manipulent d'autres types d'informations : des graphiques, des images, du texte, de la voix...

Pour pouvoir les prendre en compte, il faut des Systèmes de Bases de Données autre que ceux qui existent aujourd'hui, autre par les concepts de base, qui devraient profiter des travaux réalisés en intelligence artificielle, autre aussi par l'architecture, qui devrait intégrer les technologies nouvelle (machines bases de données, disque optique numérique, base de données réparties). De nombreuses équipes de recherche se sont attelées à cette tâche difficile, laissant espérer pour les années à venir cette 3^e génération de Systèmes de Bases de Données.

Acteurs de cette aventure nouvelle, après avoir fait leurs armes dans l'étude et la réalisation de systèmes relationnels, S. MIRANDA et J.M. BUSTA sont donc particulièrement bien placés pour présenter les concepts, décrire les réalisations et aborder le devenir des Systèmes de Bases de Données. Leur livre couvre toute l'histoire de ces systèmes et donne les lignes de force des futurs développements. A ce titre il est tout à fait complet. C'est une mine d'informations que les qualités pédagogiques des auteurs, qui sont non seulement chercheurs mais également enseignants, ont permis de rendre claires et digestes. De nombreux exemples, une bibliographie très riche, font de ce livre un outil indispensable pour les étudiants, les chercheurs, les ingénieurs qui veulent être à jour dans ce domaine.

Enfin, ce livre fournit aussi une certaine image des activités de recherche menées en France sur les bases de données. Ces recherches sont loin d'être négligeables. Sous l'impulsion du projet SIRIUS, auquel S. MIRANDA et J.M. BUSTA ont participé, un nouvel élan a été donné. Des actions de recherche sur les bases de données de 3^e génération ont été initialisées. Il n'en reste pas moins que le potentiel de chercheurs et d'ingénieurs est nettement insuffisant. Et on ne peut donc que souhaiter que ce livre contribue à orienter de nombreux étudiants vers ce domaine passionnant.

Jean LE BIHAN.

Ex-directeur du projet SIRIUS
sur les BD réparties et ex-directeur
de l'Aide à la Recherche à l'ADI.

TABLE DES MATIÈRES

PRÉFACE	V
AVANT-PROPOS à la 3 ^e édition	XIII
AVANT-PROPOS	XV
REMERCIEMENTS	XX
I. — INFORMATISATION ET BASE DE DONNÉES	1
1. — Les développements technologiques en cours	1
2. — L'informatisation de la société et ses dangers	6
3. — La ressource « donnée » ou l'or « gris »	9
II. — L'OUTIL BASE DE DONNÉES	11
1. — Concepts de base	12
1.1. — <i>Dualité donnée-information</i>	12
1.2. — <i>Dualité base de données-fichier</i>	14
1.3. — <i>Dualité base de données-banque de données</i>	14
2. — Définitions	15
3. — Objectifs de l'approche BD	17
3.1. — <i>Intégration et corrélation</i>	18
3.2. — <i>Flexibilité (Indépendance)</i>	20

