

Méthode de conception d'une base multidimensionnelle contrainte

Faiza Ghozzi, Franck Ravat,
Olivier Teste, Gilles Zurfluh

IRIT/SIG
Université Paul Sabatier
118, Route de Narbonne
31062 TOULOUSE cedex 04

{ghozzi, ravat, teste, zurfluh}@irit.fr

Résumé. Ce papier présente notre méthode de conception de bases de données multidimensionnelles (BDM) contraintes. Nous proposons une méthode mixte basée sur une démarche descendante et une démarche ascendante. La démarche descendante est composée de trois étapes: la collecte des besoins des décideurs, la spécification et la formalisation de ces besoins suivant un modèle multidimensionnel en constellation qui intègre l'expression de contraintes sémantiques. La démarche ascendante collecte les données à partir du schéma de la source opérationnelle et construit un schéma multidimensionnel candidat pour l'aide à la décision suivant un processus semi-automatique. La phase de confrontation permet d'intégrer le schéma multidimensionnel des besoins et le schéma candidat obtenu à partir des sources. L'avantage de notre méthode est qu'elle permet de tirer profit des connaissances présentes dans le schéma conceptuel de la source tout en tenant compte des besoins des décideurs. En outre, notre méthode est basée sur un schéma multidimensionnel contraint permettant de valider l'intégrité sémantique des BDM.

1. Introduction

De nos jours, la plupart des systèmes décisionnels repose sur une approche OLAP facilitant l'analyse interactive et la synthèse d'un grand volume de données [Codd *et al*, 1993]. Dans cette approche, les données sont stockées au travers d'une modélisation multidimensionnelle, autrement dit, les données sont représentées au travers de centre d'intérêt analysés selon différents axes d'analyse [Kimball et Ross, 2002]. L'objet de cet article est la proposition d'une méthode de conception de base décisionnelle établie suivant un schéma multidimensionnel intégrant un ensemble de contraintes sémantiques. En effet, la conception d'une base décisionnelle doit répondre de façon spécifique aux besoins de l'entreprise et apporter les bonnes réponses aux requêtes des décideurs. Aussi, la définition d'une méthode de conception d'une telle base est cruciale [Bonifati *et al*, 2002]. Or, les méthodes classiques proposées dans le cadre des systèmes transactionnels ne permettent ni de prendre en compte les spécificités des applications décisionnelles ni du modèle multidimensionnel [List *et al*, 2002]. Les méthodes de conception proposées dans le domaine de la modélisation multidimensionnelle, reposent sur un modèle multidimensionnel suivant un paradigme objet [Trujillo *et al*, 2002] [Moody et Kortink, 2000], relationnel [Kimball et

Ross, 2002] ou purement multidimensionnel [Golfarelli *et al*, 2002]. Une première catégorie de ces méthodes propose une démarche descendante qui part des besoins des décideurs pour construire le schéma multidimensionnel [Kimball et Ross, 2002] [Tsois *et al*, 2001]. Une deuxième catégorie propose une démarche ascendante qui part uniquement des données issues de sources, négligeant le rôle du décideur dans le processus de construction de la base décisionnelle [Golfarelli *et al*, 2002] [Moody et Kortink, 2000]. Une troisième catégorie, les méthodes mixtes, intègre les données des sources aux spécifications des besoins des décideurs lors de la modélisation de la base décisionnelle permettant ainsi d'exploiter les données existantes et de fournir une réponse succincte et de qualité aux requêtes des décideurs [Bonifati *et al*, 2002] [Carneiro et Brayner, 2002] [Lujan-Mora *et al*, 2004] [Trujillo *et al*, 2002].

Dans [Kimball et Ross, 2002] différentes études de cas de BDM sont proposées. Une méthode, appelée architecture en matrice de BUS, est proposée pour la construction d'un schéma multidimensionnel à partir de la définition des besoins des utilisateurs. Ces travaux ne proposent pas de méthode formelle de conception et de construction d'une base dimensionnelle. [Tsois *et al*, 2001] propose une méthode, permettant de construire le schéma multidimensionnel à partir des besoins décideurs, établie sur un modèle conceptuel des données multidimensionnelles, appelé MAC. La méthode proposée fournit un schéma conceptuel partiel qui ne décrit que les hiérarchies de dimensions. Ainsi, nous ne retrouvons pas un formalisme qui décrit le modèle MAC en totalité. [Prat et Akoka, 2002] propose une méthode de développement à partir d'une BDM basée sur les notations UML. Le concepteur part d'une spécification informelle des besoins des utilisateurs pour créer un diagramme de classes UML. Ce diagramme est enrichi afin d'obtenir un schéma pivot facilitant le passage des concepts UML vers les concepts dimensionnels. La transformation du modèle pivot en un modèle multidimensionnel est réalisée au niveau logique. [Golfarelli *et al*, 2002] définit un modèle de Fait Dimensionnel et décrit une méthode ascendante de conception d'une BDM en fonction de ce modèle. [Moody et Kortink, 2000] explore les techniques de dérivation de schémas multidimensionnels à partir d'un schéma Entité-Relation.

[Carneiro et Brayner, 2002] propose X_META, une méthode de développement des BDM suivant une démarche en spirale. Les auteurs insistent sur le besoin d'intégrer les décideurs dans le processus de développement du schéma multidimensionnel en se basant sur les données collectées à partir des sources opérationnelles. Cette méthode, traitant de la gestion du projet, reste théorique pour la plupart de ses phases. [Trujillo *et al*, 2002] propose une méthodologie basée sur l'approche orienté objet qui met l'accent sur l'évolution des besoins des utilisateurs dans des stades avancées de l'analyse des besoins.

Les méthodes de conception proposées s'appuient, pour la plupart, sur un modèle multidimensionnel logique ou conceptuel [Lujan-Mora *et al*, 2004]. Aucun de ces modèles n'exprime les contraintes entre les données multidimensionnelles. Or, ces contraintes permettent une définition des schémas multidimensionnels interdisant des incohérences lors des analyses [Hümmer *et al*, 2002] et une meilleure prise de décision en offrant une information valide aux décideurs.

Afin de palier ces insuffisances, nous proposons une méthode de conception des BDM intégrant l'expression des contraintes. Cette méthode est basée sur une approche mixte intégrant à la fois la description des données sources et la définition des besoins utilisateurs. Ce choix est motivé par le fait que ce type de méthode permet d'intégrer les données pertinentes pour la prise de décision extraites à la fois des sources ainsi qu'à partir des besoins des décideurs [Bonifati *et al*, 2002] [Trujillo *et al*, 2002].

Dans les sections suivantes, nous présentons tout d'abord les concepts de notre modèle multidimensionnel contraint puis les étapes de notre méthode de conception de schéma multidimensionnel. Dans cette méthode, nous suivons en parallèle une démarche descendante basée sur la spécification des besoins des décideurs et une démarche ascendante basée sur la description des données de la source. Les schémas multidimensionnels résultants de ces deux démarches sont confrontés dans une dernière étape pour produire un schéma conceptuel multidimensionnel contraint.

2. Concepts fondamentaux

2.1. Modèle multidimensionnel contraint

Cette section définit brièvement notre modèle multidimensionnel établi sur les concepts de constellation, de fait, de dimension multi-instanciable, de hiérarchies multiples et de contraintes sémantiques. Une plus ample description de notre modèle est présentée dans l'article [Ghozzi *et al*, 2004].

2.1.1. Constellation.

Le schéma en constellation est une généralisation du schéma en étoile [Kimball et Ross, 2002]. Une constellation regroupe plusieurs sujets d'analyse (faits) étudiés selon différents axes (dimensions) éventuellement partagés. Une constellation C est définie par le n -uplet $(N^C, F^C, D^C, Star^C, Cons^C)$ pour lequel N^C est le nom de la constellation, $F^C = \{F_1, F_2, \dots, F_p\}$ est un ensemble de faits, $D^C = \{D_1, D_2, \dots, D_q\}$ est un ensemble de dimensions, $Star^C : F^C \rightarrow 2^{D^C}$ est une fonction associant les faits aux dimensions afin de spécifier les sujets d'analyses et les axes d'étude associés et $Cons^C$ représente l'ensemble des contraintes associées à la constellation.

