



HAL
open science

Opérateurs OLAP dans les bases de données multidimensionnelles multifonctions

Ali Hassan, Franck Ravat, Olivier Teste, Ronan Tournier, Gilles Zurfluh

► **To cite this version:**

Ali Hassan, Franck Ravat, Olivier Teste, Ronan Tournier, Gilles Zurfluh. Opérateurs OLAP dans les bases de données multidimensionnelles multifonctions. 9èmes Journées Francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2013), Jun 2013, Blois, France. pp.73-83. hal-04083744

HAL Id: hal-04083744

<https://hal.science/hal-04083744>

Submitted on 27 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 12624

Official URL: <http://editions-rnti.fr/?inprocid=1001896>

To cite this version : Hassan, Ali and Ravat, Franck and Teste, Olivier and Tournier, Ronan and Zurfluh, Gilles *Opérateurs OLAP dans les bases de données multidimensionnelles multifonctions*. (2013) In: 9èmes Journées Francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2013), 13 June 2013 - 14 June 2013 (Blois, France).

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Opérateurs OLAP dans les bases de données multidimensionnelles multifonctions

Ali Hassan*, Franck Ravat*, Olivier Teste**, Ronan Tournier*, Gilles Zurfluh*

*Université Toulouse 1 Capitole - IRIT (UMR 5505)
118, Route de Narbonne - 31062 Toulouse cedex 9 (France)
**Université Toulouse 2 / IUT Blagnac - IRIT (UMR 5505)
1, Place Georges Brassens BP 60073 31703 Blagnac cedex
{hassan, ravat, teste, tournier, zurfluh}@irit.fr

Résumé. Notre modèle conceptuel des bases de données multidimensionnelles permet d'associer à chaque mesure plusieurs fonctions d'agrégation en fonction des dimensions, des hiérarchies et des niveaux de granularité. Cet article étudie les impacts de ce modèle sur la table multidimensionnelle (TM) et l'algèbre OLAP. A cause des caractéristiques de ce modèle, il y a des changements qui se produisent sur l'algorithme interne de certains opérateurs OLAP. Les changements peuvent affecter la TM même. Ils peuvent nécessiter d'adapter la TM à présenter plusieurs fonctions d'agrégation.

1 Introduction

Dans les modèles multidimensionnels, les mesures sont analysées selon plusieurs dimensions. Les décideurs peuvent utiliser les opérateurs OLAP (Ravat et al., 2008) pour étudier les mesures en fonction de différents niveaux de granularité (paramètres). Les données sont ainsi regroupées selon le niveau sélectionné et calculées en utilisant des fonctions d'agrégation. Les bases de données multidimensionnelles classiques n'offrent que la possibilité d'utiliser la même fonction pour agréger une mesure dans tout l'espace multidimensionnel. Cette possibilité n'est pas suffisante pour faire face aux situations qui nécessitent l'utilisation de plusieurs fonctions pour agréger la même mesure. Par exemple, la température moyenne mensuelle ou annuelle est obtenue par le calcul de la moyenne (AVG) des températures quotidiennes. Par contre, on peut calculer les températures moyennes des départements de deux manières. La première, **simple**, adopte la méthode utilisée par les présentateurs de la météo en choisissant une ville considérée comme représentative du département (préfecture). La deuxième, **scientifique**, prend en compte les températures de toutes les villes du département.

1.1 Etat de l'art

La plupart des propositions existantes considèrent qu'une mesure est associée à une fonction d'agrégation qui sera utilisée à tous les niveaux d'agrégation modélisés. Les travaux de Gyssens et Lakshmanan (1997), Vassiliadis et Skiadopoulos (2000) ne précisent pas les fonctions d'agrégation au niveau des mesures, mais ils laissent la possibilité d'utiliser pour chaque

mesure plusieurs fonctions d'agrégation au cours du processus OLAP. Cela donne une grande flexibilité, mais laisse la possibilité à l'utilisateur de commettre des erreurs en utilisant des fonctions inappropriées. Les travaux de Pedersen et al. (2001) proposent de lier à chaque mesure un ensemble de fonctions qui ne comprend que les fonctions valides. Néanmoins, chaque fonction est utilisée uniformément pour toutes les dimensions et tous les niveaux des hiérarchies qui constituent l'espace multidimensionnel. Des travaux plus récents (Abelló et al., 2006) permettent d'utiliser une fonction d'agrégation différente pour chaque dimension, sans donner la possibilité de faire évoluer la fonction avec les niveaux de hiérarchies. Cette limite a été levée par les modèles d'agrégation des travaux les plus récents (Prat et Akoka, 2010) et (Boulil et al., 2011). Ces travaux nous permettent d'associer à chaque mesure une fonction d'agrégation pour chaque dimension, chaque hiérarchie ou chaque niveau d'agrégation. Toutefois ils ne traitent pas le cas des fonctions d'agrégation non-commutatives.