3.3. -- <i>Disponibilité</i>	21
3.4. -- <i>Sécurité</i>	22
4. -- Architecture fonctionnelle d'un SGBD : l'architecture de l'ANSI-SPARC	22
4.1. -- <i>Architecture de l'ANSI-SPARC comme outil d'implantation d'un système d'information</i>	23
4.2. -- <i>Architecture de l'ANSI-SPARC comme conséquence de l'intégration du concept d'indépendance</i>	27
4.3. -- <i>Architecture de l'ANSI-SPARC</i>	31
4.3.1. -- <i>Proposition de l'ANSI-SPARC</i>	31
4.3.2. -- <i>Analyse de la proposition de l'ANSI-SPARC</i>	31
4.3.3. -- <i>Génération des bases de données en fonction de l'architecture de l'ANSI-SPARC</i>	34
5. -- Le schéma conceptuel	35
5.1. -- <i>Approche descendante ou ascendante de définition d'un schéma conceptuel</i>	35
5.2. -- <i>Caractéristiques d'un bon schéma conceptuel</i>	36
6. -- Cadre uniforme de représentation logique des données	37
6.1. -- <i>Approche intuitive du modèle conceptuel</i>	38
6.2. -- <i>Modèle conceptuel OLE</i>	39
6.2.1. -- <i>OBJET</i>	39
6.2.2. -- <i>LIEN</i>	40
6.2.2.1. -- <i>Types de lien.</i> -- 6.2.2.2. -- <i>Lien direct.</i> -- 6.2.2.3. -- <i>Couverture minimale d'un ensemble de liens</i> -- 6.2.2.4. -- <i>Optimisation du schéma OLE.</i> -- 6.2.2.5. -- <i>Validation du schéma OLE</i>	44
6.2.3. -- <i>ENTITE - objets + lien fonctionnel</i>	47
6.2.4. -- <i>Visualisation du schéma conceptuel</i>	49
6.2.5. -- <i>Traduction du schéma OLE par un modèle de données existant</i>	50
6.2.6. -- <i>Langages de manipulation de données (modèle OLE)</i>	51
7. -- État de l'art des recherches concernant le modèle conceptuel	51
III. -- SURVOL DES PRINCIPAUX MODÈLES LOGIQUES	55
Présentation de l'entreprise modélisée	55
1. -- Le modèle hiérarchique de base	56
1.1. -- <i>Représentation du schéma : un arbre ordonné</i>	57
1.2. -- <i>Mise à jour des données : les anomalies de stockage</i>	61
1.2.1. -- <i>Anomalie de destruction</i>	62
1.2.2. -- <i>Anomalie d'insertion</i>	62
1.2.3. -- <i>Anomalie de modification</i>	62
1.3. -- <i>Recherche d'informations : dissymétrie de traitement</i>	62
1.3.1. -- <i>Primitive d'accès direct à la racine d'un sous-arbre</i>	63
1.3.2. -- <i>Primitive d'accès séquentiel aux dépendants d'un segment</i>	64

1.3.3. — Primitive d'accès séquentiel au segment suivant ...	64
1.4. — <i>Avantages et inconvénients du modèle hiérarchique</i> ...	65
2. — Le modèle en réseau	66
2.1. — <i>Réprésentation du schéma : un réseau d'entités et de pointeurs logiques</i>	66
2.2. — <i>Mise à jour des données : une absence d'anomalies de stockage</i>	74
2.3. — <i>Recherche d'informations : une « navigation » de l'utilisateur</i>	74
2.4. — <i>Avantages et inconvénients du modèle en réseau</i>	76
3. — Le modèle relationnel	77
3.1. — <i>Réprésentation du schéma : une vue tabulaire des données</i>	77
3.1.1. — Concepts de base	79
3.1.2. — Visualisation d'une relation (définie d'une manière extensive)	82
3.1.3. — Différences fichier-relation	83
3.1.4. — Schéma et sous-schéma relationnel — Langage de définition de données	84
3.1.5. — Intégrité d'entité ; Intégrité de référence ; SGBD totalement relationnel	88
3.1.6. — Terminologie ANSI-SPARC	89
3.2. — <i>Mise à jour des données : normalisation et « quasi absence » des anomalies de stockage</i>	90
3.3. — <i>Recherche d'informations : non-procéduralité des langages de requête</i>	91
3.4. — <i>Avantages et inconvénients du modèle relationnel</i>	93
4. — Critique des modèles logiques présentés	95
4.1. — <i>Choix d'un modèle logique</i>	95
4.2. — <i>Comparaison des trois approches de base</i>	95
IV. — SURVOL DES PRINCIPAUX MODÈLES PHYSIQUES DE DONNÉES	97
1. — Définitions	97
1.1. — <i>Piste / Cylindre / Volume / Cellule-mémoire</i>	97
1.2. — <i>Enregistrement logique / physique, fichier, index, chemin d'accès</i>	98
2. — Modèles physiques de données	101
2.1. — <i>Organisation séquentielle</i>	101
2.1.1. — Méthode d'accès séquentiel simple	102
2.1.2. — Méthode d'accès séquentiel indexé	103
2.1.3. — Comparaison des deux méthodes séquentielles	107
2.2. — <i>Organisation à accès direct</i>	107
2.2.1. — Méthodes d'accès direct	108
2.2.2. — Hachage statique	108
2.2.3. — Hachage dynamique	110
2.3. — <i>Organisation par liste de pointeurs</i>	117

