

MODÉLISATION DES SYSTÈMES D'INFORMATION DÉCISIONNELS

Techniques de modélisation conceptuelle
et relationnelle des entrepôts de données

Emmanuel Ferragu



Vuibert

Table des matières

Préface	XV
----------------------	----

Avant-propos	XVII
---------------------------	------

Objectif du livre et public visé	XVIII
--	-------

Prérequis à la lecture du livre	XVIII
---------------------------------------	-------

Organisation du livre	XVIII
-----------------------------	-------

Introduction : systèmes d'information décisionnels et modélisation multidimensionnelle	1
---	---

Système d'information opérationnel et système d'information décisionnel	1
---	---

Limites du modèle entité-association pour les SID	4
---	---

Avantages du modèle multidimensionnel pour les SID	5
--	---

Architecture des SID	7
----------------------------	---

Data marts en silo	8
--------------------------	---

Corporate information factory (CIF) ou enterprise data warehouse (EDW)	11
--	----

Dimensional data warehouse ou bus architecture	13
--	----

Synthèse	15
----------------	----

PARTIE I - MODÉLISATION CONCEPTUELLE MULTIDIMENSIONNELLE	17
---	----

Chapitre 1

Cas d'école : l'analyse des ventes d'une entreprise de distribution	23
--	----

1.1 Description de l'entreprise	23
---------------------------------------	----

1.2 Besoins d'analyse	24
-----------------------------	----

1.3 Modèle conceptuel multidimensionnel correspondant au cas d'école	24
--	----

Chapitre 2

Dimensions	27
2.1 Définition d'une dimension	27
2.2 Niveaux	28
2.3 Hiérarchies	29
2.3.1 Définition d'une hiérarchie	29
2.3.2 Relations hiérarchiques	30
2.3.3 Forage au sein des hiérarchies	31
2.3.4 Hiérarchies simples et hiérarchies alternatives	33
2.3.5 Hiérarchies équilibrées	34
2.3.6 Hiérarchies déséquilibrées	35
2.3.7 Hiérarchies généralisées	40
2.4 Niveaux interdimensionnels	45
2.5 Relation d'héritage entre niveaux	46

Chapitre 3

Gestion des modifications temporelles intervenant au sein des dimensions	51
3.1 Gestion temporelle de type 1	51
3.2 Gestion temporelle de type 2	54
3.3 Gestion temporelle de type 3	54
3.4 Gestion temporelle hybride ou gestion temporelle de type 12	55
3.5 Absence de gestion temporelle (type 0)	55
3.6 Cas courants d'utilisation des différents types de gestion temporelle	56
3.7 Cas des corrections d'erreurs	56
3.8 Représentation graphique de la gestion temporelle dans un diagramme MCMD	57

Chapitre 4

Relations factuelles	59
4.1 Définition d'une relation factuelle	59
4.2 Grain d'une relation factuelle	60
4.3 Rôle d'un niveau pour une relation factuelle	61
4.4 Cardinalités d'un niveau pour une relation factuelle	61

4.5	Additivité des indicateurs d'une relation factuelle	62
4.5.1	Indicateurs additifs	62
4.5.2	Indicateurs semi-additifs	63
4.5.3	Indicateurs non additifs	63
4.6	Indicateurs élémentaires et indicateurs dérivés	64
4.7	Formules de calcul des indicateurs	64
4.8	Relations factuelles sans indicateurs	66
4.9	Différents types de relations factuelles	68
4.9.1	Relations factuelles élémentaires de type Transaction	68
4.9.2	Relations factuelles élémentaires de type Instantané	69
4.9.3	Relations factuelles élémentaires de type Instantané Accumulé	74
4.9.4	Relations factuelles de synthèse	82
4.9.5	Représentation graphique des types de relations factuelles dans un diagramme MCMD	83
4.10	Ordre d'application des formules d'agrégation d'un indicateur	84

