

# **Introduction pratique aux bases de données relationnelles**

Deuxième édition

**Springer**

*Paris*

*Berlin*

*Heidelberg*

*New York*

*Hong Kong*

*Londres*

*Milan*

*Tokyo*

Andreas Meier

# **Introduction pratique aux bases de données relationnelles**

Deuxième édition

*Traduit de l'allemand par : Dac Hoa Nguyen*

 Springer

**Andreas Meier**  
Département d'Informatique  
Université de Fribourg  
Boulevard de Pérolles 90  
CH-1700 Fribourg  
Suisse

Traduction de l'ouvrage allemand :

*Relationale Datenbanken* par Andreas Meier

Copyright © Springer-Verlag Berlin Heidelberg 1992,1995,1998

Tous droits réservés

---

ISBN-10 : 2-287-25205-3 Springer Paris Berlin Heidelberg New York  
ISBN-13 : 978-2-287-25205-1 Springer Paris Berlin Heidelberg New York

© Springer-Verlag France 2002, 2006  
Imprimé en France

Springer-Verlag France est membre du groupe Springer Science + Business Media

Cet ouvrage est soumis au copyright. Tous droits réservés, notamment la reproduction et la représentation, la traduction, la réimpression, l'exposé, la reproduction des illustrations et des tableaux, la transmission par voie d'enregistrement sonore ou visuel, la reproduction par microfilm ou tout autre moyen ainsi que la conservation des banques données. La loi française sur le copyright du 9 septembre 1965 dans la version en vigueur n'autorise une reproduction intégrale ou partielle que dans certains cas, et en principe moyennant les paiements des droits. Toute représentation, reproduction, contrefaçon ou conservation dans une banque de données par quelque procédé que ce soit est sanctionnée par la loi pénale sur le copyright.

L'utilisation dans cet ouvrage de désignations, dénominations commerciales, marques de fabrique, etc., même sans spécification ne signifie pas que ces termes soient libres de la législation sur les marques de fabrique et la protection des marques et qu'ils puissent être utilisés par chacun.

La maison d'édition décline toute responsabilité quant à l'exactitude des indications de dosage et des modes d'emplois. Dans chaque cas il incombe à l'utilisateur de vérifier les informations données par comparaison à la littérature existante.

SPIN: 11403517

*À Heiri, Ramani et Tina*

# Avant-propos

Nous observons tous une constante mutation structurelle de l'emploi, marquée au cours du temps par le transfert de la main-d'oeuvre du secteur agricole vers le secteur des industries manufacturières, puis vers celui des services et du traitement de l'information. De nos jours, les métiers de l'information prédominent indéniablement sur les autres branches professionnelles. La main-d'oeuvre occupée à produire, traiter et diffuser l'information gagne toujours plus en importance. Dans plusieurs pays européens, le besoin en informaticiens de gestion qualifiés reflète cette évolution vers la société de l'information et une économie numérique.

Avec les infrastructures de télécommunication, les systèmes de gestion de bases de données, qui mettent à notre portée leurs langages et leurs fonctionnalités d'analyse de données, constituent les technologies clés de la société de l'information. Dans le commerce électronique, nous consultons les catalogues de produits orientés Web pour récolter des informations dans la phase préparatoire de vente. Nous accédons aux bases de données clientèle pour préparer des offres, élaborer et exécuter des contrats. Grâce aux bases de données en ligne nous supervisons la distribution de nos produits, quelle que soit leur nature, numérique ou matérielle. Un marketing personnalisé selon la clientèle est inconcevable sans base de données ou entrepôt de données.

Le présent ouvrage a pour but d'exposer un panorama de la technologie des bases de données avec une approche orientée vers la pratique, depuis la conception des bases de données relationnelles jusqu'au développement des systèmes de gestion de bases de données post-relationnelles. Après cette *introduction avancée aux systèmes de bases de données relationnelles*, l'informaticien qui désire oeuvrer dans ce domaine aura acquis les aptitudes suivantes :

- appliquer la technologie relationnelle à la modélisation des données et la conception des bases de données ;
- mettre en pratique les connaissances acquises sur les langages relationnels de requête et de manipulation de données ;
- comprendre les opérations internes et les mécanismes implantés dans un système de bases de données relationnelles ;
- identifier et intégrer dans ses réflexions les atouts et faiblesses de la technologie relationnelle, évaluer avec justesse les futurs développements.