2.1.2. Fait.

Tout sujet d'analyse est représenté par un fait. Chaque fait est caractérisé par une ou plusieurs mesures représentant les indicateurs analysés. Un fait F est défini par le quadruplet $(N^F, M^F, I^F, IStar^F)$ où N^F est le nom du fait, $M^F = \{m_1, m_2, \dots, m_w\}$ est un ensemble de mesures (ou indicateurs), $I^F = \{I^F_1, I^F_2, \dots\}$ est l'ensemble des instances de F et $IStar^F$ est une fonction associant les instances de I^F aux instances des dimensions liées au fait.

2.1.3. Dimension

Les dimensions représentent les axes d'analyse. Une dimension est formée d'attributs exprimant les caractéristiques en fonction desquelles sont analysées les mesures d'activité. Les attributs d'une dimension peuvent être organisés en hiérarchies, de la granularité la plus fine à la plus générale. On pose IDE un ensemble d'identifiants. Une dimension D est définie par (N^D, P^D, H^D, I^D) où N^D est le nom de la dimension, $A^D = \{a_1, a_2, \dots, a_u\}$ est un ensemble d'attributs, $H^D = \{h^D_1, h^D_2, \dots, h^D_v\}$ est un ensemble de hiérarchies et $I^D = \{I^D_1, I^D_2, \dots\}$ est l'ensemble des instances de D . Toute dimension possède, en plus des attributs définis par l'utilisateur, les attributs All et Id tels que $dom(All) = \{all\}$ et $dom(Id) \in IDE$; All désigne la granularité de plus haut niveau tandis que Id représente la granularité la plus fine.

2.1.4. Hiérarchie.

Les hiérarchies permettent de déterminer les niveaux de granularité auxquels peuvent être manipulées les mesures des faits. Chaque hiérarchie propose un ordonnancement des

attributs d'une dimension afin de spécifier ces différents niveaux de granularité. Une hiérarchie h^{Di} est un chemin élémentaire acyclique débutant par Id et se terminant par All. Elle est définie par $(N^{hDi}, Param^{hDi}, Suppl^{hDi}, Cond^{hDi})$ où N^{hDi} est le nom de la hiérarchie, $Param^{hDi} : P \rightarrow P$ ($P \subseteq A^D$) est une fonction décrivant la hiérarchie des attributs (chaque attribut est appelé paramètre de la hiérarchie), $Suppl^{hDi} : P \rightarrow 2^{(AD-P)}$ est une fonction spécifiant les attributs faibles qui complètent la sémantique des paramètres (chaque paramètre est associé à un ensemble d'attributs faibles) et $Cond^{hDi}$ est une expression booléenne définissant la condition d'appartenance des instances de la dimension à une hiérarchie.

Nous notons $I_k^D \in_{(cond)} h_i^D$ pour indiquer que l'instance k de I^D satisfait la condition $Cond^h$ et par conséquent I_k^D appartient à la hiérarchie h_i^D . Cette caractéristique confère aux dimensions de notre modèle leur propriété de multi-instanciation.

2.1.5. Contraintes

Les contraintes sont des expressions qui précisent le rôle ou la portée d'un élément de modélisation (elles permettent d'étendre ou de préciser sa sémantique). [Hümmer *et al.*, 2002] dresse une liste de dix problématiques dans le domaine de modélisation dimensionnelle. Plusieurs de ces problématiques dénotent le besoin d'un mécanisme de contraintes dans le modèle multidimensionnel. L'expression de ces contraintes permet, d'une part, de valider la structure dimensionnelle (structure hiérarchique des axes d'analyse) et d'autre part de vérifier la sémantique des données analysées et notamment de désambiguïser les valeurs vides (ou nulles) qui peuvent provenir de la combinaison de perspectives d'analyses incompatibles [Carpani et Ruggia, 2001] [Hurtado et Mendelzon, 2002]. Ainsi, nous proposons un ensemble de contraintes associées aux concepts et aux données de notre modèle afin de donner l'interprétation la plus précise possible de la réalité pour une meilleure prise de décision [Carpani et Ruggia, 2001] [Hümmer *et al.*, 2002]. Nous définissons des contraintes à deux niveaux. Les contraintes intra-dimensions sont des contraintes concernant une seule dimension, autrement dit, il s'agit de contraintes entre les hiérarchies d'une même dimension et les contraintes inter-dimensions sont des contraintes portant sur plusieurs dimensions.

Contraintes intra-dimension. Nous définissons cinq types de contraintes entre les hiérarchies d'une même dimension : l'exclusion \otimes , l'inclusion \odot , la simultanéité \ominus , la totalité Θ et la partition \oplus . Nous rappelons dans ce qui suit la définition de la contrainte d'exclusion, les autres contraintes sont présentées dans des travaux antérieurs [Ghozzi *et al.*, 2004].

L'exclusion entre deux hiérarchies h_1 et h_2 de D , notée \otimes , traduit qu'une instance de D appartenant à h_1 n'appartient pas à h_2 (et réciproquement) :

$$h_1 \otimes h_2 \text{ ssi } \forall I_{k1}^D \in_{(cond)} h_1 \wedge \forall I_{k2}^D \in_{(cond)} h_2 \Rightarrow I_{k1}^D \neq I_{k2}^D$$

Contraintes inter-dimensions. De même que pour les contraintes intra-dimension, nous définissons cinq types de contraintes entre les hiérarchies de différentes dimensions. Ces contraintes sont relatives aux instances du fait en fonction des instances des hiérarchies associées. Nous distinguons les contraintes d'exclusion, d'inclusion, de simultanéité, de totalité et de partition. Nous rappelons la définition de la contrainte d'exclusion, une description exhaustive de ces contraintes est présentée dans nos travaux précédents [Ghozzi *et al.*, 2004].

L'exclusion entre deux hiérarchies h_1 d'une dimension D_1 et h_2 d'une dimension D_2 indique qu'une instance du fait (relié à D_1 et à D_2) ne peut être associée simultanément aux instances de h_1 et de h_2 :

$$h_1 \otimes h_2 \text{ ssi } (\forall I_j^F \in I^F \mid \exists I_{k1}^{D1} \in (cond) h_1 \wedge I_{k1}^{D1} \in IStar^F(I_j^F)) \Rightarrow \neg (\exists I_{k2}^{D2} \in (cond) h_2 \mid I_{k2}^{D2} \in IStar^F(I_j^F)).$$

2.2. Méthode de conception de base multidimensionnelle

Notre objectif est de spécifier les schémas de BDM intégrant toutes les données d'aide à la prise de décision qui peuvent être extraites à partir de la source ou bien à partir des besoins des décideurs [Bonifati *et al*, 2002] [Lujan-Mora *et al*, 2004] [Trujillo *et al*, 2002]. Pour répondre à ce besoin, nous avons suivi une approche mixte intégrant la description des besoins des décideurs à l'aide d'une démarche descendante et la description des données sources à l'aide d'une démarche ascendante [Bonifati *et al*, 2002]. Le résultat de la phase conceptuelle est un schéma multidimensionnel contraint obtenu après confrontation des schémas multidimensionnels résultant des deux démarches.

Néanmoins, la mise au point de ces deux démarches nécessite la définition en amont du cadre général de l'application décisionnelle. En effet, nous avons été confrontés au problème de divergence des résultats des deux démarches dans un contexte de système décisionnel multi-domaines. Pour éviter cette divergence, nous avons proposé une étape de définition du **domaine de l'analyse** en amont des deux démarches descendante et ascendante. Lors de cette étape, nous proposons de fixer le champ de l'analyse en choisissant de travailler sur un métier ou une analyse spécifique de l'entreprise tels que le commercial, le marketing, la gestion des stocks, etc. La définition de ce cadre d'analyse permet au concepteur de choisir le groupe des décideurs à interviewer lors de la démarche descendante. Au niveau de la démarche ascendante, la spécification d'un domaine d'analyse évite au concepteur d'étudier un ensemble trop important et inutile de données sources.

3. Démarche descendante

L'objectif de cette démarche est de concevoir un schéma multidimensionnel à partir des besoins des décideurs et des règles de gestion relatives aux données décisionnelles. Elle comporte trois étapes ; (1) la collecte des données, (2) la spécification des besoins et (3) la formalisation des besoins (cf. FIG 1).