Chapitre 5

Métamodèle et synthèse de la notation graphique des MCMD

5.1	Métamodèle	89
5.2	Synthèse de la notation graphique utilisée pour la représentation des MCMD	92

Chapitre 6

Bonnes pratiques de modélisation d'un MCMD

6.1	Bonne pratique n° 1 : tout schéma factuel d'un MCMD doit comporter une dimension temporelle	95
6.2	Bonne pratique n° 2 : assurer la conformité des dimensions sur l'ensemble d'un MCMD	95
6.3	Bonne pratique n° 3 : assurer la conformité des indicateurs	98
6.4	Bonne pratique n° 4 : définir un MCMD unique pour l'ensemble des activités de l'entreprise	101
6.5	Bonne pratique n° 5 : définir le grain le plus fin possible pour les relations factuelles élémentaires	102
6.6	Bonne pratique n° 6 : ne pas « mélanger » différents types (Transaction, Instantané, Instantané Accumulé) dans une même relation factuelle élémentaire	104

6.7	Bonne pratique n° 7 : définir un grain unique (ou homogène) pour l'ensemble des faits d'une relation factuelle	104
6.8	Bonne pratique n° 8 : modéliser en utilisant des dimensions correspondant à un réel concept métier et non pas pour « coller » à des besoins figés de restitutions	107
6.9	Bonne pratique n° 9 : faire en sorte que chaque indicateur soit pertinent pour l'ensemble des dimensions associées à la relation factuelle à laquelle il appartient	108

PARTIE II - MODÉLISATION RELATIONNELLE MULTIDIMENSIONNELLE

113

Chapitre 7

Principes de dérivation logique d'un MCMD

117

7.1	Principes généraux de dérivation logique	117
7.1.1	Principes généraux de dérivation logique d'un niveau.	117
7.1.2	Principes généraux de dérivation logique d'une relation factuelle	120
7.2	Illustration des principes généraux de dérivation logique	122
7.3	Modèle en étoile et modèle en flocon	123
7.3.1	Différences entre le modèle en étoile et le modèle en flocon.	123
7.3.2	Avantages et inconvénients du modèle en étoile par rapport au modèle en flocon.	125
7.4	Utilisation de clés techniques.	126
7.4.1	Principe et illustration de l'utilisation des clés techniques	126
7.4.2	Intérêt de l'utilisation des clés techniques.	128

Chapitre 8

Modélisation logique des hiérarchies

131

8.1	Modélisation logique des hiérarchies équilibrées	131
8.1.1	Approche « en flocon »	131
8.1.2	Approche « en étoile »	132
8.2	Modélisation logique des hiérarchies déséquilibrées	132
8.2.1	Modélisation logique des hiérarchies à niveau manquant	133
8.2.2	Modélisation logique des hiérarchies incomplètes	142
8.2.3	Modélisation logique des hiérarchies récursives.	148
8.3	Modélisation logique des hiérarchies généralisées	164
8.4	Modélisation logique des hiérarchies alternatives.	167

Chapitre 9

Intégration de la gestion temporelle dans la modélisation logique des dimensions..... 169

9.1	Cas d'une dimension avec uniquement des attributs et relations hiérarchiques dont la gestion temporelle est de type 1.....	169
9.1.1	Gestion de type 1 avec une approche « en étoile ».....	170
9.1.2	Gestion de type 1 avec une approche « en flocon ».....	174
9.2	Cas d'une dimension avec uniquement des champs dont la gestion temporelle est de type 2.....	175
9.2.1	Gestion de type 2 avec une approche « en étoile ».....	176
9.2.2	Gestion de type 2 avec une approche « en flocon ».....	179
9.3	Cas d'une dimension contenant à la fois des attributs et/ou relations hiérarchiques de type 1 et de type 2.....	182
9.4	Cas d'une dimension avec attribut(s) ou relation(s) hiérarchique(s) de type 3.....	183
9.5	Cas d'une dimension avec attribut(s) de type hybride (type 12).....	184
9.6	Cas des corrections d'erreurs.....	185
9.7	Gestion temporelle des hiérarchies récursives.....	186
9.7.1	Gestion temporelle d'une hiérarchie récursive avec la 1 ^{re} approche (table parent-enfant) ..	186
9.7.2	Gestion temporelle d'une hiérarchie récursive avec la 2 ^e approche (table intermédiaire)...	189
9.7.3	Gestion temporelle d'une hiérarchie récursive avec la 3 ^e approche (nested sets ou LR)....	195

