

Bases de données

Concepts, utilisation et développement

Cours et exercices corrigés

Jean-Luc Hainaut

Professeur à la Faculté d'informatique
Université de Namur

2^e édition

Dunod - 20012



Grand éditeur de SGBD relationnel observant avec intérêt l'émergence d'une nouvelle technologie de gestion de données¹.

1. Référence allégorique à la section 1.7.8. *Linogravure*, B. Hainaut, 2004

Table des matières

AVANT-PROPOS	27
CHAPITRE 1 • MOTIVATION ET INTRODUCTION	33
1.1 Des données et des hommes	33
1.2 Une première approche des données	34
1.3 Utiliser une base de données	38
1.4 Premières conclusions	39
1.5 Les systèmes de gestion de bases de données	40
1.6 Les défis des bases de données d'aujourd'hui	41
1.7 Un peu d'histoire	42
1.7.1 Le modèle collectiviste : CODASYL DBTG	43
1.7.2 La pensée unique : IMS	44
1.7.3 La naissance du modèle relationnel	44
1.7.4 Apparition d'un nouvel acteur : l'utilisateur final	45
1.7.5 La guerre des modèles	45
1.7.6 La chute d'Icare	46
1.7.7 Le modèle relationnel : la suite de l'histoire	46
1.7.8 Cronos dévore ses enfants	47
PARTIE 1	
CONCEPTS DES BASES DE DONNÉES	
CHAPITRE 2 • CONCEPTS DES BASES DE DONNÉES	49
2.1 Introduction	49
2.2 Tables, lignes et colonnes	50
2.3 La valeur null	51
2.4 Les identifiants	52
2.5 Les clés étrangères	54

2.6	Schéma et contenu d'une base de données	55
2.7	Représentation graphique d'un schéma	55
2.8	Un exemple de base de données	57
2.9	Note sur les clés étrangères	58
2.10	Modification et contraintes d'intégrité	61
2.10.1	Les colonnes obligatoires	61
2.10.2	Les contraintes d'unicité (identifiants)	61
2.10.3	Les contraintes référentielles (clés étrangères)	61
2.11	Redondances internes dans une table	63
2.11.1	Le phénomène de redondance interne	64
2.11.2	Normalisation par décomposition	65
2.12	Les structures physiques	66
2.13	Que retenir ?	68
2.14	Pour en savoir plus	69
2.15	Pratiquer les bases de données	69
2.16	Exercices	70
CHAPITRE 3 • MODÈLE RELATIONNEL ET NORMALISATION		73
3.1	Le phénomène de redondance interne	73
3.2	Le modèle relationnel de bases de données	74
3.2.1	Concepts de base	74
3.2.2	Opérateurs d'extraction de données	75
3.3	Identifiants	78
3.4	Dépendances fonctionnelles	79
3.4.1	Le phénomène de dépendance fonctionnelle	79
3.4.2	Un exemple plus complexe	80
3.4.3	Graphe ADF d'une relation (attributs et DF)	81
3.4.4	Propriétés des dépendances fonctionnelles	81
3.5	Contraintes d'inclusion et clés étrangères	84
3.6	Calcul des identifiants d'une relation	86
3.6.1	Relation entre identifiants et dépendances fonctionnelles	86
3.6.2	La relation possède un seul identifiant	87
3.6.3	La relation possède plusieurs identifiants	88
3.7	Décomposition d'une relation	89
3.8	Normalisation d'une relation	91
3.8.1	Réexamen du phénomène de redondance interne	91
3.8.2	Analyse du phénomène	92
3.8.3	Relation normalisée	93

3.8.4	Normalisation d'une relation	93
3.8.5	Normalisation d'un noyau irréductible	96
3.8.6	SGBD et dépendances fonctionnelles	98
3.9	Que retenir ?	99
3.10	Pour en savoir plus	100
3.11	Exercices	100
CHAPITRE 4 • IMPLÉMENTATION DES STRUCTURES DE DONNÉES		103
4.1	Introduction	103
4.2	Les composants d'un ordinateur	104
4.3	Mémoire externe : le disque magnétique	105
4.3.1	Géométrie d'un disque	105
4.3.2	Lecture et écriture de données sur un disque	107
4.3.3	Optimisation : tampon d'échange et lecture anticipée	109
4.4	Organisation d'un espace de stockage	111
4.5	Traitement séquentiel d'un fichier	115
4.5.1	Lecture séquentielle d'un fichier	116
4.5.2	Tri d'un fichier séquentiel	118
4.6	Les index	122
4.7	Organisation séquentielle indexée	124
4.7.1	Structure d'un fichier séquentiel indexé	124
4.7.2	Lecture et modification d'un fichier séquentiel indexé	125
4.7.3	Caractéristiques et performances d'un fichier séquentiel indexé	127
4.7.4	Étude de cas	134
4.8	Organisation calculée	135
4.8.1	Principes de l'accès calculé	135
4.8.2	Fonction de calcul d'adresse	136
4.8.3	Gestion des débordements	138
4.8.4	Caractéristiques et performances d'un fichier à accès calculé	139
4.8.5	Comparaison des deux techniques d'index primaires	142
4.9	Les index secondaires	143
4.9.1	Structure d'un index secondaire par liste de pointeurs	143
4.9.2	Performances d'un index secondaire par liste de pointeurs	144
4.9.3	Structure d'un index bitmap	147
4.9.4	Les index secondaires augmentés	147
4.9.5	Disponibilité des techniques d'indexation primaires et secondaires	148
4.10	Accès par index ou recherche séquentielle ?	148
4.11	Les techniques d'agrégation (clustering)	149
4.11.1	Clustering index	149

4.11.2	Clusters	150
4.11.3	Performances des techniques d'agrégation	150
4.12	Que retenir ?	151
4.13	Pour en savoir plus	152
4.14	Exercices	152
CHAPITRE 5 • LES SYSTÈMES DE GESTION DE BASES DE DONNÉES		155
5.1	Introduction	155
5.2	Les fonctions des systèmes de gestion de bases de données	156
5.3	Architecture des SGBD	158
5.4	Les transactions	160
5.5	Régulation de la concurrence	162
5.5.1	Mise à jour perdue	163
5.5.2	Régulation de la concurrence par verrouillage	163
5.5.3	Sérialisabilité et verrouillage en 2 phases	164
5.5.4	Interblocage	165
5.6	Protection contre les accidents	165
5.6.1	Sauvegardes et journaux	165
5.6.2	Reprise à froid	166
5.6.3	Reprise à chaud	166
5.7	Contrôle des accès	167
5.7.1	Les modèles de contrôle d'accès	167
5.7.2	Protection des données privées - Inférence statistique	168
5.8	Le marché des SGBD	170
5.9	Pour en savoir plus	171