Ce manuel s'adresse à la fois aux *praticiens*, aux *responsables de la formation* en entreprises, aux *enseignants* et *étudiants* des universités et des Grandes Écoles, et à tous ceux qui désirent *s'initier* à la technologie des bases de données relationnelles *par une approche pratique*. Une place centrale sera faite aux concepts essentiels qui sont souvent mal compris et incorrectement appliqués. Le livre est richement illustré par des figures simples et éloquentes. À la fin de chaque chapitre, une bibliographie sélective proposera au lecteur intéressé des ouvrages avancés pour approfondir les différents sujets abordés.

Ce livre d'introduction traitera de la technologie des bases de données relationnelles sous différents angles, englobant les méthodes de conception, les langages et les concepts fondamentaux de l'architecture des systèmes de bases de données relationnelles. Les thèmes seront développés indépendamment des produits offerts par les éditeurs de logiciels de bases de données pour *accorder une large place aux méthodes et techniques fondamentales* et mettre l'accent sur la compréhension liée à l'usage des bases de données relationnelles. Cette publication comble ainsi une lacune dans la littérature axée sur la pratique dans ce domaine.

La cinquième édition, revue et augmentée, présente un nouveau chapitre sur l'intégration des bases de données sur le Web qui comprend aussi la transformation des schémas de bases de données et la migration de données. Le chapitre 6, consacré aux bases de données postrelationnelles, contient une nouvelle section sur l'application de

la logique floue (bases de données floues). Une étude de cas dans le domaine du tourisme sera proposée : il s'agit d'une application de bases de données dont les étapes de conception et de mise en oeuvre avec Access seront expliquées en détail. Le chapitre de révision a été mis à jour et augmenté de nouvelles questions de compréhension accompagnées de solutions modèles.

Ce livre résulte d'un programme de formation en entreprise dans le domaine bancaire, enrichi par les discussions dans le cadre des cours académiques : «Praxis relationaler Datenbanken» (pour informaticiens et ingénieurs) à l'École Polytechnique Fédérale de Zürich, «Informationssysteme und Datenbanken» et «Informatique de gestion I» (pour étudiants en Informatique de gestion) à l'Université de Fribourg en Suisse. De nombreux experts d'entreprises et collègues universitaires ont contribué à la clarté du texte et la pertinence des figures. Je tiens à remercier Urs Bebler, Eirik Danielsen, Bernardin Denzel, Samuel Charles Fasel, Emmerich Fuchs, Peter Gasche, Caroline Grässle-Mutter, Michael Hofmann, Stefan Hüsemann, Günther Jakobitsch, Hans-Peter Joos, Klaus Küspert, Gitta Marchand, Michael Matousek, Thomas Myrach, Mikael Norlindh, Michel Patcas, Fabio Patocchi, Ernst-Rudolf Patzke, Thomas Rätz, Werner Schaad, August Scherrer, Walter Schnider, Max Vetter et Gerhard Weikum. Hartmut Wedekind a étudié l'ouvrage en détail et contribué, par ses précieuses suggestions, à la structure et au contenu de la cinquième édition. Mes compliments vont à Andreea Ionas dont le travail patient et minutieux a permis d'adapter le texte et les symboles à la nouvelle mise en page du livre. Anke Hees a assuré la relecture du livre dans un court délai. Mes remerciements s'adressent tout particulièrement à mon collègue, Dac Hoa Nguyen, qui, par ses compétences, son esprit critique et son inlassable travail, a réalisé de manière professionnelle la traduction de cet ouvrage. Je suis également reconnaissant à Springer-Verlag, en particulier à Nathalie Huilleret et Nicolas Puech pour leur collaboration très appréciée et notamment pour leur relecture attentive et leur correction du manuscrit.