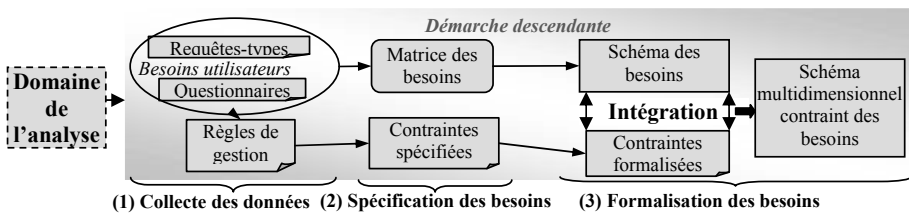


FIG 1. Démarche descendante

3.1. Collecte des données

Dans cette étape, nous déterminons les besoins décisionnels initiaux en définissant les types d'informations susceptibles d'intéresser chaque groupe de décideurs. On procède alors à la collecte des **requêtes-types** pertinentes en interviewant les décideurs, à la mise au point d'un **questionnaire** permettant de mieux caractériser leurs besoins et à l'analyse des données décisionnelles et des résultats des interviews et des questionnaires afin de dégager les **règles de gestion** appliquées à ces données.

3.1.1. Requêtes-types

Les requêtes-types peuvent être collectées à partir des anciens rapports d'analyse ou en interrogeant les décideurs. Nous proposons un pseudo-langage de structuration des requêtes qui permet de faciliter la définition des besoins. Ce pseudo-langage est établi sur trois clauses qui répondent à différents types de questions. La clause **Analyser** répond à la question **Quoi ?** Elle définit les données à analyser. La clause **En fonction** répond aux questions **Qui ? Où ? Quand ?** Elle indique les paramètres de l'analyse. Enfin, la clause **Pour** répond à la question **Pour Qui ou Quelles données ?** Elle comporte des restrictions sur les données à analyser en fixant par exemple la valeur d'un paramètre d'analyse.

Exemple 1 Nous souhaitons spécifier une BDM permettant d'analyser les locations de véhicules. L'analyse des rapports d'activité existants et une première interview des décideurs a permis de spécifier six requêtes-types (**R1** à **R6**). La FIG 2 présente un exemple de rapports d'activité que nous avons collecté pendant l'analyse. Le premier rapport décrit les montants des locations réalisées par mois et par agence. Le deuxième présente les montants des locations par ville. A partir de ces rapports, nous avons extrait les requêtes-types **R2** et **R3**.

Listing des montants des locations par Mois et par Agence :			<p>R1 : Analyser le montant des locations En fonction des mois et des véhicules Pour les véhicules de type sport.</p> <p>R2 : Analyser le montant des locations En fonction des mois et des agences</p> <p>R3 : Analyser le montant des locations En fonction des villes</p> <p>R4 : Analyser le chiffre d'affaire des employés En fonction des mois Pour l'employé Paul et l'année 2002.</p> <p>R5 : Analyser les marques des véhicules En fonction de la durée de location Pour l'état de Floride.</p> <p>R6 : Analyser les montants des locations En fonction des véhicules Pour la marque Peugeot.</p>
Mois Location	Agence	Montant Location	
Janvier-00	Ag_tlse	90	
Janvier-00	Ag_bord	120	
Juin-00	Ag_tlse	70	
juin-00	Ag_Dallas	110	
Juin-00	Ag_Gren	200	
Janvier-01	Ag_tlse	250	
...	
Listing des montants des locations par Ville :			
Ville	Montant Location		
Toulouse	410		
Bordeaux	120		
Grenoble	200		
Dallas	110		

FIG 2. Un exemple de rapport d'activité

3.1.2. Questionnaires

Le questionnaire est un outil de collecte de données dans lequel le décideur est guidé à travers d'un ensemble de questions dans le but d'obtenir des informations précises, nécessaires à la modélisation multidimensionnelle. Ce questionnaire est établi en fonction des résultats de la première étape, les requêtes-types, afin de compléter la description des besoins des décideurs. Ainsi, en remarquant que les véhicules ont une classification prédéfinie, le concepteur pose la question : quel est la classification des véhicules de location ?

Ce questionnaire doit permettre de modifier et éventuellement, d'ajouter de nouveaux éléments dans les requêtes-types. Au niveau des axes de l'analyse, par exemple, il permet de modifier ou d'ajouter de nouveaux axes, de vérifier s'il n'y a pas de nouvelles perspectives analytiques (hiérarchies). Au niveau de celles-ci, le questionnaire doit permettre d'organiser les niveaux de granularité et de compléter leur sémantique par des descripteurs (par exemple, ajouter un nom d'agence si son code n'est pas assez explicite pour le décideur).

3.1.3. Règles de gestion

Les règles de gestion régissent le système d'information. Ces règles sont souvent représentées sous forme de contraintes qui doivent être respectées pour permettre le bon fonctionnement du système opérationnel. Cette étape est réalisée en parallèle avec les deux étapes précédentes : les requêtes et les questionnaires. Dans un premier temps, nous dégageons les règles de gestion relatives aux données décisionnelles. Ces règles sont spécifiées dans la documentation et les rapports d'activité du système opérationnel. Parmi les règles de gestion dans un système opérationnel, nous retrouvons l'organisation des produits dans des rayons par catégorie ou la classification des tarifs des véhicules par type et par niveaux de confort. L'interview des décideurs pour la définition des requêtes peut révéler d'autres règles de gestion notamment celles relatives au bon fonctionnement du système opérationnel. Dans un second temps, nous validons et complétons la liste de ces règles en proposant dans le questionnaire des questions sur les contraintes relatives à l'organisation des entités du système opérationnel.

Exemple 2 Dans notre exemple de société de location, nous avons spécifié une première règle relative à la classification des véhicules dans les agences de location. En effet, l'analyse de la documentation de la société nous a permis de détecter que les agences françaises utilisent une classification de véhicules différente des agences américaines. La validation des règles définies dans notre exemple nous a amené à ajouter une nouvelle règle relative à la localisation des agences : les villes françaises sont regroupées par région tandis que les villes américaines sont regroupées par état.

3.2. Spécification des besoins

En sortie de l'étape précédente, nous obtenons une liste de requêtes-types formulées dans notre pseudo-langage et un ensemble de règles de gestion. L'étape de spécification permet d'analyser les données collectées afin de spécifier les besoins des décideurs. Ces besoins seront organisés en termes de paramètres et de mesures afin de préparer la définition du schéma multidimensionnel.

3.2.1. Matrice des besoins

A partir de la formulation simplifiée des requêtes décrivant les besoins des décideurs, nous définissons la matrice des besoins en trois étapes. Une étape de **construction** d'une matrice carrée en fonction des propriétés de l'analyse extraites à partir des requêtes-types (voir Tab 1). Une étape de **remplissage** de la matrice afin de caractériser les propriétés reliées par une relation d'analyse en se basant sur les requêtes-types ; la clause **Analyser** permet d'indiquer les propriétés analysées alors que les clauses **En fonction** et **Pour**, permettent d'extraire les propriétés qui paramètrent l'analyse. Dans la matrice, chaque case cochée (√) indique que la propriété en ligne est analysée en fonction de celle en colonne. L'étape de **simplification** de la matrice permet de supprimer les lignes et les colonnes vides. Ainsi, la matrice simplifiée contient en lignes les mesures et en colonnes les paramètres d'analyse.

Après cette simplification, nous pouvons retrouver des propriétés qui sont à la fois des indicateurs et des paramètres (exemple : Nb jours et Marque). Pour traiter ce cas particulier, nous procédons à une nouvelle lecture de la matrice permettant de montrer si ces propriétés sont plus utilisées en tant qu'indicateur ou en tant que paramètre. Nous notons que le type de la propriété peut déterminer sa nature ; souvent les indicateurs sont des propriétés numériques alors que les paramètres sont des descripteurs de type textuel [Kimball et Ross, 2002].

Exemple 3 A partir de la requête *R1*, nous déduisons que la propriété « Mt-locations » (dans la clause *Analyser*) est analysée en fonction des propriétés mois, immatriculation et type (dans les clauses *En fonction* et *Pour*). Nous cochons les cases correspondantes à l'intersection de la ligne montant des locations et des colonnes mois, immatriculation et type de véhicule. L'analyse de toutes les requêtes permet de construire la matrice des besoins suivante.