PARTIE 2

UTILISATION DES BASES DE DONNÉES

CHAPITRE 6 • LE LANGAGE SQL DDL		173
6.1	Introduction	173
6.2	Le langage SQL DDL	174
6.3	Création d'un schéma	174
6.4	Création d'une table	175
6.4.1	Les colonnes et leurs types	175
6.4.2	Les identifiants	177
6.4.3	Les clés étrangères	178

6.4.4	Caractère obligatoire/facultatif d'une colonne	178
6.4.5	Exemple complet	178
6.4.6	Forme synthétique des contraintes	179
6.5	Suppression d'une table	179
6.6	Ajout, retrait et modification d'une colonne	180
6.7	Ajout et retrait d'une contrainte	180
6.8	Les structures physiques	181
6.9	Que retenir ?	182
6.10	Pour en savoir plus	182
CHAPITRE 7 • LE LANGAGE SQL DML (1)		183
7.1	Introduction	183
7.2	Requêtes élémentaires	185
7.2.1	Extraction simple	185
7.2.2	Extraction de lignes sélectionnées	186
7.2.3	Lignes dupliquées dans le résultat	187
7.2.4	Des conditions de sélection plus complexes	188
7.3	Données extraites et données dérivées	190
7.3.1	Les fonctions SQL	191
7.3.2	Les registres du système	193
7.4	Les fonctions agrégatives (ou statistiques)	194
7.5	Les sous-requêtes	196
7.5.1	Condition par sous-requêtes	196
7.5.2	Sous-requête et clé étrangère multi-composants	197
7.5.3	Attention aux conditions d'association négatives	197
7.5.4	Références multiples à une même table	200
7.6	Les quantificateurs ensemblistes	202
7.7	Exercices	204
CHAPITRE 8 • LE LANGAGE SQL DML (2)		207
8.1	Introduction	207
8.2	Extraction de données de plusieurs tables (jointure)	208
8.2.1	La jointure de plusieurs tables	208
8.2.2	Conditions de jointure et conditions de sélection	209
8.2.3	La jointure et les lignes célibataires	210
8.3	Les opérateurs ensemblistes	211
8.4	Le produit relationnel	212
8.5	Requêtes sur des structures de données cycliques	213

8.6	Compléments sur les jointures	217
8.6.1	Sous-requête ou jointure ?	218
8.6.2	Valeurs dérivées dans une jointure	220
8.6.3	Les jointures généralisées	220
8.6.4	Interprétation du résultat d'une jointure	221
8.7	Extraction de données groupées	224
8.7.1	Notion de groupe de lignes	224
8.7.2	Sélection de groupes et sélection de lignes	226
8.7.3	Groupes et jointures	227
8.7.4	Composition du critère de groupement	228
8.7.5	Attention aux groupements multi-niveaux	229
8.7.6	Peut-on éviter l'utilisation de données groupées ?	230
8.8	Ordre des lignes d'un résultat	230
8.9	Interprétation d'une requête	232
8.10	Modification des données	232
8.10.1	Ajout de lignes	233
8.10.2	Suppression de lignes	234
8.10.3	Modification de lignes	234
8.10.4	Mise à jour et contraintes référentielles	235
8.10.5	Déterminisme des requêtes de modification	237
8.11	Modification des structures de données	240
8.12	Que retenir ?	241
8.13	Pour en savoir plus	241
8.14	Exercices	241
CHAPITRE 9 • LE LANGAGE SQL AVANCÉ		245
9.1	Introduction	245
9.2	Le contrôle d'accès	246
9.3	Les vues SQL	248
9.3.1	Principes et objectifs des vues	248
9.3.2	Définition et utilisation d'une vue	248
9.3.3	Les vues comme interface pour des besoins particuliers	249
9.3.4	Les vues comme mécanisme de contrôle d'accès	250
9.3.5	Les vues comme mécanisme d'évolution de la base de données	250
9.3.6	Les vues comme aide à l'expression de requêtes complexes	250
9.3.7	Mise à jour des données via une vue	251

9.4	Extension de la structure des requêtes SFW	252
9.4.1	Extension de la clause select	252
9.4.2	Extension de la clause from	252
9.5	Les requêtes récursives	258
9.6	Les extensions objet de SQL3	260
9.6.1	Types de données complexes (row et array)	261
9.6.2	Type défini par l'utilisateur (TDU)	262
9.6.3	Table typée	262
9.6.4	Hiérarchie de types	262
9.6.5	Hiérarchie de tables typées	264
9.6.6	Références entre tables	265
9.6.7	Remarques finales	266
9.7	Les prédicats (check)	267
9.8	Les procédures SQL (stored procedures)	268
9.9	Les déclencheurs (triggers)	269
9.10	Le catalogue	274
9.11	Les interfaces SQL des programmes d'application	277
9.11.1	L'interface SQL statique classique	278
9.11.2	SQLJ : une interface statique pour Java	282
9.11.3	L'interface SQL dynamique classique	284
9.11.4	JDBC : une interface CLI-SQL dynamique pour Java	287
9.11.5	Comparaison des modèles d'interaction	289
9.11.6	Un problème de sécurité : l'injection de code SQL	290
9.12	SQL et l'information incomplète	292
9.12.1	Introduction	292
9.12.2	La valeur null de SQL	292
9.12.3	La logique ternaire de SQL	292
9.12.4	La propagation de null en SQL	293
9.12.5	La propagation de unknown en SQL	293
9.12.6	Les problèmes de l'information incomplète en SQL	295
9.12.7	Deux recommandations	299
9.13	Que retenir ?	300
9.14	Pour en savoir plus	301