2.3.1. — Organisation multi-liste	118
2.3.2. — Organisation par liste inverse	122
2.4. — Organisation arborescente	127
2.4.1. — B-arbre	128
2.4.2. — Variantes des B-arbres : B*-arbre, B+-arbre, VSAM	134
2.5. — Classification des organisations de stockage à accès rapide	139

V. — SYSTÈMES DE GESTION DE BASE DE DONNÉES HIÉRARCHIQUES COMMERCIALISÉES

143

1. — IMS (« Information Management System »)	143
1.1. — Définition des données	144
1.1.1. — Définition du schéma	146
1.1.2. — Définition des sous-schémas : les BD logiques ou LBD	148
1.2. — Manipulation des données	150
1.2.1. — DL-1	151
1.2.1.1. — Opérateurs de recherche de données. — 1.2.1.2. — Opérateurs de mise à jour	151
1.2.1.3. Exemples de requêtes DL-1	
1.2.2. — Utilisation du langage hôte	155
1.2.2.1. — Structure d'interface avec COBOL. — 1.2.2.2. — Exécution d'un programme par DL-1	
1.3. — Liaison logiques : un réseau d'ordre deux	158
1.3.1. — Représentation interne des liens 1:N	159
1.3.1.1. — Implantation sous forme de PDB indépendants. — 1.3.1.2. — Implantation sous-forme de PDB liés logiquement	
1.3.2. — Représentation interne des liens N:M	163
1.3.2.1. — Pairage physique. — 1.3.2.2. — Pairage logique. — 1.3.2.3. — BD logique et réseau d'ordre deux	
1.3.3. — Résolution des anomalies de stockage avec IMS	167
1.4. — Organisation physique des données	167
1.4.1. — HSAM	169
1.4.2. — HISAM	170
1.4.2.1. — Chargement initial (avec redondance physique). — 1.4.2.2. — Opérations de stockage. — 1.4.2.3. — DSG secondaire	
1.4.3. — HDAM	175
1.4.3.1. — Chargement initial. — 1.4.3.2. — Opérations de stockage	
1.4.4. — HIDAM	178
1.4.4.1. — Chargement initial. — 1.4.4.2. — Opérations de stockage. — 1.4.4.3. — DSG secondaire	
1.4.5. — Résumé comparatif des différentes méthodes d'accès IMS	182
2. — S2000 (System 2000)	183
2.1. — Définition des données	183

2.2. — <i>Manipulation des données</i>	185
2.3. — <i>Organisation physique des données</i>	187
2.4. — <i>Résumé</i>	188
VI. — SYSTÈMES DE GESTION DE BASE DE DONNÉES EN RÉSEAU COMMERCIALISÉS	189
1. — SGBD commercialisé qui suit le modèle DBTG : IDS-II ...	189
1.1. — <i>Définition des données</i>	189
1.1.1. — Exemple de définition d'un schéma	191
1.1.2. — Commentaires sur la définition du schéma	192
1.2. — <i>Manipulation des données</i>	197
1.2.1. — Fonctions principales du LMD	197
1.2.2. — Indicateurs de positionnement et table d'interface .	198
1.2.3. — Exemples de traitement d'une requête	199
1.2.3.1. — Exemple détaillé. — 1.2.3.2. — Autres exemples	
1.3. — <i>Organisation physique des données</i>	206
1.3.1. — Techniques générales d'implantation d'un coset ...	206
1.3.1.1. — Contiguïté physique. — 1.3.1.2. — Pointeurs. —	
1.3.1.3. — Contiguïté et pointeurs. — 1.3.1.4. Tableau de bits.	
— 1.3.1.5. — Arbre	
1.3.2. — Technique de chaînage par pointeur dans IDS-II ..	209
1.3.3. — Le langage de contrôle des supports physiques (DMCL)	209
2. — Interface de manipulation non procédurale : BIPS	210
2.1. — <i>Premier niveau de traitement : requêtes intra-enregistrement</i>	211
2.2. — <i>Deuxième niveau de traitement : requêtes inter-enregistre-</i>	
<i>ment suivant des liens réels</i>	211
2.2. — <i>Troisième niveau de traitement : requêtes inter-enregistre-</i>	
<i>ment suivant des liens virtuels</i>	212
VII. — LEXIQUE FRANÇAIS/ANGLAIS/ESPAGNOL	215
VIII. — RÉFÉRENCES	219