
Editorial

Les entrepôts de données et les bases multidimensionnelles sont maintenant nombreux dans les entreprises. Ils sont « supportés » par les outils du marché tels qu'Oracle Express, Business Objects ou encore Impromptu de Cognos pour ne citer que les produits leaders. Toutefois, la recherche ne s'épuise pas dans ce domaine dans la mesure où de nombreux aspects restent insuffisamment couverts par ces produits. Ce numéro spécial a pour objectif d'explicitier quelques-uns de ces problèmes et de proposer des solutions au moins partielles et, en tout cas, innovantes.

Le premier article de F. Ravat, O. Teste et G. Zurfluh décrit un langage pour les bases multidimensionnelles. Ce langage, inspiré de SQL et qui va au-delà des extensions proposées par Oracle, est à la fois langage de définition de données (LDD), langage de manipulation des données (LMD) et langage de définition des droits pour les utilisateurs. Le langage de définition de données permet la création, la modification et la suppression des dimensions, des faits et des hiérarchies du schéma multidimensionnel. Le langage de manipulation de données permet l'interrogation et la mise à jour des données dans les tables de dimensions et dans les tables de faits. La commande d'interrogation SELECT est enrichie d'une clause <according to> qui permet de définir les paramètres servant au calcul des mesures affichées par la requête. Enfin, le langage de définition des droits permet d'autoriser ou non un utilisateur ou un groupe d'utilisateurs à accéder à une table de faits, de dimension ou à une hiérarchie. A noter que l'allocation d'un droit sur un composant donne les mêmes droits sur les éléments le composant. Un droit sur un fait assure un droit sur les dimensions et les hiérarchies associées.

Le deuxième article, de M. Bouzeghoub, Z. Kedad et A. Soukane, aborde une seconde classe de problèmes de toute importance, l'alimentation des entrepôts. Ce processus nécessite, dans la plupart des cas, la mise en place d'outils de médiation permettant de rendre cohérentes toutes les données issues de sources hétérogènes. On appelle nettoyage les opérations de transformations des données qui rendent effective cette cohérence. L'article décrit une approche consistant à construire un schéma global, ou schéma de médiation. Les utilisateurs expriment leurs besoins

sous forme de requêtes interrogeant le schéma global. Ces requêtes contiennent des règles de nettoyage et de transformation nécessaires à l'obtention de réponses satisfaisantes. Pour ce faire, les auteurs ont choisi de définir des opérateurs algébriques qui intègrent des fonctions de transformation d'attributs ou de similarité. Par exemple, un opérateur de sélection approximative filtre les données d'une table en appliquant une fonction aux attributs concernés. Une jointure approximative étend la notion de jointure au cas où les domaines de valeurs ne sont pas compatibles mais peuvent être liés par une fonction d'appariement.

Le troisième article, de S. Goutier, G. Hébrail et V. Stéphan, aborde un autre problème lié à la qualité de l'alimentation de l'entrepôt de données, celui des données manquantes. Deux cas sont considérés : celui des lignes manquantes dans une table de faits et celui des valeurs manquantes dans une table de dimension. La solution commune proposée pour ces deux problèmes consiste à affecter chaque type de la table de faits d'un poids calculé soit à partir de données externes, soit à partir des données non manquantes. La méthode proposée utilise des techniques classiques d'échantillonnage mais ces techniques sont adaptées de façon à permettre, d'une part, de s'assurer de la robustesse des réponses (si l'échantillonnage n'est pas statistiquement correct, la requête ne fournit pas un résultat fiable) et, d'autre part, de minimiser le nombre de calculs de poids. Pour ce faire, un algorithme permet de définir et de stocker « a priori » quelques systèmes de poids dont l'utilisation est supposée fréquente.

Enfin, le quatrième article de M. Miquel, Y. Bédard et A. Brisebois présente trois originalités principales. D'abord, il traite d'un type d'entrepôt spécifique, à savoir les entrepôts de données géospatiales, en abordant les problèmes induits par la multiplicité des dimensions spatiales qu'il faut prendre en compte quand on souhaite représenter dans un même entrepôt des données spatiales de différentes natures. De plus, il traite du problème des données temporelles qui doivent être consolidées dans une même structure, même si elles qualifient des durées ou des échelles de temps très différentes. Enfin, l'ensemble de ces aspects est illustré à l'aide d'une application réelle qui a le mérite de montrer de façon concrète la nature des problèmes de compatibilité des données dans un entrepôt de données géospatiales.

Nous espérons que vous prendrez plaisir à lire ces articles et remercions leurs auteurs, mais aussi les lecteurs qui ont travaillé avec sérieux et dévouement pour assurer la qualité de ce numéro spécial.

Jacky Akoka
CNAM et INT

Isabelle Comyn-Wattiau
CNAM et ESSEC

COMITE DE LECTURE

Michel Adiba
Université Joseph Fourier de Grenoble, France

Mokrane Bouzeghoub
Université de Versailles Saint Quentin, France

Christine Collet
ENSIMAG et INPG, France

Mathias Jarke
Aachen University, Allemagne

Manfred Jeusfeld
Tilburg University, Pays-Bas

Elisabeth Métais
Conservatoire National des Arts et Métiers, France

Nicolas Prat
Ecole Supérieure des Sciences Economiques et Commerciales, France

Il-Yeol Song
Drexel University, Etats-Unis

Gilles Zurfluh
Université de Toulouse 1, France

RELECTEURS ADDITIONNELS

Edgard Benitez-Guerrero
Laboratoire IMAG, Grenoble, France

Agnès Conte
Université Pierre Mendès-France, Grenoble, France

Stéphane Lopes
Université de Versailles Saint Quentin, France