PARTIE 3
DÉVELOPPEMENT DES BASES DE DONNÉES

CHAPITRE 10 • CONSTRUCTION D'UNE BASE DE DONNÉES	303
10.1 De l'artisanat à la méthode	303
10.2 Les acteurs et les métiers	304
10.3 Deux parcours	305
10.4 Une démarche simplifiée	305
10.5 Les outils	306
CHAPITRE 11 • LE MODÈLE ENTITÉ-ASSOCIATION DE BASE	307
11.1 Introduction	307
11.2 Les types d'entités	308
11.3 Les attributs	309
11.4 Les types d'associations	310
11.4.1 Rôles d'un type d'associations	310
11.4.2 Propriétés d'un type d'associations	311
11.4.3 Classe fonctionnelle d'un type d'associations	311
11.4.4 Type d'associations (ou rôle) obligatoire ou facultatif	314
11.4.5 Cardinalités d'un type d'associations	314
11.4.6 Type d'associations cyclique	314
11.4.7 Représentation graphique des populations	316
11.5 Une question de vocabulaire	316
11.6 Les identifiants	317
11.6.1 Identifiants hybrides	317
11.6.2 Composition des identifiants	319
11.6.3 Identifiants minimaux et identifiants implicites	320
11.6.4 Importance du concept d'identifiant	321
11.7 Les contraintes d'intégrité	322
11.7.1 Contraintes de base	322
11.7.2 Contraintes additionnelles	322
11.8 Contenu informationnel d'un schéma	322
11.9 Exemples	324
11.9.1 Une structure administrative	324
11.9.2 Gestion d'une bibliothèque	325
11.9.3 Voyages en train	326
11.10 Que retenir ?	327
11.11 Pour en savoir plus	328

11.12 Dessiner un schéma conceptuel à l'aide d'un outil	328
11.13 Exercices	330
CHAPITRE 12 • ÉLABORATION D'UN SCHÉMA CONCEPTUEL	333
12.1 Introduction	333
12.2 Un premier exemple	335
12.3 Décomposition de l'énoncé	336
12.3.1 Notion de proposition élémentaire	336
12.3.2 Cardinalité	337
12.3.3 Propositions générales et propositions particulières	338
12.3.4 Attention aux propositions complexes irréductibles !	339
12.3.5 Les propositions non binaires réductibles	341
12.4 Pertinence d'une proposition	342
12.5 Représentation d'une proposition	343
12.5.1 Nouveau type d'entités et son attribut	343
12.5.2 Type d'associations entre types d'entités existants	343
12.5.3 Attribut d'un type d'entités existant	344
12.5.4 Nouveau type d'entités et ses deux attributs	344
12.5.5 Restructuration pour intégration	344
12.5.6 Propriétés multivaluées	348
12.5.7 Recommandations pratiques	351
12.6 Non-redondance des propositions	352
12.6.1 Redondance explicite	352
12.6.2 Variantes d'expression	352
12.6.3 Redondance indirecte	352
12.6.4 Redondance apparente	353
12.6.5 Synonymes et homonymes	354
12.7 Non-contradiction des propositions	354
12.8 Les contraintes d'intégrité	356
12.9 Normalisation du schéma	356
12.9.1 Simplification du schéma	356
12.9.2 Élimination des redondances internes	357
12.10 Documentation du schéma	359
12.11 Complétude du schéma	359
12.12 Que retenir ?	360
12.13 Pour en savoir plus	361
12.14 Exercices	361

CHAPITRE 13 • PRODUCTION DU SCHÉMA DE LA BASE DE DONNÉES	365
13.1 Introduction	365
13.2 Représentation des types d'entités	366
13.3 Représentation des attributs	366
13.4 Représentation des types d'associations	366
13.4.1 Types d'associations un-à-plusieurs (1:N)	366
13.4.2 Types d'associations un-à-un (1:1)	368
13.4.3 Types d'associations plusieurs-à-plusieurs (N:N)	370
13.4.4 Types d'associations cycliques	371
13.5 Représentation des identifiants	371
13.6 Traduction des noms	372
13.7 Synthèse des règles de traduction	373
13.8 Les structures physiques	374
13.9 Traduction des structures en SQL	376
13.10 Que retenir ?	377
13.11 Pour en savoir plus	378
13.12 Produire un schéma de base de données à l'aide d'un outil	378
13.13 Exercices	379
CHAPITRE 14 • MÉTHODOLOGIE DES BASES DE DONNÉES	381
14.1 Introduction	381
14.2 Méthodes, processus et produits	382
14.3 Méthodes de développement d'une base de données	383
14.4 Les modèles de développement	388
14.5 Techniques transformationnelles	390
14.6 Rôle des outils	390
CHAPITRE 15 • LE MODÈLE ENTITÉ-ASSOCIATION ÉTENDU	391
15.1 Introduction	391
15.2 Le modèle de base - rappel	392
15.3 Sémantiques	393
15.4 Modèle, schéma et population	394
15.5 Les stéréotypes	396
15.6 Schéma de base de données	397
15.7 Attributs	397
15.7.1 Attributs composés	397
15.7.2 Attributs multivalués	398

15.7.3	Attributs complexes	398
15.7.4	Domaine et type d'un attribut	399
15.7.5	Cardinalité d'un attribut	400
15.8	Types d'associations	401
15.8.1	Types d'associations fonctionnels	401
15.8.2	Types d'associations n-aires	401
15.8.3	Types d'associations à rôle multitypes d'entités	402
15.8.4	Attributs de type d'associations	403
15.8.5	Types d'associations de types d'associations	403
15.8.6	Généralisation des cardinalités des rôles	404
15.8.7	Taille de la population d'un type d'associations	404
15.9	Types d'entités	405
15.9.1	Relation de généralisation-spécialisation (relation <i>is-a</i>)	405
15.9.2	Propriétés et contraintes ensemblistes des sous-types	406
15.9.3	L'héritage	407
15.9.4	Surtypes et héritage multiples	409
15.9.5	Contraintes portant sur des composants hérités	411
15.9.6	Répartitions multiples	412
15.9.7	Règle de répartition en sous-types	413
15.9.8	Types d'entités et classes Java	414
15.9.9	Du mauvais usage des relations <i>is-a</i>	415
15.9.10	Relations <i>is-a</i> entre d'autres composants du modèle	416
15.10	Identifiants	416
15.10.1	Identifiants d'un type d'entités	416
15.10.2	Identifiant cyclique d'un type d'entités	418
15.10.3	Identifiant global de types d'entités	419
15.10.4	Identifiants d'un type d'associations	420
15.10.5	Identifiants d'un attribut complexe	422
15.11	Contraintes d'intégrité	423
15.11.1	Contraintes de domaine	424
15.11.2	Contraintes entre valeurs	424
15.11.3	Contraintes d'existence	424
15.11.4	Dépendances fonctionnelles	426
15.12	Contraintes d'intégrité additionnelles	427
15.12.1	Expression des contraintes	427
15.12.2	Remarques sur les langages d'expression de contraintes	428
15.12.3	Cycles et contraintes cycliques	429
15.12.4	Propriétés des graphes d'instances	430
15.12.5	Contraintes de redondance	431