Fribourg, juillet 2005

Andreas Meier



# Table des matières

<b>1</b>	<b>Vers un système de gestion de données</b> .....	1
1.1	Principes fondamentaux du modèle relationnel .....	1
1.2	SQL, langage normalisé au niveau international.....	4
1.3	Les composants d'un système de bases de données relationnelles .....	8
1.4	Organisation de la mise en œuvre des bases de données.....	11
1.5	Notes bibliographiques .....	14
<b>2</b>	<b>Les phases de la construction d'un modèle de données</b> .....	17
2.1	De l'analyse à la base de données .....	17
2.2	Le modèle entité-association .....	20
2.2.1	Entités et associations.....	20
2.2.2	Les types d'associations .....	22
2.2.3	Généralisation et agrégation.....	25
2.3	Le schéma d'une base de données relationnelle .....	30
2.3.1	Le passage du modèle entité-association au schéma de base de données relationnelle .....	30
2.3.2	Règles de passage pour les ensembles de liens .....	33
2.3.3	Règles de passage pour la généralisation et l'agrégation .....	37
2.4	Les dépendances entre données et les formes normales.....	41
2.4.1	La signification et le but des formes normales.....	41
2.4.2	Les dépendances fonctionnelles .....	44
2.4.3	Les dépendances transitives .....	47
2.4.4	Les dépendances multivaluées .....	50
2.5	Les contraintes d'intégrité structurelles.....	53
2.6	L'architecture de données d'entreprise est vitale .....	57
2.7	Guide de la construction d'une base de données .....	61
2.8	Notes bibliographiques .....	64
<b>3</b>	<b>Langages de requête et de manipulation des données</b> .....	67
3.1	Exploitation d'une base de données .....	67
3.2	Les bases de l'algèbre relationnelle .....	69
3.2.1	Vue d'ensemble des opérateurs .....	69

3.2.2	Les opérateurs ensemblistes .....	71
3.2.3	Les opérateurs relationnels .....	75
3.3	Les langages relationnels complets .....	81
3.4	Aperçu des langages relationnels .....	83
3.4.1	SQL.....	83
3.4.2	QUEL.....	87
3.4.3	QBE .....	89
3.5	Les langages immergés.....	92
3.6	Traitement des valeurs nulles .....	94
3.7	La protection des données .....	96
3.8	La formulation des contraintes d'intégrité .....	100
3.9	Notes bibliographiques .....	103
<b>4</b>	<b>Les composants de l'architecture d'un système de bases de données .....</b>	<b>105</b>
4.1	Vue d'ensemble de l'architecture du système.....	105
4.2	Traduction et optimisation des requêtes .....	108
4.2.1	Construction d'un arbre d'interrogation.....	108
4.2.2	Optimisation des requêtes par transformation algébrique.....	111
4.2.3	Évaluation de l'opérateur de jointure .....	114
4.3	Fonctionnement d'un système de bases de données multi-utilisateur.....	118
4.3.1	Le concept de transaction .....	118
4.3.2	La sérialisabilité .....	120
4.3.3	Approches pessimistes .....	124
4.3.4	Approches optimistes .....	129
4.4	Structures de stockage et d'accès .....	131
4.4.1	Structures arborescentes .....	131
4.4.2	Méthodes de hachage .....	135
4.4.3	Structures de données multidimensionnelles .....	138
4.5	Traitement des erreurs .....	142
4.6	Architecture détaillée du système.....	145
4.7	Notes bibliographiques .....	148
<b>5</b>	<b>Intégration et migration des bases de données .....</b>	<b>149</b>
5.1	Exploitation d'ensembles de données hétérogènes .....	149
5.2	Les bases de données sur le Web .....	151
5.2.1	Création d'un système d'information orienté web .....	151
5.2.2	Documents et schémas XML.....	153
5.2.3	Le langage de requête XQuery.....	156

---

5.3	Règles de conversion pour l'intégration et la migration des données .....	158
5.3.1	Conversion des ensembles d'entités simples et des groupes répétitifs .....	159
5.3.2	Conversion des ensembles d'entités dépendants.....	161
5.3.3	Les conversions indirectes pour l'intégration et la migration des données.....	164
5.4	Variantes de migration des bases de données hétérogènes .....	167
5.4.1	Caractérisation des variantes de migration.....	168
5.4.2	Duplication des bases de données sous contrôle du système .....	171
5.5	Principes de la planification de l'intégration et de la migration.....	174
5.6	Notes bibliographiques.....	178
<b>6</b>	<b>Les systèmes de bases de données post-relationnelles .....</b>	<b>181</b>
6.1	Évolution future : pourquoi et dans quelle direction .....	181
6.2	Les bases de données réparties .....	182
6.3	Les bases de données temporelles .....	188
6.4	Les bases de données relationnelles-objet .....	192
6.5	Les bases de données multidimensionnelles .....	197
6.6	Les bases de données floues .....	203
6.7	Les bases de connaissances .....	210
6.8	Notes bibliographiques.....	214
	<b>Révision.....</b>	<b>219</b>
	<b>La mise en œuvre d'une base de données avec Access :</b>	
	<b>l'agence de voyage travelblitz.....</b>	<b>231</b>
	<b>Glossaire .....</b>	<b>259</b>
	<b>Lexique anglais-français .....</b>	<b>267</b>
	<b>Bibliographie.....</b>	<b>273</b>
	<b>Index.....</b>	<b>285</b>