Matrice des besoins

	<i>Mt locations</i>	<i>Ville</i>	<i>Région</i>	<i>Etat</i>	<i>Année</i>	<i>Mois</i>	<i>Nb jours</i>	<i>CA</i>	<i>Id_Emp</i>	<i>Id_Client</i>	<i>Immat</i>	<i>Type</i>	<i>Marque</i>
<i>Mt locations</i>		√	√	√	√	√				√	√	√	√
<i>Ville</i>													
<i>Région</i>													
<i>Etat</i>													
<i>Année</i>													
<i>Mois</i>													
<i>Nb jours</i>		√	√	√	√	√				√	√	√	√
<i>CA</i>					√	√			√				
<i>Id_Emp</i>													
<i>Id_Client</i>													
<i>Immat</i>													
<i>Type</i>													
<i>Marque</i>				√			√						

TAB 1. Matrice carrée des propriétés de notre exemple

Dans notre exemple, le processus de simplification engendre la matrice des besoins décrite au tableau Tab 2.

Matrice des besoins

<i>Paramètres / Indicateur</i>	<i>Ville</i>	<i>Région</i>	<i>Etat</i>	<i>Année</i>	<i>Mois</i>	<i>Id_Emp</i>	<i>Id_Client</i>	<i>Immat</i>	<i>Type</i>	<i>Marque</i>
<i>Mt locations</i>	√	√	√	√	√		√	√	√	√
<i>Nb jours</i>	√	√	√	√	√		√	√	√	√
<i>CA</i>				√	√	√				

TAB 2. Matrice simplifiée des besoins.

3.2.2. Contraintes spécifiées.

Cette étape permet d'énumérer les contraintes spécifiées à partir des règles de gestion définies dans l'étape de collecte de besoins. Cette énumération est faite en langage naturel. En effet, une règle de gestion peut donner lieu à plusieurs contraintes appliquées à différentes données multidimensionnelles. Cette énumération permet de préparer l'étape suivante de formalisation.

Exemple 4 Dans notre application de location de véhicule et à partir des règles de gestion définies dans l'exemple 2, nous avons spécifié les contraintes suivantes :

- L'organisation géographique décrivant la localisation des agences française est différente de l'organisation américaine.
- La classification des véhicules appliquée en France est différente de celle appliquée aux Etats-Unis.
- La classification française des véhicules appliquée par les agences Françaises n'est pas utilisée par les agences Américaines et inversement.

3.3. Formalisation des besoins

Après avoir collecté et spécifié les besoins des décideurs, nous réalisons dans cette étape la formalisation de ces besoins sous forme d'un schéma multidimensionnel contraint. La conception de ce schéma est basée sur la matrice des besoins définie dans l'étape précédente. Les contraintes sémantiques sont formalisées puis intégrées dans le schéma multidimensionnel des besoins.

3.3.1. Transformation de la matrice des besoins

A ce niveau, nous proposons de formaliser les besoins spécifiés dans les étapes précédentes sous forme de schéma multidimensionnel. Nous considérons que ce schéma est le plus adapté à représenter les besoins des décideurs. En outre, cette représentation permet de faciliter l'étape de confrontation entre le résultat de la démarche descendante et celui de la démarche ascendante. Pour obtenir ce schéma nous proposons les étapes suivantes :

Définition des faits. Durant cette étape nous regroupons les mesures dans des faits (sujets d'analyse). La définition des faits peut être réalisée d'une manière automatique en regroupant les mesures analysées au travers de paramètres identiques. Au niveau de cette étape, nous définissons aussi les fonctions d'agrégation compatibles avec chaque mesure en se basant sur les questionnaires.

Exemple 5 Pour notre exemple, nous regroupons les mesures « montant » et « nombre de jours » des locations dans le fait *Location* et nous définissons le fait *Performance* pour la mesure « CA ». Pour chacune de ces mesures, nous définissons les fonctions d'agrégation nécessaires à l'analyse.

- Location ((Montant_Loc, {sum, avg}), (Nb_jours, {sum, avg}))
- Performance ((CA, {sum, avg, max, min}))

Définition des dimensions. Cette étape réalise le **regroupement** automatique des paramètres dans des dimensions (axes d'analyse), l'**enrichissement** de ces dimensions à travers de nouvelles propriétés tels que les attributs faibles qui seront greffés aux différentes dimensions de l'analyse et la définition de la **granularité** de l'analyse en choisissant le niveau d'analyse le plus fin pour chaque dimension..

Exemple 6 Dans notre exemple, le regroupement automatique des paramètres engendre les dimensions (*D1, D2, D3*) (cf. Tab 3). L'analyse de ces dimensions permet de diviser la dimension D1 en trois dimensions *Agences, Clients* et *Véhicules*. L'affectation des noms aux dimensions est réalisée par le concepteur en fonction de ces informations. Au niveau de l'étape de la collecte des données, nous avons enrichi les dimensions de l'analyse par les informations collectées dans les questionnaires. Ainsi, un client est caractérisé par son nom, son prénom et sa ville. Les agences sont identifiées par un code (*Code_Ag*) et caractérisées par leur nom et par leur localisation (*Ville, Région, Etat, Pays*). Les employés sont caractérisés par leur nom, leur prénom, leur âge et la tranche d'âge à laquelle ils appartiennent.

L'étude de la granularité nous a permis de détecter que l'analyse des locations est réalisée par agence (*Code_Ag*), jour (*IdT*), client (*Id_Client*) et véhicule (*Immat*), l'analyse des chiffres d'affaire des employés est effectuée par employé (*Id_employé*) et par jour (*IdT*). Ces paramètres représentent la granularité la plus fine de l'analyse. Les dimensions obtenues après chaque étape sont présentées dans le tableau suivant.

D1 (Région, Etat, Id_Client, Immat Véhicule, Type, Marque)	Agences (Ville, Région, Etat) Clients (Id_Client) Véhicules (Immat véhicule, Type, Marque)	Agences (Code_Ag, Nom, Ville, Département, Région, Etat, Pays) Clients (Id_Client, Nom, Prénom, Ville) Véhicules (Immat, véhicule, Type, Marque)
D2 (Id_Employé)	Employés (Id_Employé)	Employés (Id_Employé, Nom, Prénom, Tranche)
D3 (Mois, Année)	Temps (Mois, Année).	Temps (IdT, jour, Mois, Année).
a- Regroupement automatique	b- Enrichissement	c- granularité de l'analyse

TAB 3. Transformation des dimensions

Définition des hiérarchies. Cette étape consiste à organiser les paramètres de chaque dimension dans des hiérarchies. Cette opération est réalisée en se basant sur le questionnaire et les règles de gestion définies dans l'étape de collecte de données. Pour la dimension *Agences*, par exemple, nous définissons deux hiérarchies : une hiérarchie relative à la localisation des agences françaises et une deuxième relative à la localisation des agences américaines.

Définition du schéma multidimensionnel. Nous réalisons l'affectation des dimensions aux faits d'une manière automatique à partir de la matrice des besoins. Chaque dimension comportant des paramètres qui ont des intersections avec un indicateur d'un fait, est affectée à ce fait.

Exemple 7 Dans notre exemple, les locations sont analysées en fonction des dimensions *Agences, Temps, Clients* et *Véhicules*. Le fait Performance est analysé en fonction des dimensions *Employés* et *Temps*. Le résultat de cette étape de formalisation est présenté dans le schéma en constellation (FIG 3).

3.3.2. Formalisation des contraintes

Cette étape permet de formaliser les contraintes extraites à partir des règles de gestion définies dans l'étape de collecte des besoins. En effet, une règle de gestion peut donner lieu à plusieurs contraintes appliquées à différentes données multidimensionnelles. Cette

spécification permet de préparer l'étape suivante d'intégration des contraintes dans le schéma multidimensionnel.