15.13 Les noms	432
15.13.1 Règles de dénomination	432
15.13.2 Noms identiques, noms absents et stéréotypes	432
15.13.3 Dénomination des attributs	433
15.13.4 Dénomination des types d'associations	434
15.14 Classes d'objets génériques	435
15.14.1 Les types d'entités génériques	435
15.14.2 Les types d'associations génériques	436
15.15 Opérations et méthodes	439
15.16 Exemple de schéma Entité-association	440
15.17 Métadonnées et métaschémas	440
15.17.1 Une première approche : le catalogue d'une base de données	441
15.17.2 Métaschéma du modèle Entité-association	442
15.18 Sémantique du modèle Entité-association	443
15.18.1 Types d'entités et relations <i>is-a</i>	444
15.18.2 Types d'associations	444
15.18.3 Attributs	445
15.18.4 Attributs composés et multivalués	445
15.18.5 Attributs et rôles obligatoires	446
15.18.6 Identifiants	446
15.18.7 Autres contraintes	448
15.18.8 Exemple complet	449
15.18.9 Application de la théorie relationnelle au modèle Entité-association	449
15.18.10 Interprétation relationnelle et traduction relationnelle	451
15.19 Que retenir ?	451
15.20 Pour en savoir plus	452
15.21 Créer un schéma conceptuel à l'aide d'un outil (compléments)	453
15.22 Exercices	454
CHAPITRE 16 • LES DIAGRAMMES DE CLASSES UML	459
16.1 Introduction	459
16.2 UML et les bases de données	460
16.3 Les mécanismes d'extension	461
16.4 Les classes d'objets	462
16.5 Les associations binaires	462
16.6 Les associations n-aires	463
16.6.1 Multiplicité maximale (max)	464
16.6.2 Multiplicité minimale (min)	464

16.7	Les attributs	466
16.8	Les opérations	467
16.9	Les associations qualifiées	467
16.10	Les classes associations	469
16.11	Les associations d'agrégation et de composition	469
16.12	Les contraintes	470
16.13	Le langage de contraintes OCL	471
16.14	Une question de vocabulaire	474
16.15	Exemple de schéma de classes en UML	474
16.16	Le modèle de classes d'UML revisité (DB-UML)	474
16.17	Sémantique du modèle de classes DB-UML	476
16.18	Que retenir ?	477
16.19	Pour en savoir plus	478
16.20	Créer un schéma conceptuel UML à l'aide d'un outil	478
CHAPITRE 17 • ANALYSE CONCEPTUELLE DU DOMAINE D'APPLICATION		479
17.1	Introduction	479
17.2	Analyse des énoncés textuels	482
17.2.1	Proposition existentielle	482
17.2.2	Proposition relationnelle binaire	483
17.2.3	Proposition relationnelle n-aire	486
17.2.4	Intégration d'une proposition dans un schéma	490
17.3	Les patrons de conception	491
17.4	Analyse des documents structurés	493
17.4.1	Analyse des documents de l'organisation	494
17.4.2	Analyse d'un écran de saisie de données	495
17.4.3	Analyse d'un document papier	496
17.5	Dénomination des objets d'un schéma	497
17.6	Processus de correction d'un schéma	498
17.6.1	Constructions syntaxiquement incorrectes	499
17.6.2	Constructions non satisfiables	500
17.7	Processus de normalisation du schéma conceptuel	506
17.7.1	Simplification des constructions non minimales	508
17.7.2	Explicitation des constructions insuffisamment expressives	512
17.7.3	Réduction des sur-spécifications	517
17.7.4	Traitement des constructions anormales	519
17.7.5	Uniformisation des constructions irrégulières	521
17.7.6	Élimination des redondances de contrainte	522

17.7.7	Élimination des redondances structurelles d'instances	524
17.7.8	Élimination des redondances internes	527
17.7.9	Amélioration de la présentation graphique	530
17.7.10	Mise en conformité aux standards en vigueur	533
17.8	Évaluation d'un schéma	536
17.8.1	Paraphrase du schéma	537
17.8.2	Prototypage de la base de données	537
17.8.3	Inférence structurelle	538
17.8.4	Les heuristiques d'erreurs	538
17.8.5	Remarque	539
17.9	Intégration de schémas	540
17.9.1	Scénarios d'intégration	540
17.9.2	Principes de l'intégration binaire de schémas	541
17.9.3	Recherche des correspondances	542
17.9.4	Unification des schémas	546
17.9.5	Fusion des schémas unifiés	551
17.9.6	Résolution des conflits sémantiques	551
17.9.7	Compléments	552
17.9.8	Étude de cas	553
17.10	Analyse conceptuelle à l'aide d'un atelier	558
17.10.1	Transformation de schémas	558
17.10.2	Analyse de schémas	560
17.10.3	Intégration de schémas	560
17.11	Que retenir ?	561
17.12	Pour en savoir plus	562
17.13	Exercices	563
CHAPITRE 18 • CONCEPTION LOGIQUE D'UNE BASE DE DONNÉES RELATIONNELLE		569
18.1	Introduction	569
18.2	Le modèle logique relationnel	571
18.2.1	Le modèle relationnel de base	571
18.2.2	Le modèle relationnel enrichi	572
18.3	Représentation des attributs	574
18.3.1	Représentation d'un attribut simple	574
18.3.2	Représentation d'un attribut composé	574
18.3.3	Représentation d'un attribut multivalué	577
18.3.4	Représentation d'un attribut composé multivalué	578
18.3.5	Représentation d'un attribut objet	578
18.3.6	Traduction des types et domaines de valeurs	579
18.3.7	Compléments	580