Afin de lever ces limites, nous avons développé un modèle conceptuel de représentation des données multidimensionnelles (Hassan et al., 2012b,a). Ce modèle permet d'associer une mesure avec différentes fonctions d'agrégation en fonction des dimensions, des hiérarchies et des niveaux d'agrégation appelés paramètres. En outre, le modèle permet de contrôler la validité des calculs des fonctions en définissant l'ordre d'exécution des fonctions non-commutatives. Il traite également le cas où la mesure ne peut pas être calculée à partir du niveau de base en utilisant des contraintes d'agrégation. Les contraintes d'agrégation définissent le niveau à partir duquel l'agrégation doit être calculée.

L'application de cette proposition a des impacts sur la table multidimensionnelle (TM) et les opérateurs d'OLAP. Certains effets peuvent être visibles par l'utilisateur. Dans cet article, nous présentons les modifications à apporter pour adapter la TM et les opérateurs OLAP à notre modèle multifonctions.

1.2 Organisation de l'article

La section 2 présente notre modèle conceptuel pour intégrer plusieurs fonctions d'agrégation pour la même mesure. La section 3 décrit les extensions nécessaires à la TM et l'algèbre multidimensionnelle pour s'adapter à notre modèle.

2 Modèle conceptuel de données multifonctions

Un *schéma multidimensionnel*, noté S , est défini par $(F, D, Star, Aggregate)$. Ce schéma est une extension de nos précédents travaux (Hassan et al., 2012b,a) où nous ajoutons un nouveau type d'agrégation (multiple hiérarchique) et nous revisitons le mécanisme de l'ordre d'exécution afin de rendre notre modèle de représentation plus expressif.

- $F = \{F_1, \dots, F_n\}$ un ensemble fini de faits ;
- $D = \{D_1, \dots, D_m\}$ un ensemble fini de dimensions ;
- $Star : F \rightarrow 2^D$ est une fonction qui associe chaque fait à un ensemble de dimensions en fonction desquelles il peut être analysé ;
- $Aggregate : M \rightarrow 2^{N^* \times \mathcal{F} \times 2^D \times 2^H \times 2^P \times N^-}$ associe chaque mesure à un ensemble de fonctions d'agrégation (\mathcal{F} est un ensemble fini de fonctions d'agrégation, M, H, P sont respectivement les ensembles de *mesures*, de *hiérarchies*, de niveaux d'agrégation appelés *paramètres*). Elle permet de définir quatre types de fonction d'agrégation :

- Générale (si $2^D = \emptyset$, $2^H = \emptyset$ et $2^P = \emptyset$) pour agréger la mesure entre n'importe quel paramètre. Elle représente la fonction d'agrégation dans le modèle classique ;
- Multiple dimensionnelle (si $2^H = \emptyset$ et $2^P = \emptyset$) pour agréger la mesure sur toute la dimension considérée ;
- Multiple hiérarchique (si $2^P = \emptyset$) pour agréger la mesure sur toute la hiérarchie ;
- Différentiée (si $2^D \neq \emptyset$, $2^H \neq \emptyset$ et $2^P \neq \emptyset$) pour agréger la mesure entre le paramètre considéré et celui directement supérieur.

Pour traiter le cas des fonctions non-commutatives, \mathbb{N}^* associe à chaque fonction un numéro d'ordre qui représente la priorité d'exécution. La fonction avec l'ordre le plus petit a la priorité la plus élevée. Si les fonctions sont commutatives, elles sont du même ordre.

\mathbb{N}^- sert à contraindre une agrégation en indiquant un niveau d'agrégation spécifique à partir duquel l'agrégation considérée doit se calculer. Une agrégation non contrainte sera associée à 0 tandis qu'une agrégation contrainte sera associée à une valeur négative pour forcer le calcul à partir d'un niveau inférieur choisi par rapport au niveau considéré.