Afin d'exprimer les contraintes relatives au contexte multidimensionnel, nous proposons un langage qui s'inspire du langage des contraintes objet OCL¹. Nous appelons ce langage LCD pour Langage de Contraintes multiDimensionnelles. Ce langage permet de spécifier les contraintes sous un langage simple et facile à décrire par le concepteur. Pour la spécification des contraintes, nous reprenons un concept de base du langage OCL, à savoir le concept d'invariant. Dans notre contexte, les invariants sont définis sur les concepts multidimensionnels, à savoir les faits, les dimensions et les hiérarchies. Ces concepts représentent dans notre langage le contexte d'une contrainte énoncé par le mot **context**. Ainsi, nous pouvons définir, par exemple, la dimension Agences en tant que contexte d'une contrainte invariante énoncée par le mot clé **inv**.

Exemple 8 Afin de spécifier les contraintes définies dans l'étape précédente, nous devons les analyser sémantiquement afin de pouvoir les exprimer dans notre langage LCD. Ces contraintes sont souvent basées sur les domaines des paramètres et des mesures de l'analyse. Une solution possible pour la détermination de ces contraintes est l'analyse des relations entre les domaines des paramètres et des mesures. Nous avons adopté cette solution pour la traduction en langage LCD des contraintes définies dans notre application de location de véhicules

1. L'organisation géographique française est différente de l'organisation américaine. Cette contrainte est appliquée sur les domaines des paramètres de l'axe d'analyse *Agences*. Une analyse de ces domaines permet de constater que les agences qui sont localisées dans un état américain ont des valeurs nulles pour le paramètre *Région*. Inversement, les agences françaises dont le paramètre *Région* est renseigné ont un paramètre *Etat* vide.
2. Les agences françaises utilisent la nomenclature des véhicules française et les agences américaines appliquent la nomenclature américaine.

<p>Context <i>a</i> : <i>Agences</i> Inv : IF <i>a.Etat</i> -> <i>notEmpty()</i> then <i>a.Région</i> -> <i>IsEmpty()</i> Else <i>a.Région</i> -> <i>notEmpty()</i> EndIF</p>	<p>Context <i>l</i>.<i>Location inv</i> Inv : IF <i>l.Agences.Etat</i> -> <i>notEmpty()</i> then <i>l.Véhicules.Marque</i> -> <i>IsEmpty()</i> Else <i>l.Véhicules.Type</i> -> <i>IsEmpty()</i> EndIF</p>
---	---

3.3.3. Intégration des contraintes

Dans cette étape nous réalisons l'intégration des contraintes dans le schéma des besoins. Ainsi, nous transformons les contraintes exprimées dans le langage LCD en contraintes intra et inter-dimensions.

Exemple 9 La conception du schéma multidimensionnel avec les différentes dimensions et hiérarchies, nous permet de définir les contraintes sémantiques. La première contrainte définie dans notre application relative à l'organisation géographique française et à l'organisation américaine, est exprimée par une contrainte d'exclusion intra-dimension entre les hiérarchies "*geo_fr*" et "*geo_us*" de la dimension *Agences*. La deuxième contrainte caractérisant le fait que les agences françaises utilisent la nomenclature des véhicules

¹ <http://www.omg.org/docs/formal/03-03-13.pdf>

française et les agences américaines appliquent la nomenclature américaine, est exprimé par une contrainte d'exclusion intra-dimension entre les hiérarchies "clas_fr" et "clas_us" de la dimension *Véhicules*. Nous obtenons à la fin de cette étape le schéma des besoins suivant.

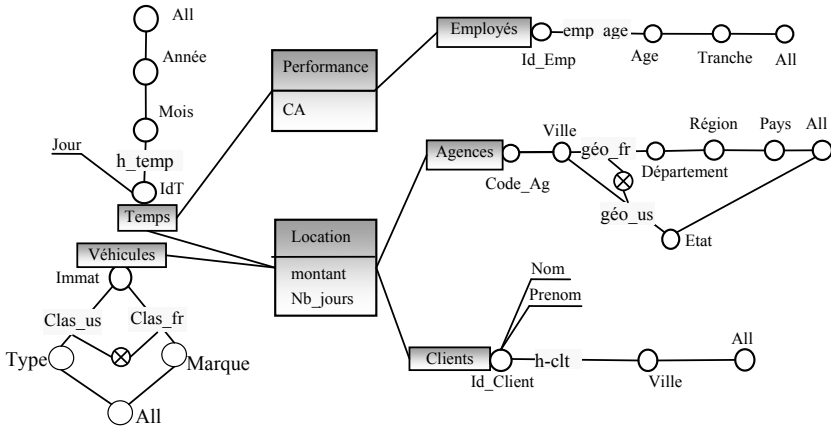


FIG 3. Représentation graphique du schéma des besoins

La démarche *descendante* réalise progressivement la transformation des besoins des décideurs exprimés dans un langage naturel en un schéma multidimensionnel structuré intégrant l'expression des contraintes. Cette transformation progressive permet au concepteur d'exprimer les besoins décisionnels de façon exacte et complète. Cette démarche descendante constitue une première étape dans la conception du schéma multidimensionnel au niveau conceptuel. Basée uniquement sur les besoins des décideurs, elle ne tient pas compte des données sources stockées dans la source. Nous proposons dans la section suivante, une démarche ascendante pour construire un schéma multidimensionnel à partir du schéma de la source.

4. Démarche ascendante

Afin de tenir compte des informations contenues dans la source de données, nous proposons une démarche ascendante incrémentale. Cette démarche part du schéma conceptuel de la source de données pour construire le schéma multidimensionnel contraint d'un magasin de données. Dans cette démarche, nous supposons que le concepteur du schéma multidimensionnel a une double compétence informatique et métier. En effet, c'est le concepteur du schéma multidimensionnel qui doit détecter les différents centres d'intérêt de l'organisation en analysant le schéma de la source. En outre, l'étape de la définition du domaine d'analyse permet de fournir au concepteur une vision générale des besoins décisionnels de l'entreprise. Dans notre méthode, nous considérons que la démarche ascendante est réalisée séparément mais en parallèle avec la démarche descendante. Toutefois, le concepteur peut, au niveau de la phase ascendante, tenir compte des informations collectées lors de la démarche descendante. Nous décrivons dans les sections suivantes les étapes de notre démarche ascendante.

4.1. Détermination des faits

La détermination des sujets de l’analyse permet de dresser la liste des faits du modèle multidimensionnel. Selon notre démarche, un fait est projeté à partir d’une Classe Représentative (CR) de la source. Une Classe Représentative (CR) enregistre les valeurs numériques et les événements particuliers comme le montant en devise, le poids, les volumes. Ce ces valeurs et ces «événements que les décideurs analysent Cette classe est choisie par le concepteur de la BDM. Celui-ci analyse la source et définit les classes susceptibles d’être transformées en sujets d’analyse. Le concepteur choisit comme classes représentatives les classes qui comportent des propriétés numériques transformables en indicateurs d’analyse. Les mesures du fait sont obtenues en appliquant une fonction de calcul sur un ou plusieurs attributs de la classe représentative. Cette fonction peut être une simple fonction de projection (fonction identité) ou une expression arithmétique (montant = prix unitaire * quantité).

Afin de permettre l’agrégation de ce fait, nous définissons pour chaque mesure l’ensemble des fonctions d’agrégation sémantiquement compatibles. Ces fonctions seront utilisées lors de la manipulation des données multidimensionnelles en agrégeant les mesures à différents niveaux hiérarchiques.

Exemple 10 Notre source de données représente une application de gestion de location de véhicules d’une société qui possède plusieurs agences en France et aux États-unis. Une location est réalisée selon plusieurs types (forfait journalier, heure, kilométrage, ...). Elle concerne un seul client et un seul véhicule. Un véhicule est caractérisé par sa catégorie et son modèle. Chaque type de location possède un tarif de location qui dépend de la catégorie et du modèle du véhicule. Cette source appelée Location est présentée par le diagramme de classes UML de la FIG 4.