18.4	Retour aux types d'associations fonctionnels	581
18.5	Représentation des types d'associations complexes	582
18.5.1	Types d'associations n-aires, avec attributs et/ou contraintes	582
18.5.2	Représentation d'un type d'associations à rôle multitypes	582
18.5.3	Clé étrangère et identifiant absent	583
18.6	Représentation des relations <i>is-a</i>	584
18.6.1	Les trois techniques de base	585
18.6.2	Représentation de relations <i>is-a</i> par matérialisation	586
18.6.3	Représentation de relations <i>is-a</i> par héritage descendant	588
18.6.4	Représentation de relations <i>is-a</i> par héritage ascendant	590
18.6.5	Évaluation comparative	591
18.6.6	Hiérarchies multiples et clés étrangères liées	592
18.7	Compléments	593
18.7.1	Les identifiants facultatifs	593
18.7.2	Les identifiants primaires complexes	594
18.7.3	Traduction des contraintes conceptuelles	594
18.7.4	Noms des objets	595
18.7.5	<i>Delete</i> et <i>update modes</i> des clés étrangères	595
18.7.6	Au sujet des rôles de cardinalité [1-N]	596
18.8	Représentation d'un schéma conceptuel	597
18.8.1	Choix des représentations privilégiées	597
18.8.2	Construction du plan de transformation	598
18.8.3	Limitations d'un plan de transformation	599
18.8.4	Validation d'un plan de transformation	600
18.9	Définition des vues	601
18.10	Conception logique : un exemple	601
18.11	Exploitation des constructions SQL3	602
18.11.1	Le modèle relationnel objet	605
18.11.2	Représentation des attributs complexes	605
18.11.3	Représentation des hiérarchies <i>is-a</i> de types d'entités	606
18.11.4	Conception logique relationnelle objet : un exemple	607
18.11.5	Extension de la méthode	607
18.12	Produire un schéma logique à l'aide d'un outil	607
18.12.1	Transformation de schémas	607
18.12.2	Analyse de schémas	609
18.13	Que retenir ?	609
18.14	Pour en savoir plus	610
18.15	Exercices	611

CHAPITRE 19 • CONCEPTION PHYSIQUE D'UNE BASE DE DONNÉES RELATIONNELLE	617
19.1 Introduction	617
19.2 Notion d'efficacité d'une base de données	619
19.3 Conception physique d'une base de données en 16 règles	620
19.4 Détermination des index	620
19.4.1 Index induits par les structures logiques	620
19.4.2 Index induits par l'utilisation des données	621
19.4.3 Index à écarter	622
19.4.4 Maintenance du jeu d'index	623
19.5 Technologie des index	624
19.6 Sélection des espaces de stockage	624
19.7 Optimisation avancée	625
19.8 Estimation des volumes et des temps de traitement	628
19.9 Que retenir ?	629
19.10 Pour en savoir plus	629
19.11 Exercices	630
CHAPITRE 20 • PRODUCTION DU CODE D'UNE BASE DE DONNÉES	633
20.1 Introduction	633
20.2 Expression déclarative des structures	634
20.3 Les concepts SQL d'expression de contraintes	635
20.4 Les comportements non standard	638
20.5 Les contraintes d'intégrité de base	639
20.6 Les contraintes d'intégrité additionnelles	640
20.7 Les contraintes dérivées du modèle Entité-association	643
20.8 Les contraintes d'intégrité dynamiques	645
20.9 Traduction des relations <i>is-a</i>	645
20.10 Traduction des vues	649
20.11 Que retenir ?	649
20.12 Génération du code à l'aide d'un outil	650
20.13 Exercices	651
CHAPITRE 21 • RÉTRO-INGÉNIERIE D'UNE BASE DE DONNÉES	653
21.1 Introduction	653
21.2 Méthode générale de rétro-ingénierie	657
21.3 Extraction physique	658

21.4	Reconstruction du schéma logique	659
21.4.1	Les sources et leurs techniques d'analyse	659
21.4.2	Recherche des constructions implicites	663
21.5	Conceptualisation du schéma logique	665
21.6	Que retenir ?	669
21.7	Pour en savoir plus	670
21.8	Rétro-ingénierie à l'aide d'un outil	670
CHAPITRE 22 • ÉTUDES DE CAS		671
22.1	Introduction	671
22.2	Les animaux du zoo	671
22.2.1	Énoncé	672
22.2.2	Construction du schéma conceptuel	672
22.2.3	Production du schéma de tables	674
22.2.4	Production du code SQL	674
22.2.5	Exploitation	675
22.3	Circuits intégrés	676
22.3.1	Énoncé	676
22.3.2	Construction du schéma conceptuel	676
22.3.3	Schéma de tables et contenu	681
22.4	Rétro-ingénierie d'une base de données	682
22.4.1	Extraction du schéma physique	682
22.4.2	Extraction du schéma logique	683
22.4.3	Conceptualisation	684
22.5	Autres études de cas	686
BIBLIOGRAPHIE		687
INDEX		695

ANNEXES

Les annexes sont disponibles en ligne sur le site de l'ouvrage, à l'adresse <http://www.info.fundp.ac.be/libd>, onglet Books/Ouvrages ou sur le site de Dunod <http://www.dunod.com>, sur la page dédiée à l'ouvrage.

ANNEXE A - EXERCICES ET SOLUTIONS

A.1	Motivation et introduction	1
A.2	Concepts des bases de données	1
A.3	Modèle relationnel et normalisation	4
A.4	Implémentation des structures de données	10
A.5	Les systèmes de gestion de bases de données	15
A.6	Le langage SQL DDL	16
A.7	Le langage SQL DML (1)	18
A.8	Le langage SQL DML (2)	29
A.9	Le langage SQL avancé	42
A.10	Construction d'une base de données	44
A.11	Le modèle Entité-association de base	45
A.12	Élaboration d'un schéma conceptuel	49
A.13	Production du schéma de la base de données	67
A.14	Méthodologie des bases de données	71
A.15	Le modèle Entité-association étendu	71
A.16	Les diagrammes de classes UML	83
A.17	Analyse conceptuelle du domaine d'application	84
A.18	Conception logique d'une base de données relationnelle	102
A.19	Conception physique d'une base de données relationnelle	116
A.20	Production du code d'une base de données	119

ANNEXE B - SQL, LES ENSEMBLES ET LA LOGIQUE

B.1	SQL et les mathématiques	1
B.2	Ensemble c'est tout	2
B.3	Les multi-ensembles	7
B.4	Un peu de logique	10
B.5	Que retenir ?	21
B.6	Pour en savoir plus	21
B.7	Exercices	22

ANNEXE C - APPLICATIONS AVANCÉES DES BASES DE DONNÉES

C.1	Introduction	1
C.2	Les structures ordonnées	1
C.3	Les bases de données actives	4
C.4	Les données temporelles	8
C.5	La génération de code	18
C.6	Exercices	24

ANNEXE D - CONVENTIONS GRAPHIQUES DES MODÈLES

D.1	Introduction	1
D.2	Le modèle Entité-Association de base	1
D.3	Le modèle Entité-Association étendu	2
D.4	Le modèle logique relationnel	5
D.5	Le modèle logique relationnel objet	6
D.6	Le modèle physique relationnel	8

