

Vers l'auto-administration des entrepôts de données

Kamel Aouiche*, Jérôme Darmont*
Le Gruenwald**

*Laboratoire ERIC, Équipe BDD
Université Lumière – Lyon 2
5 avenue Pierre Mendès-France
69676 BRON Cedex
{kaouiche, jdarmont}@eric.univ-lyon2.fr
**School of Computer Science
University of Oklahoma
Norman, OK 73019, USA
ggruenwald@ou.edu

Résumé. Avec le développement des bases de données en général et des entrepôts de données (*data warehouses*) en particulier, il est devenu primordial de réduire la fonction d'administration de base de données. L'idée d'utiliser des techniques de fouille de données (*data mining*) pour extraire des connaissances utiles des données elles-mêmes pour leur administration est avancée depuis quelques années. Pourtant, peu de travaux de recherche ont été entrepris dans ce domaine. L'objectif de cette étude est de rechercher une façon d'extraire, à partir des données stockées, des connaissances utilisables pour appliquer de manière automatique des techniques d'optimisation des performances, et plus particulièrement d'indexation. Nous avons réalisé un outil qui effectue une recherche de motifs fréquents sur une charge donnée afin de calculer une configuration d'index permettant d'optimiser le temps d'accès aux données. Les expérimentations que nous avons menées ont montré que les configurations d'index générées par notre outil permettent des gains de performance de l'ordre de 15% à 25% sur une base et un entrepôt de données tests.

1 Introduction

L'utilisation courante de bases de données requiert un administrateur qui a pour rôle principal la gestion des données au niveau logique (définition de schéma) et physique (fichiers et disques de stockage), ainsi que l'optimisation des performances de l'accès aux données. Avec le déploiement à grande échelle des systèmes de gestion de bases de données (SGBD), minimiser la fonction, d'administration est devenu indispensable (Weikum et al. 2002).

L'une des tâches importantes d'un administrateur est la sélection d'une structure physique appropriée pouvant améliorer les performances du système en minimisant les temps d'accès aux données (Finkelstein et al. 1988). Les index sont des structures physiques permettant un accès direct aux données. Le travail d'optimisation des performances de l'administrateur se porte en grande partie sur la sélection d'index et de