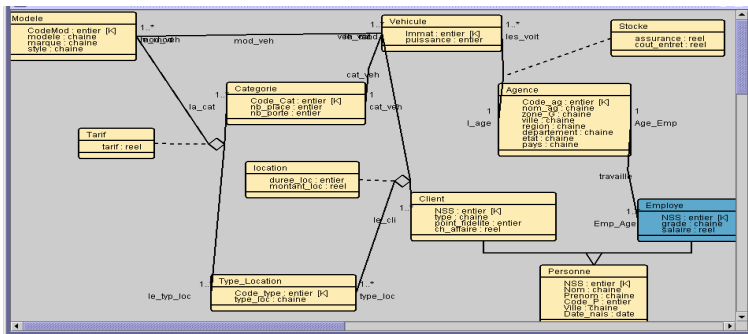


FIG 4. Diagramme de classes UML du schéma de la source de données

En se basant sur le schéma de la source et sur le domaine d’analyse de l’activité commerciale des agences de location, le concepteur souhaite dégager les sujets d’analyse susceptibles d’intéresser les décideurs. Dans le domaine commercial, l’analyse du schéma permet de détecter une classe représentative qui comporte des indicateurs d’analyse ; la classe *Location* comportant les montants des locations des agences. Nous souhaitons, maintenant, définir un fait analysant les montants et les durées des locations. Ce fait répond aux besoins décisionnels du domaine d’analyse défini. Pour répondre à ce besoin, nous définissons le fait ‘*Location_Veh*’ à partir de la classe représentative *Location*. Ce fait est

défini par le couple (N^F, M^F) représentant le nom du fait et les mesures accompagnées de leurs fonctions d'agrégation.

- NLocation_Veh = "Location_Veh",
- MLocation_Veh = {(montant = SD.Location.Montant_Loc, {sum, avg}), (nbjours = SD.Location.durée_Loc, {sum, avg})} avec SD le nom de la source de notre exemple.

4.2. Détermination des dimensions

Les dimensions sont élaborées à partir des classes sources reliées à la classe représentative, appelées classes **déterminantes**. La détermination de ces classes est réalisée automatiquement en suivant le principe de dépendance fonctionnelle entre classes. Une classe Ci est déterminante d'une classe CR noté Ci => CR si (i) Ci = CR, (ii) Ci hérite de CR, (iii) Ci représente une classe d'association et CR entre dans la formation de l'association, (iv) Ci est reliée à CR par une relation d'association monovaluée (X, 1) ou (v) Ci est reliée à CR par une relation de composition de type (1, x) si Ci est composée de CR ou (x, 1) si Ci compose CR.

La contrainte sur les cardinalités des relations d'association et de composition permet de garantir l'unicité entre les valeurs d'une classe C1 et les valeurs liées d'une classe C2. Cette propriété est essentielle, car elle permet de relier dans le magasin les mesures issues de la classe représentative aux paramètres issus des classes déterminantes. On peut déterminer l'ensemble **Determin (CR) = {CD1, CD2,..., CDm}** des classes déterminantes d'une classe représentative CR afin de spécifier l'ensemble des classes sources à partir desquelles peuvent être créées les dimensions.

Exemple 11 L'application du principe de dépendance sur l'exemple de notre source de données nous amène à déterminer l'ensemble des classes déterminantes de la classe représentative Location (classes grisées dans la figure FIG 5)

Détermin (Location) = {Location, Véhicule, Agence, Client, Personne, Type_Location, Catégorie, Modèle}

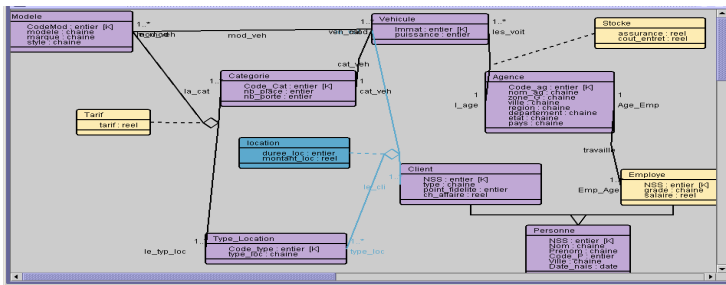


FIG 5. Détection automatique des classes déterminantes de la classe représentative Location

Les dimensions du fait Location_Veh sont définies à partir de ces classes déterminantes. En se basant sur le domaine de l'analyse, le concepteur définit toutes les dimensions susceptibles d'intéresser les décideurs. Dans le domaine commercial, le concepteur choisit de définir les dimensions Agences, Véhicules, Clients respectivement à partir des classes sources Agence, Véhicule et Client.

4.3. Définition de la granularité de l'analyse

La quatrième étape consiste à spécifier le niveau le plus détaillé suivant lequel les données multidimensionnelles sont analysées. Par exemple, pour une dimension géographique, les mesures peuvent être définies au niveau Département ou bien Ville. Au niveau de la démarche ascendante, nous choisissons la granularité d'analyse la plus détaillée possible au niveau de chaque dimension en se basant sur le schéma source. La détermination de la granularité de l'analyse permet de définir les paramètres les plus détaillés de chaque dimension. Chacun de ces paramètres représente le début d'une structure hiérarchique que nous définissons dans l'étape suivante.

4.4. Hiérarchisation des dimensions

Les paramètres des dimensions sont organisés en une structure hiérarchique qui permet d'analyser les mesures à différents niveaux de détail. La définition d'une hiérarchie est réalisée par la détection des **dépendances hiérarchiques** entre les paramètres d'une même dimension. Une dépendance hiérarchique entre deux paramètres p_i et p_j implique que la valeur de p_i détermine d'une manière unique la valeur de p_j (dépendance fonctionnelle) et que p_j ne dépend pas de p_i .

En se basant sur ce principe, nous pouvons définir différentes hiérarchies dans la même dimension. Au niveau de cette démarche, nous proposons de définir les hiérarchies les plus complètes que possible en fonction du schéma de la source.

Exemple 12

L'analyse des dépendances fonctionnelles entre les paramètres de la dimension *Agences* nous a permis de définir les trois hiérarchies suivantes :

- *geo_fr* = ('géo. française', {Param^{geo_fr}(CodeAg) = Ville, Param^{geo_fr}(Ville) = Département, Param^{geo_fr}(Département) = Région, Param^{geo_fr}(Région) = Pays, Param^{geo_fr}(Pays) = All}, {Suppl^{geo_fr}(CodeAg) = {Raison}, Suppl^{geo_fr}(Département) = {Nom_dpt}}, Pays = "France"),
- *geo_us* = ('géo. américaine', {Param^{geo_us}(CodeAg) = Ville, Param^{geo_us}(Ville) = Etat, Param^{geo_us}(Etat) = Pays, Param^{geo_us}(Pays) = All}, {Suppl^{geo_us}(CodeAg) = {Raison}}, Pays = "Etats-Unis" \wedge Etat \neq NULL),
- *geo_zn* = ('zone .agence', {Param^{geo_zn}(CodeAg) = Ville, Param^{geo_zn}(Ville) = Zone, Param^{geo_zn}(Zone) = Pays, Param^{geo_zn}(Pays) = All}, {Suppl^{geo_zn}(CodeAg) = {Raison}}, Zone \neq NULL).

4.5. Expression des contraintes

Nous rappelons les différentes contraintes qui peuvent être définies au niveau d'une dimension : exclusion \otimes , inclusion \odot , simultanéité \ominus , totalité \oplus et partition \oplus . La détection de ces contraintes est basée sur l'analyse des données de la source. Cette analyse au niveau de la classe Agence, nous a permis de constater que, pour toutes les agences françaises, l'attribut Etat est non renseigné et que, pour toutes les agences américaines, les attributs Département et Région sont non renseignés. Ce fait est exprimé au niveau de notre schéma conceptuel à l'aide de la condition d'appartenance aux hiérarchies. Dans un second temps, nous constatons que toutes les agences de la dimension font partie des agences françaises ou bien des agences américaines.

De la même manière que pour les contraintes intra-dimension, les contraintes inter-dimensions sont définies suite à l'analyse des instances des classes représentatives et des classes déterminantes de la source. Par exemple, l'analyse des données de la classe représentative *Location* nous permet de constater que les instances de cette classe, reliées aux instances des agences françaises, ne sont pas associées aux instances de la classe véhicule organisées suivant une classification américaine.