ANNEXE E - TRANSFORMATIONS DE SCHÉMAS

E.1	Introduction	1
E.2	Caractéristiques générales des transformations	2
E.3	Préservation de la sémantique	4
E.4	Catalogue de transformations	5
E.5	Transformations de mutation	9
E.6	Transformation d'attributs d'un type d'entités	10
E.7	Transformation d'attributs d'un type d'associations	23
E.8	Transformations de types d'entités	26
E.9	Transformation de relations is-a	35
E.10	Transformation de types d'associations	48
E.11	Transformations relatives aux contraintes	58
E.12	Pour en savoir plus	59
E.13	Références et bibliographie	60

ANNEXE F - ÉTUDES DE CAS - SAISON 2

F.1	Introduction	1
F.2	Le porte-feuille d'études de cas	2
F.3	Voyages aériens	3
F.4	Intégration de schémas pour un établissement scolaire	8

F.5	Tournées de distribution de colis	17
F.6	Représentation des graphes	22
F.7	Gestion de clés	44
F.8	Migration de fichiers	50
F.9	Dictionnaire de données	79
F.10	Analyse d'un stack Hypercard	87
F.11	SQL et les ontologies	98

ANNEXE G - CONCEPTION LOGIQUE D'UNE BASE DE DONNÉES RELATIONNELLE OBJET

G.1	Introduction	1
G.2	Le modèle logique relationnel objet	2
G.3	Types d'entités et identifiants	4
G.4	Représentation des attributs complexes	5
G.5	Représentation des types d'associations binaires	6
G.6	Représentation des types d'associations complexes	10
G.7	Représentation des relations is-a	12
G.8	Compléments	17
G.9	Transformation d'un schéma conceptuel	17
G.10	Conception logique : un exemple	19
G.11	Que retenir ?	20
G.12	Pour en savoir plus	21
G.13	Exercices	23

ANNEXE H - COMPLÉMENTS DE TECHNOLOGIE (version provisoire)

H.1	Introduction	
H.2	Traitement séquentiel d'un fichier	
H.3	Organisation séquentielle indexée	
H.4	Organisation calculée	
H.5	Les index secondaires	
H.6	Un exemple : SQL Server de Microsoft	
22.7	Exercices	

Avant-propos

Je voudrais avant tout situer et justifier cet ouvrage. Celui-ci est le fruit de près de trente ans d'enseignement, de recherche et de pratique, tant à l'université que dans les entreprises et les administrations. J'y ai rassemblé et structuré les concepts, les techniques et les méthodes qui me sont apparus les plus *utiles* dans le domaine des bases de données. Je pourrais dire que cet ouvrage ne devrait contenir que ce que j'estime *nécessaire* à l'informaticien, depuis l'étudiant jusqu'au praticien en passant par l'utilisateur motivé.

Le lecteur ne trouvera ici que ce qu'il est susceptible de rencontrer dans sa pratique. Même les sections un peu plus théoriques ont leur utilité, soit pour résoudre un problème concret de manière rigoureuse, soit pour mieux comprendre l'un ou l'autre concept délicat ou controversé. En revanche, le lecteur intéressé par les détails techniques propres à un SGBD particulier trouvera ceux-ci dans l'abondante littérature qui leur est consacrée, tant sur le Web que dans les librairies.

J'ai donc résisté à la tentation de l'*encyclopédisme*, qui voudrait que je traite ici, comme il est souvent d'usage dans un ouvrage généraliste, de thèmes tels que les dépendances multivaluées, la 4^e forme normale, le calcul relationnel (ou Datalog), l'optimisation des requêtes, les bases de données orientées objet, déductives, XML, multimédias ou NoSQL. Malgré leur incontestable intérêt je les ai écartés de la table des matières dans la mesure où ils sont aujourd'hui étrangers, ou à tout le moins marginaux, au cœur de métier de l'informaticien.

Avant de décrire plus en détail les objectifs et le mode d'emploi de l'ouvrage, je voudrais prendre un peu de temps pour remercier tous ceux qui, jadis ou aujourd'hui, de loin ou de près, peu ou prou, ont contribué à son élaboration. Certains ont été, ou sont encore, des collaborateurs de recherche. D'autres sont des collègues avec lesquels j'ai à l'occasion des discussions sur des points particuliers techniques ou théoriques.

Je pense en particulier, dans le désordre, à Véronique Goemans, Yves Delvaux, Lysiane Goffaux, Jean-Pierre Thiry, Alain Delcourt, Benoît Vanhoute, Carine Charlot, Bertrand Mignon, Catherine Tonneau, Muriel Chandelon, Michel Joris, Mario Cadelli, Bernard Decuyper, Olivier Marchand, Jean Henrard, Jean-Marc Hick, Vincent Englebert, Didier Roland, Alain Gofflot, Pierre Delvaux, Anne-France Brogneaux, Arnaud Deflorenne, Stéphane Bodart, Majid Chougrani, Philippe Thiran, Virginie Detienne, Denis Zampunieris, Christine Delcroix, Aurore François, Fabrice Estiévenart, Frédéric Vesentini, Jean-Roch Meurisse, Julien Vilz, Ravi Ramdoyal, Anthony Cleve, Jonathan Lemaitre, Karin Becker, Yannis Tzitzikas, Rokia Bendaoud, François Bodart, Patrick Heymans, Wim Vanhoof et Yves

Pigneur. Une pensée toute particulière à Anne-France, Vincent et Anthony, pour leur aide dans le décryptage de cet hydre capricieux qu'est SQL lorsqu'on tente de le pousser dans ses derniers retranchements. De nombreuses notes de cet ouvrage leur sont dues.

Je remercie tous les étudiants auxquels j'ai enseigné dans les facultés d'informatique, d'économie, de gestion et de droit, cette discipline particulièrement riche et passionnante qu'est celle des bases de données. Leurs questions, leurs suggestions et leurs difficultés sont à l'origine d'une partie significative de cet ouvrage. L'ouvrage leur est tout naturellement dédié. Qu'ils y pensent lorsque, en fin d'année, ils seront tentés de le vouer, lui et son auteur, aux gémonies !

Merci et bravo à ReveR et à ses aventuriers, qui ont à ce point cru en la valeur des résultats des recherches de notre laboratoire qu'ils n'ont pas hésité à fonder en 2003, avec l'aide d'autres illuminés de la même eau, une société commerciale unique au monde. Ils méritent toute mon admiration.

Merci enfin aux maisons d'édition InterEditions et Dunod pour leur aide et leur confiance.