Grâce à la fonction Star, nous obtenons un schéma structurel qui visualise les éléments structurels (faits, dimensions et hiérarchies) en masquant les mécanismes d'agrégation, (cf. figure 1 (a)). Le formalisme graphique que nous utilisons repose sur les propositions de Golfarelli et al. (1998), Ravat et al. (2001). Le schéma de la figure 1 (a) correspond à l'analyse des températures moyennes. Le fait 'Température' est associé à trois dimensions : 'Géographie', 'Dates' et 'Temps'. La dimension 'Géographie' se compose de deux hiérarchies correspondant aux manières d'agréger les données : 'Simple' et 'Scientifique'. La dimension 'Temps' a une seule hiérarchie qui ordonne les granularités des horaires auxquelles les températures ont été mesurées pendant la journée. La dimension 'Dates' comprend plusieurs hiérarchies qui organisent les niveaux des granularités de la journée.

Pour chaque mesure, un schéma d'agrégation peut être obtenu grâce à la fonction Aggregate. Ce schéma détaille les mécanismes d'agrégation s'appliquant à la mesure en simplifiant les éléments structurels autant que possible. La figure 1 (b) illustre le schéma d'agrégation de la température moyenne (Tem_Moy). Elle a une fonction d'agrégation générale (Avg), une fonction multiple hiérarchique (Select_center) sur la hiérarchie 'Hgéo_simp' et des fonctions différenciées (Avg et Avg_W) sur la hiérarchie 'Hgéo_Scien'.

La fonction Select_center(I, M) prend deux paramètres numériques. Elle rend la valeur M qui correspond au Max(I). Par exemple, si on applique Select_center(Niveau_Administratif, Temp_Moy) au niveau région ou département, elle rend les températures de la capitale régionale ou de la préfecture (ville ayant le niveau administratif maximal). La fonction Avg_W(X, Y) prend deux entrées numériques. Elle retourne la moyenne des valeurs X pondérée par Y. Autrement dit, la moyenne pondérée $Avg_W(X, Y) = \frac{\sum(X \times Y)}{\sum Y}$.

Dans les schémas d'agrégation, les hiérarchies sont présentées de manière disjointes (chaque hiérarchie est représentée par un chemin complet du paramètre racine au paramètre extrémité ALL) contrairement au schéma structurel où elles sont présentées en version compacte (les parties communes des hiérarchies sont fusionnées, formant une arborescence de paramètres). Les fonctions d'agrégation sont modélisées par des losanges. Chaque losange indique l'ordre d'exécution et la contrainte d'agrégation. Les agrégations avec des contraintes fixées à -1 sont calculées à partir du niveau directement inférieur ; par exemple, la température moyenne 'Tem_Moy' par région est obligatoirement calculée à partir des températures moyennes des

départements. Dans l'hypothèse où nous aurions choisi de calculer cette température moyenne par région à partir des températures des villes, la contrainte aurait été fixée à -2.

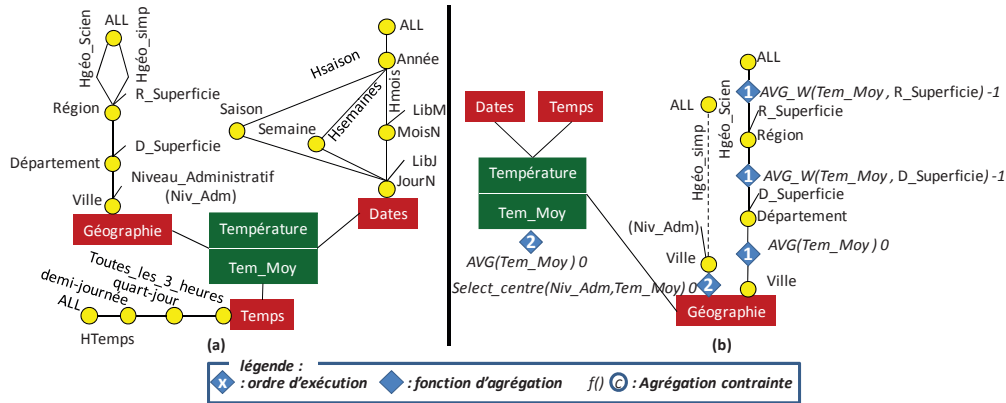


FIG. 1 – Représentation des schémas multidimensionnels.

3 Manipulation OLAP

Dans cet article, nous concentrons notre étude sur l'analyse multidimensionnelle OLAP (Codd, 1993) appliquée à une BDM multifonctions.