Exemple 13 En se basant sur l'analyse des données de la classe *Agence* de la source, nous avons défini les contraintes d'exclusion entre "geo_fr" et "geo_us" pour exprimer le fait que les agences organisées suivant l'une des deux hiérarchies ne peuvent pas appartenir à la deuxième hiérarchie. Par contre, nous remarquons que les agences situées dans les différents états (suivant la hiérarchie "geo_us") peuvent être organisées par Zone selon la hiérarchie "geo_zn". De même pour les agences décrites par la hiérarchie "geo_fr". Ainsi, nous avons défini une contrainte d'inclusion de la hiérarchie "geo_us" dans "geo_zn" et de même pour "geo_fr". Les hiérarchies "geo_fr" et "geo_us" de la dimension Agence contiennent la totalité des agences de la dimension. Ceci est exprimé par une contrainte de totalité entre les deux hiérarchies. Ces contraintes sont définies comme suit :

C1::geo_fr ⊗ geo_us, C3::geo_fr ⊗ geo_zn, C2::geo_fr ⊖ geo_us, C4::geo_us ⊗ geo_zn.

Au niveau inter-dimensions, une contrainte d'exclusion peut être définie entre la hiérarchie "geo_fr" de la dimension *Agences* et la hiérarchie "clas_us" de la dimension *Véhicules*. La contrainte d'exclusion décrit le fait que les mesures du fait ne peuvent pas être paramétrées par les deux hiérarchies à la fois. De même, une contrainte de totalité est définie entre ces deux hiérarchies exprimant le fait qu'à chaque instance du fait *Loc_Vehicule* correspond au moins une instance de l'une ou de l'autre des deux hiérarchies. La définition d'une contrainte d'inclusion de la hiérarchie "clas_us" dans la hiérarchie "geo_zn" exprime le fait que l'ensemble des instances du fait *Loc_Vehicule* associées à la hiérarchie "clas_us" sont incluses dans les instances du fait associées à la hiérarchie "geo_zn" de la dimension *Agences*.

C5::geo_fr ⊗ clas_us, C7::clas_us ⊗ geo_zn, C6::geo_us ⊖ clas_fr, C8::clas_fr ⊗ geo_zn.

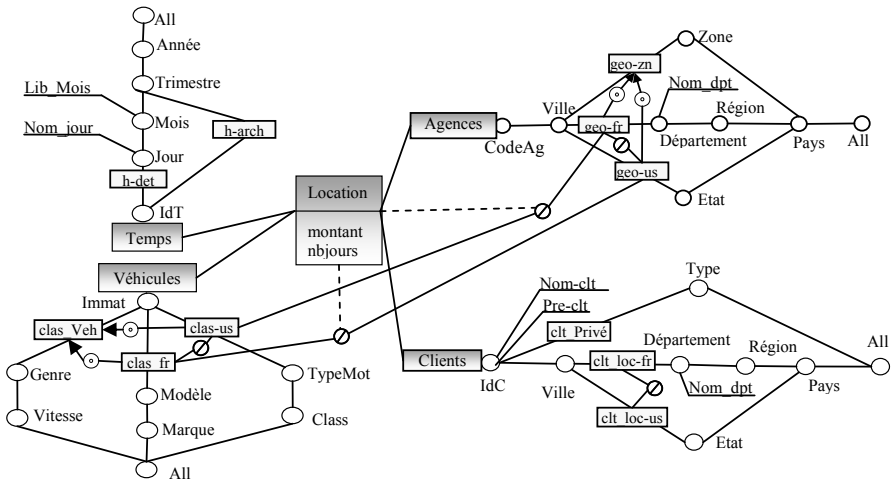


FIG 6. Schéma multidimensionnel résultant de la démarche ascendante.

Le résultat de la démarche ascendante est un schéma multidimensionnel intégrant l'expression des contraintes sémantiques (voir FIG 6).

Ce schéma représente les données extraites de la source de l'entreprise comportant toutes les données qui caractérisent son activité. Aussi, il englobe toutes les données détaillées que le décideur peut ne pas exprimer dans ses besoins (au niveau de la démarche descendante).

Pour valider les résultats des deux démarches descendante et ascendante, nous réalisons une confrontation entre les deux schémas multidimensionnels résultant de ces deux démarches. Cette confrontation est représentée dans la section suivante.

5. Confrontation

Après avoir conçu le schéma multidimensionnel des besoins en suivant la démarche descendante et le schéma multidimensionnel de l'entreprise établi à partir de la source de données, nous procédons à l'intégration de ces deux schémas. Cette fusion est réalisée en confrontant les deux schémas afin d'enlever, d'ajouter ou d'épurer quelques informations. Cette étape d'intégration nécessite l'intervention de toute l'équipe du décisionnel : informaticiens et décideurs. L'intervention des décideurs au niveau de l'intégration est primordiale afin de tester la complétude du schéma multidimensionnel. Nous notons que l'intégration de schémas relève plusieurs problématiques notamment au niveau de la gestion des conflits de nom, de contexte, d'unité de mesure et de structure. Ces problématiques ne sont pas traitées dans notre cas car nous considérons qu'au niveau de la définition du domaine de l'analyse (cf. FIG 1) nous définissons un même dictionnaire de données et un même contexte. Ainsi, l'information montant des locations aura toujours le même nom dans les deux démarches.

La confrontation est basée sur deux étapes :

- la première étape consiste à construire un schéma multidimensionnel qui comporte les données communes : fait, dimension, hiérarchies, ... ;
- la deuxième réalise l'analyse des deux schémas multidimensionnels confrontés pour intégrer les informations jugées pertinentes par l'ensemble des informaticiens et des décideurs.

Ainsi, l'intégration des données des deux schémas permet :

- la définition correcte de la granularité de l'analyse. Le choix d'un niveau de détail est très important dans la conception d'une BDM. Le choix d'une granularité trop fine risque d'augmenter la taille de la base en dérivant à partir de la source des données détaillées non pertinentes à l'analyse. Par contre, le choix d'une granularité moins fine ne permet pas d'analyser les données les plus détaillées. A ce niveau, le concepteur doit faire appel aux décideurs pour confronter leurs besoins exprimés au niveau de la démarche descendante aux granularités d'analyse détaillées obtenues lors de la démarche ascendante. Cette confrontation permet de découvrir des besoins non déclarés par les décideurs afin de déterminer une granularité adaptée aux besoins décisionnels ;
- l'épuration des données. En effet, les données extraites à partir des requêtes utilisateurs peuvent ne pas exister dans notre source de données. Dans ce cas, nous avons le choix entre enlever ces données ou bien les garder avec des valeurs vides en attendant d'avoir les informations nécessaires à leur instanciation.
- ajout des données sources. C'est le concepteur qui analyse les données de la source et détecte les informations susceptibles d'intéresser le décideur. Souvent, ce dernier

ne connaît pas le contenu de la source ni les informations susceptibles d'être analysées. Nous pouvons, par exemple, ajouter des paramètres d'analyse en détectant de nouvelles propriétés dans les dimensions définies dans notre exemple de schéma multidimensionnel ;

- ajout des mesures calculées demandées par le décideur. Ces mesures nécessitent l'utilisation d'une règle de gestion qui permet de calculer la mesure à partir d'un ensemble de données source ;
- intégration de toutes les contraintes sémantiques. Ces contraintes proviennent soit de la démarche descendante, soit de la démarche ascendante et détectées lors de l'analyse de la source.

Exemple 14 Lors de la confrontation des deux schémas que nous avons obtenus à partir des démarches descendante et ascendante, nous avons été amenés à réaliser les transformations suivantes :

- l'ajout du fait Performance qui n'a pas été considéré par le concepteur lors de l'analyse de la source ;
- l'enrichissement des dimensions Agences, Clients et Véhicules avec de nouveaux paramètres et attributs faibles ;
- la validation des contraintes entre la hiérarchie de la géographie française et celle de la géographie américaine. De même, pour les contraintes entre la hiérarchie de la classification française des véhicules et celle adoptée par les agences américaines.

La FIG 7 présente le schéma multidimensionnel contraint résultant de cette confrontation.