Objectifs et structure de l'ouvrage

L'ouvrage a pour objectif d'offrir au lecteur un texte d'introduction aux bases de données à la fois pratique, moderne et détaillé. Les concepts et les problèmes qu'il aborde sont ceux que le lecteur, quel que soit son niveau de départ et quelles que soient ses attentes (il s'adresse aussi bien à l'*étudiant* qu'au *praticien*) rencontrera sur le terrain. Sa structure générale suit une approche rationnelle : pour apprendre à *utiliser* les bases de données il faut au préalable en *comprendre* les concepts et pour apprendre à *construire* des bases de données il faut savoir les *utiliser*. L'ouvrage est par conséquent constitué de trois parties :

- La première, intitulée **Concepts des bases de données**, est constituée de quatre chapitres qui décrivent les concepts fondamentaux des bases de données : les structures de données, leur implémentation, les systèmes de gestion de bases de données et, pour les lecteurs plus curieux, des éléments du modèle relationnel et de la normalisation.
- La deuxième partie, **Utilisation des bases de données**, décrit les différents aspects du langage SQL, depuis les formes et fonctions élémentaires jusqu'aux fonctions avancées du modèle relationnel objet. Cette partie comprend quatre chapitres.
- La troisième partie, **Développement des bases de données**, décrit de manière détaillée les techniques et les méthodes de construction de bases de données relationnelles et relationnelles objet. On y trouve également une introduction à la rétro-ingénierie des bases de données. Cette partie constituée de treize chapitres est suivie d'une collection d'études de cas complètement résolues.

Des annexes complètent ces chapitres. Elle fournissent notamment un important recueil d'exercices et de solutions, une mise à niveau de certains principes plus théoriques, des applications avancées et des études de cas résolues. Ces annexes, ainsi que des matériaux complémentaires, destinés à l'étudiant, à l'enseignant et au praticien, sont disponibles en ligne sur le site de l'ouvrage.

Les publics cibles

Cet ouvrage est destiné à la fois aux *étudiants et enseignants*, et aux *professionnels* de l'informatique, auxquels il propose deux parcours différents que nous décrivons ci-après.

Les étudiants et enseignants. Dans un curriculum d'enseignement supérieur, le parcours introductif sera utile à l'étudiant de début de premier cycle. Il correspond à un module d'introduction aux bases de données de 30 heures + 15 heures de travaux pratiques. Selon ses objectifs, l'enseignant pourra se servir également de certains chapitres plus spécialisés, tels que le chapitre 3 et l'annexe B, voire les chapitres 15 et 16. Étendu aux chapitres complémentaires, il servira de support à un enseignement plus approfondi de fin de premier cycle ou de deuxième cycle.

Les professionnels. L'ouvrage s'adresse également au lecteur autodidacte désireux de s'initier à la discipline des bases de données. Il sélectionnera, depuis le parcours introductif jusqu'à l'intégralité de l'ouvrage, l'itinéraire correspondant aux exigences de son métier. Le parcours court, qui comporte 10 chapitres, sera choisi par l'informaticien débutant et pressé, le consultant ou le correspondant informatique, tandis que l'analyste et le développeur de bases de données choisiront l'ensemble des chapitres. L'administrateur de données et le programmeur d'application se constitueront des itinéraires intermédiaires individualisés.

Les options de cet ouvrage

L'ouvrage a été structuré et rédigé sur la base de quelques principes suggérés par l'expérience et l'étude des besoins actuels en matière de bases de données.

Une approche pratique et pédagogique. Destiné à un public diversifié et évolutif, l'ouvrage adopte une démarche progressive et propose plusieurs parcours de lecture. Le contenu de l'ouvrage est articulé autour de problèmes et de questions effectivement rencontrés en pratique. Les aspects théoriques sont proposés dans la mesure où ils permettent de résoudre aisément et efficacement ces problèmes. Le texte inclut de nombreux exemples, exercices et études de cas, accompagnés de leurs solutions. Les concepts sont illustrés par des outils disponibles gratuitement (logiciels libres ou logiciels industriels en version Education) que le lecteur est invité à utiliser. La plupart des chapitres se clôturent par une synthèse (*Que retenir ?*), par des pistes d'approfondissement et un état de l'art (*Pour en savoir plus*) et par un jeu d'exercices corrigés.

Un contenu et une approche modernes. L'ouvrage est destiné aux lecteurs confrontés aux problèmes que posent les systèmes d'information et les technologies d'aujourd'hui. Il inclut des résultats récents issus de la recherche et de l'industrie. Ainsi, des méthodes pratiques relatives au *calcul des identifiants*, à la *normalisation de schémas conceptuels*, à la construction de *bases de données relationnelles objet* et à la *rétro-ingénierie* apparaissent-elles pour la première fois dans cet ouvrage. Les méthodes d'ingénierie qu'il décrit sont basées sur des approches transformationnelles et peuvent être mises en œuvre à l'aide d'outils (les AGL).

Une approche rigoureuse. L'ouvrage ne se contente pas de décrire les concepts et les méthodes à la manière d'une encyclopédie. Il construit progressivement un édifice cohérent et raisonné. Les techniques et les méthodes de résolution de problèmes sont présentées, non comme des recettes prêtes à l'emploi, mais comme le résultat naturel de raisonnements objectifs et rigoureux. Le lecteur peut ainsi comprendre la rationalité de ces techniques et méthodes et en appréhender les limites. Il peut dès lors se les approprier et les adapter à ses propres besoins. Certaines parties délicates de l'exposé sont soutenues par des éléments plus théoriques qui en garantissent la validité. Ainsi en est-il de la théorie relationnelle et de la normalisation (chapitre 3).

Une approche critique. Les concepts, les technologies et les méthodes ne sont pas livrées comme allant de soi. Leur histoire, leurs points forts et leurs faiblesses sont largement discutés. Pour citer quelques exemples, les normes SQL2 et SQL3, les valeurs null, le modèle relationnel objet, le modèle Entité-association, les diagrammes de classes d'UML et les options choisies par les implémenteurs de SGBD sont analysés et critiqués, souvent de manière détaillée.