3.1 Table multidimensionnelle étendue

Dans de précédents travaux (Ravat et al., 2008), nous avons défini une algèbre orientée décideur pour expliciter les processus d'analyse en ligne. Cette algèbre repose sur le concept de Table Multidimensionnelle (TM). Une TM est une table à deux dimensions, utilisée en tant que source ou cible d'un opérateur de l'algèbre assurant ainsi la fermeture de l'algèbre OLAP. Ce choix est justifié par le fait que, lors de leurs analyses, les décideurs manipulent couramment les données au travers de tableaux à deux dimensions (ligne et colonne) de par leur simplicité d'interprétation et leur précision (Gyssens et Lakshmanan, 1997).

Dans le cadre de cet article, nous avons étendu le concept de TM pour qu'il puisse supporter les principes du modèle conceptuel multifonctions, notamment en y intégrant les fonctions d'agrégation dans sa définition. Une TM étendue est donc définie comme suit :

$$TM = (F, \langle (D_L, h_L, \langle p_{L1}, p_{L2}, \dots \rangle), (D_C, h_C, \langle p_{C1}, p_{C2}, \dots \rangle) \rangle, \langle \{Aggregate(m_1)\}, \{Aggregate(m_2)\}, \dots \rangle, Pred)$$

Où :

- F : fait à analyser ;
- D_L, D_C : dimensions affichées en ligne et en colonne respectivement ;
- h_L, h_C : hiérarchies utilisées pour naviguer respectivement en ligne et en colonne ;
- $p_{L1}, p_{L2}, \dots, p_{C1}, p_{C2}, \dots$: les paramètres affichés ;
- $Aggregate(m_1), Aggregate(m_2), \dots$: fonctions d'agrégation respectivement associées aux mesures affichées (m_1, m_2, \dots). C'est la partie que nous ajoutons à la définition de la TM ;

- Pred : le prédicat de sélection sur le fait et/ou les dimensions pour limiter l'ensemble de valeurs à analyser.

L'exemple suivant donne la définition d'une table multidimensionnelle pour analyser la température moyenne par mois et par département dans la région Rhône-Alpes :

TM = (Température, {Tem_Moy}
 <(Dates, Hmois, <Année, MoisN>),
 (Géographie, HgéO_Scien, <Région, Département>)>,
 <{(2, Avg(Tem_Moy), {}, {}, {}, 0),
 (1, Avg(Tem_Moy), {Géographie}, {HgéO_Scien}, {Ville}, 0)}>,
 Géographie.Région = 'Rhône-Alpes')

La représentation graphique de cette définition est illustrée dans la figure 2 (b) :

Température AVG(Tem_Moy)			Géographie HgéO_Scien		
			Région	Rhône-Alpes	
			Département	Rhône	Isère
Dates	Année	MoisN			
Hmois	2012	2012-1		1	2
		2012-2		5	4
Géographie.Région = 'Rhône-Alpes'					

Température <2> AVG(Tem_Moy)			Géographie HgéO_Scien		
			Région	Rhône-Alpes	
			Département	Rhône	Isère
			<1> AVG(Tem_Moy)		
Dates	Année	MoisN			
Hmois	2012	2012-1		1	2
		2012-2		5	4
Géographie.Région = 'Rhône-Alpes'					

(a) Classique

(b) étendue

FIG. 2 – Représentation graphique d'une table multidimensionnelle.

Afin d'adapter la visualisation de la TM pour présenter plusieurs fonctions d'agrégation qui peuvent être utilisées pour une seule mesure (figure 2 (b)), la TM permet d'afficher les fonctions d'agrégation (avec leurs entrées, leurs ordres d'exécution et leurs contraintes) utilisées pour obtenir les éléments affichés.

- La fonction générale est affichée au lieu de la mesure dans la cellule du fait,
- La fonction multiple dimensionnelle est affichée à côté du nom de la dimension correspondante,
- La fonction multiple hiérarchique est affichée à côté du nom de la hiérarchie correspondante,
- La fonction différenciée est affichée dans la cellule du paramètre correspondant.

La visualisation est basée sur la simplification des fonctions autant que possible :

- Simplification des contraintes d'agrégation : si une fonction n'était pas contrainte (elle a une contrainte de valeur 0), la TM ne présente pas cette contrainte,
- Simplification des ordres d'exécution : si toutes les fonctions d'agrégation affichées ont le même ordre d'exécution, la TM cache tous ces ordres d'agrégation,
- Réduire le nombre des fonctions affichées :
 - Si tous les paramètres affichés sur une dimension ont une fonction différenciée, la TM ne présente ni la fonction multiple hiérarchique ni la fonction multiple dimensionnelle sur la dimension considérée,
 - Si une hiérarchie affichée a une fonction d'agrégation, la TM ne présente pas la fonction multiple dimensionnelle de la dimension considérée,

- Si les deux dimensions affichées ont une fonction multiple dimensionnelle ou multiple hiérarchique pour la hiérarchie affichée ou encore une fonction différenciée pour chaque paramètre affiché, la TM ne présente pas la fonction générale.