# Préface

Les Systèmes de Gestion de Bases de Données (SGBD) ont supplanté les Systèmes de Gestion de Fichiers au début des années 70. Ils apportaient des concepts nouveaux et essentiels : intégration des données, concurrence, reprise et confidentialité. Les premiers systèmes, définis par la norme CODASYL du DBTG étaient basés sur le modèle dit réseau. Défini au début des années 70, le modèle relationnel a vu ses premières versions industrielles dix années plus tard, et son véritable succès commercial vingt années plus tard. De nombreuses tentatives de le remplacer par un modèle plus élaboré ont eu lieu par la suite, mais aucun essai, soit de créer un nouveau vendeur de SGBD, soit d'introduire une technologie vraiment différente n'a réussi. Ni les modèles dit sémantiques, ni les bases de données objet, ne sont parvenus à s'imposer commercialement, soit parce que les vendeurs n'ont pu réussir commercialement, soit parce que le poids des systèmes patrimoine était trop fort. Aujourd'hui, une trentaine d'années après l'apparition des premiers systèmes, la technologie des systèmes de gestion de bases de données est mature et établie et représente une industrie de plusieurs milliards d'euros de chiffre d'affaire annuel. En étudiant cette industrie, on peut dégager plusieurs tendances :

**Le modèle relationnel s'est imposé comme naturel et omniprésent.** Le modèle simple introduit par Codd en 1970 est maintenant uniformément adopté comme standard de stockage de données secondaires. Il a le grand mérite de la simplicité, et de sa correspondance directe avec le modèle des tableurs connu et adopté par tous. Il a aussi l'avantage de l'existence d'une technologie de langage de requête largement basée sur le langage SQL. Il bénéficie aussi du long travail de standardisation du comité ANSI qui a produit les versions successives de la norme du langage. Donc, même si ce modèle n'est pas le plus naturel pour représenter des données de type

nouveau (XML, objets de langages tels que Java ou C++), des extensions ont été faites pour permettre de les accommoder.

**Le marché s'est consolidé autour de quatre acteurs principaux.** Après la disparition des acteurs initiaux tels Unify ou Interbase, après l'absorption d'acteurs clé comme Informix ou Ingres, il ne reste plus sur le marché que quatre vendeurs : Oracle, Microsoft, IBM et Sybase. Chacun de ces acteurs a une taille critique importante, investit de façon significative en recherche et développement, fournit une offre complète de produits et services associés, et a dans son portefeuille d'outils logiciels de nombreux autres produits complémentaires.

**Les SGBD sont devenus des commodités.** Les systèmes de gestion de bases de données font maintenant partie du paysage traditionnel du logiciel. Ils se sont imposés irrévocablement tant pour les applications spécialisées que pour les progiciels horizontaux ou verticaux. Pratiquement aucune application n'est aujourd'hui développée au-dessus d'un gestionnaire de données autre qu'un SGBD. Dès qu'il y a besoin d'une grande quantité de données, dès qu'il y a un partage de données entre plusieurs utilisateurs différents, dès qu'il y a un besoin de fiabilité des données, dès qu'il y a besoin de sécurisation ou de confidentialité des données, l'application est développée avec un SGBD. Le même phénomène s'est produit pour les progiciels commerciaux, qu'ils soient verticaux, comme des systèmes de CAO, des systèmes de gestion de documents, ou horizontaux, comme des systèmes de gestion du personnel, des systèmes de paye ou des systèmes de gestion de type ERP (enterprise resource planning) ou SCM (supply chain management) : chacun de ces progiciels, soit contient une version incluse de SGBD, soit doit pour fonctionner être lié à l'exécutif d'un SGBD.