Chapitre 10

Modélisation logique des relations factuelles..... 199

10.1	Dérivation logique des indicateurs.....	199
10.1.1	Dérivation logique des indicateurs dérivés.....	199
10.1.2	Dérivation logique des indicateurs élémentaires.....	202
10.2	Niveaux avec rôles multiples.....	204
10.3	Tables de faits sans faits.....	205
10.4	Tables de couverture.....	209
10.5	Dérivation logique des relations factuelles en relation 1..N (ou 0..N) avec un niveau.....	212
10.5.1	Simplification de la relation 1..N par la création de plusieurs relations 1..1.....	212
10.5.2	Création d'une table intermédiaire entre la table de faits et la table de dimension.....	217

Chapitre 11

Techniques complémentaires de modélisation logique des dimensions 223

11.1	Dimensions dégénérées et dimensions fourre-tout.....	223
11.1.1	Dimensions dégénérées.....	223
11.1.2	Tables de dimension « fourre-tout ».....	225
11.2	Dimensions géantes et minidimensions.....	228
11.2.1	Dimensions géantes.....	228
11.2.2	Minidimensions.....	229
11.2.3	Plusieurs minidimensions pour une même dimension géante.....	232
11.2.4	Limitation occasionnée par les minidimensions et solutions de contournement.....	236
11.3	Attributs à forte cardinalité et regroupements par tranches.....	238

Chapitre 12

Traitement des cas particuliers des niveaux interdimensionnels et de l'héritage..... 245

12.1	Dérivation logique des niveaux interdimensionnels.....	245
12.2	Dérivation logique des schémas factuels contenant des relations d'héritage entre niveaux.....	249
12.2.1	Première méthode : une table de dimension unique.....	249
12.2.2	Deuxième méthode : une table de dimension par sous-type.....	251

Chapitre 13

Valeurs NULL et valeurs par défaut 255

13.1	Éviter les valeurs NULL.....	255
13.2	La valeur NULL dans les tables de dimension.....	255
13.3	La valeur NULL dans les clés étrangères des tables de faits.....	257
13.4	La valeur NULL dans les indicateurs.....	260

Chapitre 14

Tables de faits fusionnées et agrégats..... 263

14.1	Tables de faits fusionnées.....	263
14.2	Agrégats.....	266
14.2.1	Exemples de tables de faits agrégées.....	266
14.2.2	Choix des tables de faits agrégées.....	269

14.2.3 Gestion temporelle de type 1 et tables de faits agrégées 271
 14.2.4 Tables de faits agrégées et indicateurs non agrégeables 273
 14.2.5 Navigateur d'agrégats..... 273
 14.2.6 Vues matérialisées 275

PARTIE III - ANNEXES..... 277
Sigles 279
Bibliographie..... 281
Index 283

Préface

Par un jour où l'on se réveille, on se rend compte que le monde est devenu un peu différent. On se réveille avec une certaine inquiétude, on se réveille avec une certaine curiosité. C'est une curiosité qui n'est pas seulement intellectuelle, mais qui est aussi politique. C'est une curiosité qui nous pousse à chercher à comprendre ce qui se passe autour de nous, à chercher à comprendre les raisons de ce qui se passe. C'est une curiosité qui nous pousse à chercher à agir, à chercher à changer ce qui se passe.

Tout cela nous amène à nous interroger sur le rôle de l'information dans notre société. L'information est devenue un bien commun, un bien qui nous concerne tous. Elle est devenue un bien qui nous permet de mieux comprendre le monde, de mieux comprendre les autres, de mieux comprendre nous-mêmes. Elle est devenue un bien qui nous permet de mieux agir, de mieux changer ce qui se passe.

Il nous faut donc nous interroger sur le rôle de l'information dans notre société. Il nous faut nous interroger sur le rôle de l'information dans notre société. Il nous faut nous interroger sur le rôle de l'information dans notre société. Il nous faut nous interroger sur le rôle de l'information dans notre société.

Autre chose, il nous faut nous interroger sur le rôle de l'information dans notre société. Il nous faut nous interroger sur le rôle de l'information dans notre société. Il nous faut nous interroger sur le rôle de l'information dans notre société. Il nous faut nous interroger sur le rôle de l'information dans notre société.