3.2 Algèbre multidimensionnelle étendue

Comme indiqué précédemment, les opérateurs de l’algèbre sont appliqués sur les TM (Ravat et al., 2008). Comme nous utilisons un modèle multifonctions, il y a des changements qui se produisent sur certains opérateurs. Ces changements ne touchent pas la fonctionnalité mais ils peuvent toucher le mécanisme interne. Selon ce mécanisme interne, la requête, qui effectue l’opération demandée, sera construite. Avant d’étudier ces changements, nous présentons notre exemple de motivation.

3.2.1 Exemple de motivation

Figure 3 (a) contient un échantillon de données simplifié compatible avec notre exemple ci-dessus. Dans cet exemple, la température est enregistrée deux fois par jour dans le département ‘Rhône’ et une fois dans le département ‘Isère’. Si l’utilisateur souhaite analyser les températures moyennes des régions par mois, deux fonctions d’agrégation doivent être utilisées :

- ‘Avg_W(Tem_Moy, D_Superficie)’ pour calculer les températures moyennes des régions à partir des températures des départements pondérées par les superficies des départements ‘D_Superficie’ ;
- ‘Avg(Tem_Moy)’ pour calculer les températures mensuelles à partir des températures quotidiennes.

Si on applique d’abord la fonction ‘Avg_W(Tem_Moy, D_Superficie)’, on obtient les températures moyennes des régions par jour et temps (Figure 3 (b)). Ensuite, si on applique la fonction ‘Avg(Tem_Moy)’, on obtient les températures moyennes des régions par mois demandées (Figure 3 (c)). Mais, si on applique d’abord la fonction ‘Avg(Tem_Moy)’, on obtient les températures moyennes des départements par mois (Figure 3 (d)). Ensuite, si on applique la fonction ‘Avg_W(Tem_Moy, D_Superficie)’, on obtient les températures moyennes des régions par mois (Figure 3 (e)).

La différence entre les résultats obtenus dans les Figures 3 (c) et 3 (e) prouve la nécessité d’utiliser l’ordre d’exécution. Le rôle clé de l’ordre d’exécution est de forcer l’application des fonctions d’agrégation dans un ordre précis pour ne pas avoir un résultat erroné à cause de la non-commutativité. Le choix d’un ordre valide dépend des besoins de l’utilisateur. Il peut différer d’un cas à l’autre, même si les fonctions sont les mêmes dans les deux cas. Le modèle permet de fixer l’ordre qui donne un résultat valide pour l’utilisateur. Dans cet exemple, le résultat qui correspond aux besoins de l’utilisateur est dans la Figure 3 (c). Cela justifie pourquoi nous donnons une valeur de 1 pour l’ordre d’exécution de la fonction ‘Avg_W’ alors que nous donnons 2 pour l’ordre d’exécution de la fonction ‘Avg’ (Figure 1 (b)).

Nous concluons de l’exemple précédent que l’application d’une fonction d’agrégation avec un ordre d’exécution inférieur après une fonction d’agrégation avec un ordre d’exécution supérieure donnera un résultat erroné.

Région	Département	D Superficie	Date	Temps	Tem Moy
Rhône-Alpes	Rhône	3249	1/1/2012	00h00	-1
Rhône-Alpes	Rhône	3249	1/1/2012	12h00	1
Rhône-Alpes	Isère	7431	1/1/2012	12h00	2
Rhône-Alpes	Rhône	3249	2/1/2012	00h00	0
Rhône-Alpes	Rhône	3249	2/1/2012	12h00	2
Rhône-Alpes	Isère	7431	2/1/2012	12h00	2

(a) Températures moyennes des départements par jour et temps

Région	Date	Temps	Tem Moy
Rhône-Alpes	1/1/2012	00h00	-1
Rhône-Alpes	1/1/2012	12h00	1.7
Rhône-Alpes	2/1/2012	00h00	0
Rhône-Alpes	2/1/2012	12h00	2

(b) Températures moyennes des régions par jour et temps

Région	Mois	Tem Moy
Rhône-Alpes	2012-1	0.67

(c) Températures moyennes des régions par mois

Région	Département	D Superficie	Mois	Tem Moy
Rhône-Alpes	Rhône	3249	2012-1	0.5
Rhône-Alpes	Isère	7431	2012-1	2

(d) Températures moyennes des départements par mois

Région	Mois	Tem Moy
Rhône-Alpes	2012-1	1.54

(e) Températures moyennes des régions par mois

FIG. 3 – Exemple de motivation.