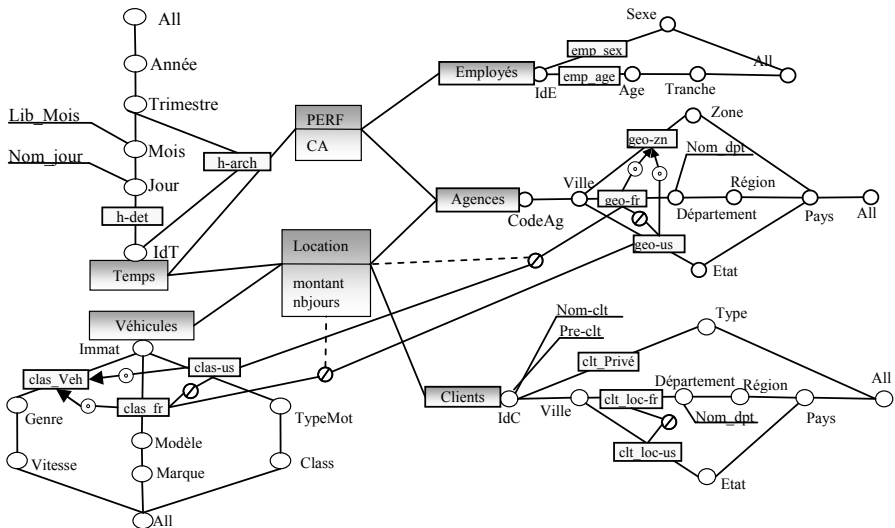


FIG 7. Résultat de la confrontation des démarches ascendante et descendante

Le concepteur doit familiariser les décideurs aux concepts du schéma multidimensionnel et tester ensuite leur niveau de satisfaction en leur montrant les analyses décisionnelles possibles à travers ce schéma. Finalement, une étape de test auprès des décideurs est nécessaire afin de valider le schéma multidimensionnel final.

6. Conclusion

Cet article propose une méthode de conception d'une BDM cohérente. Pour répondre à cet objectif, nous avons proposé une méthode mixte basée sur une démarche descendante et ascendante. Lors de la démarche descendante, nous avons défini un ensemble d'étapes qui permettent de collecter les besoins des décideurs, de les spécifier et de les formaliser sous forme de schéma multidimensionnel en constellation intégrant un ensemble de contraintes sémantiques. La démarche ascendante permet de collecter les données sources et de construire le schéma multidimensionnel dédié à l'aide à la prise de décision. Une phase de confrontation a permis de construire un schéma multidimensionnel intégrant les besoins exprimés par les décideurs tout en tenant compte des données dérivées des sources. L'avantage de notre méthode est la combinaison des démarches ascendante et descendante permettant de tirer profit de l'analyse des sources pour extraire les informations que le décideur oublie ou n'arrive pas à exprimer lors de la collecte des besoins. En outre, cette méthode est établie sur un modèle multidimensionnel intégrant un ensemble de contraintes favorisant la définition de BDMs cohérentes. En appliquant notre démarche ascendante, nous proposons d'automatiser la conception du schéma multidimensionnel à l'aide de notre système d'aide à la conception de BDM. Nous envisageons dans le futur d'intégrer dans ce prototype un module permettant de réaliser la confrontation et l'intégration automatique des deux schémas résultants de l'étape descendante et de l'étape ascendante. En outre, notre proposition permet de modéliser les besoins des décideurs sous forme de schéma multidimensionnel contraint. Or, ces besoins peuvent évoluer dans le temps (nouvel indicateur ou paramètre d'analyse, nouvel axe d'analyse, ...). Nous envisageons de poursuivre ces travaux par la prise en compte de ces nouveaux besoins au travers de la gestion de l'évolution du schéma multidimensionnel.

Références

- [Bonifati et al, 2001] A. Bonifati, F. Cattaneo, S. Ceri, A. Fuggetta, S. Paraboschi "Designing Data Marts for Data Warehouses", ACM Transactions on Software Engineering Methodology, (4)2001.
- [Carneiro et Brayner, 2002] L. Carneiro, A. Brayner, "X-META : A Methodology for Data Warehouse Design with Metadata Management". In 4th Int. Workshop on Design and Management of Data Warehouses (DMDW'02), Toronto, Canada, p. 13-22, May 2002.
- [Carpani et Ruggia, 2001] F. Carpani, R. Ruggia, "An Integrity Constraints Language for a Conceptual Multidimensional Data Model". In 13th Int. Conf. on Software Engineering & Knowledge Engineering (SEKE'01), Argentina, 2001.
- [Codd et al, 1993] E. F. Codd, S.B. Codd, C.T. Salley, "Providing OLAP (On Line Analytical Processing) to Users-Analysts: An IT Mondate", Rapport technique, E.F. Codd and Associates, 1993.
- [Ghozzi et al, 2004] F. Ghozzi, F. Ravat, O. Teste, G. Zurfluh, "*Contraintes pour modèle et langage multidimensionnels*". Dans RSTI-ISI : Fouille, transactions, évaluation dans les bases de données, Hermes –Lavoisier, Vol. 9, N. 1, p.9-34, 2004.
- [Golfarelli et al, 2002] M. Golfarelli, S. Rizzi, E. Saltarelli, "WAND: A CASE Tool for Workload-Based Design of a Data Mart". In 10th National Convention on Systems Evolution for Data Bases (SEBD'02), Portoferraio - Island of Elba, Italy, ed. Paul Ciaccia, the Faustus Rabitti, Hard Giovanni, p. 422-426, June 2002.

- [Hümmer et al, 2002] W. Hümmer, W. Lehner, A. Bauer, L. Schlesinger, "A Decathlon in Multidimensional Modeling: Open Issues and Some Solutions". Dans 4th Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK 2002), Aix-en-Provence, France, p. 275-285, September, 2002.
- [Hurtado et Mendelzon, 2002] C.A. Hurtado, A.O. Mendelzon, "OLAP Dimension Constraints". In 21st ACM SIGACT-SIGMOD Symposium on Principles of Database Systems (PODS'02), Madison, USA, p. 169-179, Jun 2002.
- [Kimball et Ross, 2002] R. Kimball, M. Ross, "The Data Warehouse Toolkit", Wiley, New York, second edition, 2002.
- [List et al, 2002] B. List, R. Bruckner, K. Machaczek, J. Sciefer, "A Comparison of Data warehouse development Methodologies case study of the process Warehouse". In 13th Int. Conf. on Database and Expert Systems Applications (DEXA'02), Aix-en-Provence, France, LNCS 2453, p 203-215, Sept 2002.
- [Lujan-Mora et al, 2004] S. Luján-Mora, J. Trujillo, P. Vassiliadis, "Advantages of UML for Multidimensional Modeling". In 6th Int. Conf. on Enterprise Information Systems (ICEIS'04), Porto, Portugal, p. 298-305, April, 2004.
- [Moody et Kortink, 2000] D. L. Moody., M. A.R Kortink. "From Enterprise Models to Dimensional Models:A Methodology for Data Warehouse and Data Mart Design". In 2nd Int. Workshop on Design and Management of Data Warehouses (DMDW'2000) Stockholm, Suede, paper 5, June 2000.
- [Prat et Akoka, 2002] N. Prat, J. Akoka, "From UML to ROLAP multidimensional databases using a pivot model". In 18th journées Bases de données avancées (BDA02), Évry, France, p. 171-195, October 2002.
- [Tsois et al, 2001] A. Tsois, N. Karayannidis, T. Sellis, "MAC: Conceptual Data Modeling for OLAP". Dans International Workshop on Design and Management of Data Warehouses (DMDW'2001) Interlaken, Switzerland, juin 2001.
- [Trujillo et al, 2002] J. C. Trujillo, S. Luján-Mora, E. Medina. "The GOLD Model CASE Tool: an environment for designing OLAP applications". In 4th Int. Conf. on Enterprise Information Systems (ICEIS'02), Ciudad Real, Spain, p. 699-707, April 2002.

Summary

This paper describes a multidimensional database design method for decisional systems. A mixed method is provided, which is based on a top-down and bottom-up processes. Top-down process enables to elicit and consolidate requirements of decision-makers. This process is composed of three steps: (i) collect of user needs, (ii) specification and (iii) formalisation of these needs in semantic constraint-based multidimensional constellation schema. The bottom-up process defines the candidate multidimensional schema from the information system. A confrontation phase integrates multidimensional schemas obtained from these two processes. The advantage of our method is the integration of two complimentary processes which both emphasize business needs and integrate the semantic of existing operational database. This method is based on a multidimensional constraint model supporting the definition of consistent multidimensional databases.