

ParAdmin: Un Outil d'Assistance à l'Administration et Tuning d'un Entrepôt de Données

Ladjet Bellatreche *, Kamel Boukhalfa *, Sybille Caffiau * **

* LISI/ENSMA - Université de Poitiers - Futuroscope
(bellatreche, boukhalk, caffiaus)@ensma.fr,

** INRIA Domaine de Volvecan, Rocquencourt

Résumé. Les entrepôts de données ont rendu les tâches d'administration et de tuning plus complexes que dans les bases de données traditionnelles. Cela est dû aux caractéristiques des entrepôts de données : la volumétrie, les requêtes complexes, les délais de réponse exigés par les décideurs et la gestion de l'évolution. Dans ce contexte, une panoplie de techniques d'optimisation ont été proposées durant la phase de conception physique. Pour chacune d'entre elles un nombre important d'algorithmes de sélection existe, chacun ayant ses propres paramètres. Un autre choix déterminant est à faire entre les techniques d'optimisation similaires - pour optimiser une requête donnée, deux techniques peuvent être candidates. Durant la tâche de conception physique, l'administrateur doit donc effectuer de multiple choix. Dans ce papier, nous montrons d'abord les difficultés qu'un administrateur peut rencontrer durant la phase de conception physique. Deuxièmement, nous présentons une méthode de tuning basée sur la fragmentation horizontale et les index de jointure binaires. Finalement, nous proposons un outil d'administration, appelé *ParAdmin*, qui permet d'assister l'administrateur en proposant la réalisation interactive des tâches de conception physique et de tuning. Elles sont définies à l'aide des choix : de la ou des structures d'optimisation, des algorithmes de sélection, des paramètres, de tuner ou non et de visualiser les recommandations de performance.

1 Introduction

Les entrepôts de données ont largement contribué à l'évolution de la conception physique traditionnelle dédiée aux applications de type OLTP, où la tâche d'un administrateur était concentrée sur la gestion des utilisateurs et d'un nombre restreint de techniques d'optimisation (les index, les différentes implémentations de jointure - boucles imbriquées, sort merge, hash, etc.). L'apparition des entrepôts de données a rendu la tâche d'administration de plus en plus complexe. Cette complexité est due à leurs caractéristiques. L'une d'entre elles est le volume de données. Par exemple, le schéma en étoile de l'entrepôt de données relationnel de la compagnie *General Motors Corporation* dépasse 1.5 téraoctets et la table des faits a plus de deux milliard d'instances (Gill, 2000). De plus, les requêtes décisionnelles sont complexes car elles contiennent des jointures, des sélections et des agrégations. De même, il est impératif d'exécuter ces requêtes en un temps raisonnable. Enfin, les entrepôts de données évoluent