Pour effectuer les opérateurs RollUp¹ et Rotate², il faut changer un niveau d'agrégation actuel vers un autre niveau supérieur :

- Pour l'opérateur RollUp : du niveau actuel au niveau cible ;
- Pour l'opérateur Rotate : du niveau actuel au niveau ALL de la dimension qui va disparaître de la TM.

Pour cela il faut appliquer de nouvelles fonctions d'agrégation après avoir appliqué les fonctions précédentes qui donnent la TM actuelle (avant d'appliquer les opérateurs RollUp ou Rotate). Or, les ordres d'exécution des fonctions précédentes peuvent être supérieurs aux ordres d'exécution des nouvelles fonctions, c'est-à-dire, qu'il n'est pas possible d'appliquer les nouvelles fonctions après avoir appliqué les fonctions précédentes. Cela nécessite la mise en place de mécanismes internes plus coûteux pour ces opérateurs RollUp et Rotate.

Dans cette section, nous étudions l'opérateur RollUp parce qu'il est parmi les opérateurs les plus utilisés. Nous commençons d'abord par présenter l'opérateur RollUp classique. Ensuite, nous détaillons les changements de RollUp adapté au modèle multifonctions.

3.2.2 RollUp classique

Afin d'étudier cet opérateur, nous supposons avoir une TM pour analyser les températures moyennes des départements par mois (Figure 4 (a)). A partir de cette TM, nous voulons effectuer une opération de RollUp pour analyser les températures par régions (Figure 4 (b)).

Pour étudier l'opérateur RollUp classique, nous prenons en compte le modèle multidimensionnel classique où il y a une seule fonction pour agréger la mesure. La requête SQL qui réalise l'analyse des températures moyennes des départements par mois (Figure 4 (a)) est :

```
R1 : Result1 = SELECT G.DEPARTEMENT, D.MOISN, AVG(TT.TEM_MOY) AS TEM_MOY
FROM DATES D, GEOGRAPHIE G, TEMPERATURE TT
WHERE TT.ID_VILLE = G.ID_VILLE AND TT.ID_DATE = D.ID_DATE
```

1. RollUp(T_{SRC} , D , p_i) : modifie le niveau de la granularité de la dimension (D) vers un niveau moins détaillé (p_i) en supprimant un (ou plusieurs) paramètre en ligne ou en colonne d'une TM (T_{SRC}).

2. Rotate(T_{SRC} , D_{old} , D_{new} [, H_{new}]) : remplace un axe d'analyse (D_{old}) par un autre (D_{new}) dans une TM (T_{SRC}). Il est possible de préciser la hiérarchie (H_{new}) de la nouvelle dimension pour l'utiliser dans la TM résultante.

GROUP BY G.DEPARTEMENT, D.MOISN ;

Pour exécuter le forage (RollUp) souhaité, on peut profiter des résultats de la TM précédente si la fonction d'agrégation est distributive ou algébrique (dans ce dernier cas, il est nécessaire de stocker des valeurs intermédiaires) (Gray et al., 1996). Dans notre exemple, la fonction est algébrique (Avg). Les valeurs intermédiaires demandées sont la somme des températures moyennes des départements (sum_Tem_Moy) et le nombre des occurrences (count_Tem_Moy). Ainsi, la requête R₂, qui réalise le RollUp correspondant à la TM de la Figure 4 (b), bénéficie des résultats (Result1) de la requête précédente R₁. Autrement dit, pour mettre en œuvre une opération de forage vers le haut (RollUp), on n'a pas besoin d'accéder et de charger les valeurs des mesures, mais on peut utiliser les valeurs intermédiaires présentées dans la TM.

R₂ : **SELECT** REGION, MOISN, SUM(sum_TEM_MOY)/SUM(count_TEM_MOY) **AS** TEM_MOY
FROM GEOGRAPHIE G, Result1 R1
WHERE R1.DEPARTEMENT = G.DEPARTEMENT
GROUP BY REGION, MOISN ;

Température Tem_Moy			Géographie Hgéó_Scien		
			Région	Rhône-Alpes	
			Département	Rhône	Isère
Dates Hmois	Année	MoisN			
	2012	2012-1		1	2
		2012-2		5	4

→

Température Tem_Moy			Géographie Hgéó_Scien	
			Région	Rhône-Alpes
Dates Hmois	Année	MoisN		
	2012	2012-1		1.67
		2012-2		4.3

(a) T RollUp(T, Géographie, Région) = T' (b) T'

FIG. 4 – RollUp.