**La technologie continue à évoluer.** Paradoxalement pour une technologie mature et parvenue au stade de la commodité, l'évolution technologique reste importante. Elle est due à des facteurs multiples : pression de la concurrence qui reste vive entre les quatre acteurs du marché, investissements technologiques importants consentis par les vendeurs, et demandes nouvelles du marché. Des extensions sont proposées régulièrement et progressivement intégrées dans les

systemes, que se soit pour prendre en compte des nouveaux phénomènes comme l'apparition de XML, de nouveaux langages de programmation comme Java dans les années 90 ou C++ dans les années 2000. Il convient donc de comprendre comment cette technologie évolue et ce que sont les nouveaux produits.

Il est donc essentiel pour tous les professionnels de l'informatique de comprendre et de maîtriser cette technologie. Ce livre présente l'ensemble des éléments nécessaires à cette maîtrise et cette compréhension. Il présente un panorama moderne et mis à jour du modèle relationnel, des notions de modélisation de données, des langages de requêtes essentiels. Il couvre de façon complète et détaillée l'ensemble des concepts essentiels des systèmes : traitement de requêtes, structures de stockage et méthodes d'accès, concurrence et reprise. Il constitue donc un outil essentiel pour tout enseignant, chercheur, ou praticien qui s'intéresse à cette technologie.

Andreas Meier allie une profonde connaissance de la théorie des bases de données (il est l'auteur de nombreux articles sur le sujet), à une expérience détaillée de l'utilisation pratique des SGBD (il a participé dans l'industrie à la mise en place de grands projets d'application). Ce double point de vue lui permet de mettre en valeur dans cet ouvrage tant les fondements théoriques essentiels à la maîtrise de cette technologie que l'aspect concret indispensable à l'utilisateur de ces systèmes.

François Bancilhon

# Préface à l'édition allemande

Pour un ouvrage spécialisé en informatique, la parution de la cinquième édition marque un événement rare qui mérite que nous nous interroguions sur les raisons de cette remarquable pérennité.

La réponse suivante s'avère appropriée, mais encore trop simplifiée : Andreas Meier propose un livre écrit de manière compréhensible aux praticiens et aux apprenants grâce à son orientation vers la pratique. Explorons la question corollaire suivante : Pourquoi le livre de Meier est-il compréhensible ? J'apporterai une réponse plus élaborée à cette interrogation :

Pour les apprenants, la bonne compréhensibilité du présent ouvrage repose sur trois caractéristiques. Premièrement, l'auteur s'adresse aux lecteurs de niveau débutant dans le domaine des bases de données. Le langage et les figures s'harmonisent tout au long du livre. L'apprenant débutant ne se perd pas dans des abstractions de haut niveau avec toute la précision requise. Deuxièmement, l'auteur présente la matière aux apprenants dans un contexte où elle devrait se placer, tant dans sa profondeur que dans son ampleur. À cet égard, la recherche académique et l'application industrielle sont deux contextes fondamentalement différents. Meier met en exergue l'application industrielle. Troisièmement, le souci de "justesse" est un point sensible pour tout auteur : elle signifie non pas simplement la vérité logique, mais la conformité aux approches de modélisation reconnues qui sont en constante évolution. Le monde des bases de données, issues des traditionnels systèmes de fichiers implantés dans les systèmes d'exploitation, a connu une évolution continue des modèles depuis plus de trois décennies. Nous étions partis du modèle CODASYL, nous avons ensuite découvert l'approche relationnelle, et l'Internet met aujourd'hui à notre portée le plus grand système de bases de données que nous puissions imaginer. Prédire les conséquences dans le futur n'est pas une tâche facile. La seule

certitude est que nous ne nous dirigeons pas vers des eaux calmes. Dans cette perspective, l'auteur a augmenté la cinquième édition d'un important chapitre traitant de l'intégration des bases de données hétérogènes sur le Web.

Je souhaite que la cinquième édition touche un large lectorat. Devant l'évolution constatée ci-haut, il n'y a pas à douter qu'une sixième édition se profilera à l'horizon.

Erlangen, juillet 2003

Hartmut Wedekind