Parcours multiples

L'ouvrage autorise donc différents parcours de lecture en fonction des objectifs et des besoins du lecteur :

- Le lecteur débutant se contentera dans un premier temps des chapitres introductifs dans chacune des trois parties de l'ouvrage (chapitre 1 suivi des chapitres 2, 6, 7, 8, 10, 11, 12, 13), lesquels constituent une première approche complète des différents thèmes abordés.
- Le lecteur poursuivra ensuite la lecture selon ses propres objectifs. Les chapitres 4, 5 et 9 (et les annexes C et H) lui donneront des connaissances techniques et applicatives plus approfondies.
- Sur le plan de la méthodologie (troisième partie), les chapitres 14, 15 et 17 apportent aux concepteurs, consultants et administrateurs de données une spécialisation en analyse conceptuelle. Celle-ci pourra être complétée par l'étude d'UML (chapitre 16).
- Les chapitres 18, 19 et 20 fournissent des méthodes approfondies destinées aux développeurs. Ceux-ci pourraient être en outre intéressés par des techniques complémentaires, telle que la production de bases de données relationnelles objet (annexe G) ou la rétro-ingénierie (chapitre 21).
- Le lecteur plus exigeant sera sans doute aussi intéressé par les compléments théoriques du chapitre 3 (théorie relationnelle). L'annexe B (ensembles et logique) constitue en outre une mise à niveau destinée à combler les éventuelles lacunes du lecteur, en particulier étudiant.
- Le chapitre 22, consacré à trois études de cas est destiné aux différentes catégories de lecteurs, selon leur niveau. L'annexe F propose une dizaine d'autres études dans des domaines très variés.

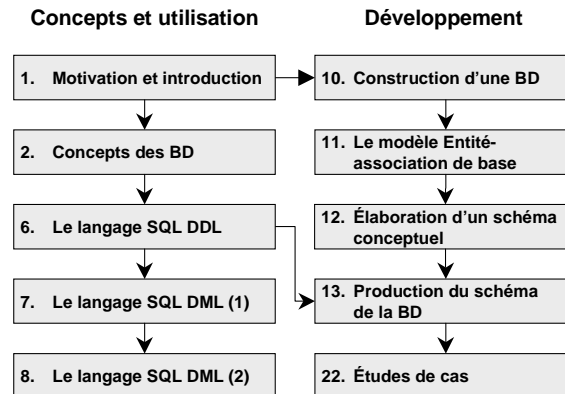


Figure 1 - Parcours de lecture introductif

Certains chapitres constituent pour d'autres des *prérequis*. Certains de ceux-ci sont indispensables, d'autres sont simplement conseillés. La figure 1 représente les chapitres du parcours minimal introductif et leurs prérequis. La figure 2 reprend l'ensemble des chapitres et leurs prérequis pour constituer un parcours standard complet. Les chapitres 16 et 21 sont facultatifs et peuvent être abordés en fonction des besoins. Il en est de même des annexes.

Cet ouvrage est accompagné et complété d'un site web. Celui-ci met en ligne les annexes A à H et des matériaux complémentaires : tutoriels, logiciels recommandés dans l'ouvrage, matériaux didactiques, documents d'approfondissement et errata (que le lecteur attentif consultera régulièrement). Ces matériaux sont disponibles sur le site de l'ouvrage, à l'adresse <http://www.info.fundp.ac.be/libd> (onglet *Books/Ouvrages*) ou sur le site de Dunod (<http://www.dunod.com>) sur la page dédiée à l'ouvrage.

Au sujet de la 2^e édition

Cette édition comporte, par rapport à la précédente, de nombreuses mises à jour, corrections et améliorations de présentation. Le chapitre 4, consacré aux techniques d'implémentation a été considérablement augmenté (il est complété de l'annexe H). Il inclut une description détaillée des technologies communément utilisées ainsi que les modèles de calcul permettant d'estimer leurs performances. Le chapitre 19, consacré à la conception logique relationnelle objet a été reporté en annexe et le chapitre 18 a été complété des éléments les plus utiles de ce modèle de données.

Le document *Dunod-2012-Nouveautés.pdf* reprend de manière plus détaillée les nouveautés de la 2^e édition.

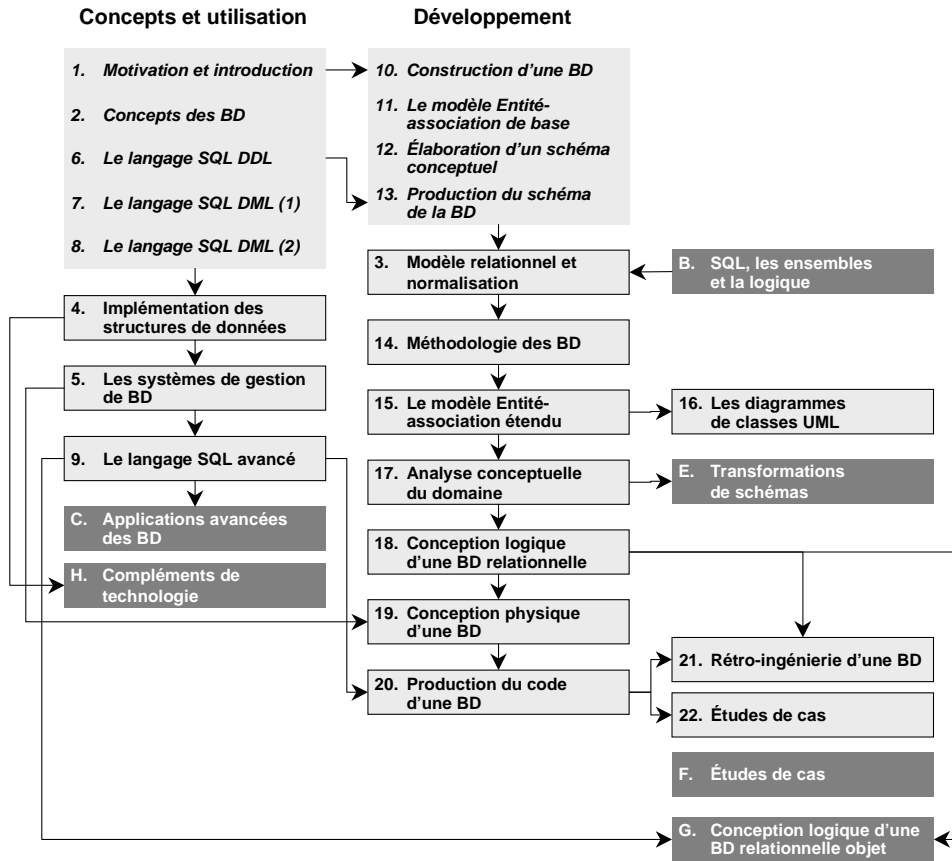


Figure 2 - Parcours de lecture complet

Jean-Luc Hainaut
 Namur, le 1^{er} janvier 2012.