3.2.3 RollUp étendu

Dans ce modèle multifonctions, puisqu'il y a plusieurs fonctions qui agrègent la mesure, la requête SQL qui réalise l'analyse des températures moyennes des départements (sur la hiérarchie 'Hgéó_Scien') par mois (Figure 4 (a)) devient plus complexe :

R₃ : **Result2** = **SELECT** MOISN, DEPARTEMENT, AVG(TEM_MOY) **AS** TEM_MOY
FROM (SELECT D.MOISN, G.DEPARTEMENT, D.JourN, T.TOUTE_LES_3_HEURES,
AVG(TT.TEM_MOY) **AS** TEM_MOY
FROM DATES D, GEOGRAPHIE G, TEMPERATURE TT, TEMPS T
WHERE TT.ID_TEMPS = T.ID_TEMPS **AND** TT.ID_VILLE = G.ID_VILLE
AND TT.ID_DATE = D.ID_DATE
GROUP BY D.MOISN, G.DEPARTEMENT, D.JourN, T.TOUTE_LES_3_HEURES)
GROUP BY MOISN, REGION, DEPARTEMENT ;

Pour effectuer une opération de forage vers le haut dans un modèle multifonctions, nous pouvons distinguer deux cas selon la fonction ou les fonctions d'agrégation correspondant au forage demandé :

Le premier cas, tous les ordres d'exécution des fonctions, qui agrègent la mesure entre le paramètre actuel et le paramètre demandé, sont supérieurs ou égaux aux ordres d'exécution des fonctions qui agrègent la mesure entre les paramètres de base et les paramètres actuels, y compris les niveaux ALL des dimensions qui n'apparaissent pas dans la TM. Par exemple, si nous voulons faire un forage vers le haut pour analyser les températures moyennes des départements par années. La fonction qui agrège les températures moyennes entre le niveau 'moisN' et 'Année' est la fonction générale 'Avg(Tem_Moy)'. Cette fonction a un ordre d'exécution de

valeur 2 qui est égal ou supérieur aux ordres d'exécution des fonctions qui agrègent les températures moyennes entre les niveaux de base et les niveaux 'Département', 'moisN', 'ALL' de la dimension 'Temps'. Dans ce cas, de la même manière que le RollUp classique, nous pouvons profiter des valeurs de la mesure présentées dans la TM comme dans la requête R₄.

```
R4 : SELECT ANNEE, DEPARTEMENT, SUM(sum_TEM_MOY)/SUM(count_TEM_MOY) AS TEM_MOY
FROM DATES D, Result2 R2
WHERE R2.MOISN = D.MOISN
GROUP BY ANNEE, DEPARTEMENT ;
```

Le deuxième cas, un ordre d'exécution d'une fonction, qui agrège la mesure entre le paramètre actuel et le paramètre demandé, est inférieur à un ordre d'exécution d'une fonction qui agrège la mesure entre les paramètres de base et les paramètres actuels. Par exemple, si nous voulons faire un forage vers le haut pour analyser les températures moyennes des régions par mois à partir des températures moyennes des départements par mois. La fonction qui agrège les températures moyennes entre les niveaux 'Département' et 'Région' sur la hiérarchie 'Hgéο_Scien' est la fonction 'Avg_W(Tem_Moy, D_Superficie)'. Cette fonction a un ordre d'exécution de valeur 1 qui est inférieur à l'ordre d'exécution de la fonction générale qui agrège les températures moyennes entre le niveau de base et le niveau 'moisN'. Cela signifie qu'il faut calculer les températures moyennes par région avant de calculer les températures moyennes par mois. Dans ce cas, nous ne pouvons pas profiter des valeurs de la mesure présentées dans la TM. Il faut donc recalculer les valeurs de la mesure à partir des niveaux de base comme la requête R₅.

```
R5 : SELECT MOISN, REGION, AVG(TEM_MOY) AS TEM_MOY
FROM (SELECT MOISN, REGION, TOUTE_LES_3_HEURES, JourN,
AVG_W(DATA_WEIGHTED(TEM_MOY, D_SUPERFICIE)) AS TEM_MOY
FROM (SELECT D.MOISN, G.REGION, G.DEPARTEMENT, T.TOUTE_LES_3_HEURES, D.JourN,
G.D_SUPERFICIE, AVG(TT.TEM_MOY) AS TEM_MOY
FROM DATES D, GEOGRAPHIE G, TEMPERATURE TT, TEMPS T
WHERE TT.ID_TEMPS = T.ID_TEMPS AND TT.ID_VILLE = G.ID_VILLE
AND TT.ID_DATE = D.ID_DATE
GROUP BY D.MOISN, G.REGION, G.DEPARTEMENT, T.TOUTE_LES_3_HEURES, D.JourN,
G.D_SUPERFICIE)
GROUP BY MOISN, REGION, TOUTE_LES_3_HEURES, JourN)
GROUP BY MOISN, REGION ;
```

4 Conclusion

Dans cet article nous étudions les impacts de notre modèle conceptuel de représentation des données multidimensionnelles sur la TM et l'algèbre multidimensionnelle. Ce modèle permet d'associer à chaque mesure plusieurs fonctions d'agrégation en fonction des dimensions, des hiérarchies et des paramètres. Afin d'adapter la visualisation de la TM pour présenter plusieurs fonctions d'agrégation, la TM simplifie l'illustration des fonctions en termes de contraintes d'agrégation, d'ordres d'exécution et de nombre des fonctions affichées. Les fonctions d'agrégation non-commutatives nécessitent l'utilisation de l'ordre d'exécution. Cela peut influencer le mécanisme interne des opérateurs OLAP surtout Rotate et RollUp.

Nous envisageons de poursuivre nos travaux en optimisant les algorithmes des opérateurs étendus et en effectuant des expérimentations basées sur un prototype.

Références

- Abelló, A., J. Samos, et F. Saltor (2006). Yam2 : A multidimensional conceptual model extending uml. *Information Systems* 31, 541–567.
- Bouilil, K., S. Bimonte, et F. Pinet (2011). Un modèle uml et des contraintes ocl pour les entrepôts de données spatiales. de la représentation conceptuelle à l’implémentation. *Ingénierie des Systèmes d’Information (ISI)* 16(6), 11–39.
- Codd, E. (1993). Providing olap (on-line analytical processing) to user analysts : an it mandate. *Technical Report, E.F. Codd and Associates*.
- Golfarelli, M., D. Maio, et S. Rizzi (1998). Conceptual design of data warehouses from e/r schemes. *Intl. Conf. HICSS’98* 7, 334–343.
- Gray, J., A. Bosworth, A. Layman, et H. Pirahesh (1996). Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-total. *Intl. Conf. ICDE* 96, 152–159.
- Gyssens, M. et L. V. S. Lakshmanan (1997). A foundation for multi-dimensional databases. *Intl. Conf. VLDB* 97, 106–115.
- Hassan, A., F. Ravat, O. Teste, R. Tournier, et G. Zurfluh (2012a). Differentiated multiple aggregations in multidimensional databases. *DAWAK LNCS 7448*, 93–104.
- Hassan, A., F. Ravat, O. Teste, R. Tournier, et G. Zurfluh (2012b). Modélisation des bases de données multidimensionnelles à agrégations multiples et différenciées. *EDA*, 57–71.
- Pedersen, T., C. Jensen, et C. Dyreson (2001). A foundation for capturing and querying complex multidimensional data. *Information Systems* 26, n 5, 383–423.
- Prat, N. and Wattiau, I. et J. Akoka (2010). Representation of aggregation knowledge in olap systems. *The 18th European Conference on Information Systems (ECIS)*.
- Ravat, F., O. Teste, R. Tournier, et G. Zurfluh (2008). Algebraic and graphic languages for olap manipulations. *Intl. Journal of Data Warehousing and Mining, IGI Publishing* 4(1), 17–46.
- Ravat, F., O. Teste, et G. Zurfluh (2001). Modélisation multidimensionnelle des systèmes décisionnels. *1ère journée Extraction et Gestion des Connaissances (EGC) 1*, 201–212.
- Vassiliadis, P. et S. Skiadopoulos (2000). Modelling and optimisation issues for multidimensional databases. *Intl. Conf CAISE 2000 LNCS 1789*, 482–497.

Summary

Our conceptual model for multidimensional databases allows associating each measure to several aggregation functions according to dimensions, hierarchies, and levels of granularity. This article studies the impacts of this model on the multidimensional table MT and OLAP algebra. Because of the characteristics of this model, there are changes that occur on the internal algorithms of some OLAP operators. Changes may affect the MT. They may require adapting the MT to presenting several aggregation functions.