

Vers une Conception des Entrepôts de Données Parallèles Autonomes

Soumia Benkrid*, Ladjel Bellatreche**

* Ecole nationale Supérieure d'Informatique (ESI), Alger, Algérie
s_benkrid@esi.dz

**LIAS/ISAE-ENSMA, Futuroscope, Poitiers, France
bellatreche@ensma.fr

Résumé. Les systèmes de stockage de données parallèles (SSDPs) sont devenus une des solutions incontournables pour traiter les données massives pour des fins d'analyse. L'efficacité de ces systèmes dépend fortement des processus de fragmentation et d'allocation des partitions sur l'ensemble des nœuds. En examinant les travaux existants, nous avons constaté qu'ils se basent sur les requêtes, une chose usuelle pour toute méthodologie de conception des systèmes complexes. Les requêtes d'entrée dans le contexte de SSDPs sont souvent connues à l'avance (et statiques) et peuvent évoluer. Les approches utilisées pour concevoir des SSDP sont alors réactives. Dans la BI 2.0, où les utilisateurs (décideurs) sont au centre du système avec *leurs requêtes en lots et ad-hoc*, les approches réactives peuvent facilement montrer leur limite. Pour faire face à cette situation, le recours aux techniques issues des systèmes autonomes *intelligents* (comme l'informatique proactive) est nécessaire. Dans cet article, nous proposons une approche proactive de conception des SBDPs comportant deux phases principales : (i) une phase hors ligne qui génère des schémas de fragmentation et d'allocation à partir des requêtes supposées connues et vues comme une base d'apprentissage et (ii) une phase en ligne qui augmente les schémas obtenus par la phase hors ligne afin de prendre en compte les nouvelles requêtes ad-hoc par lots. La prise en charge de ces requêtes est assurée par l'exploitation des résultats intermédiaires identifiés dans la phase hors ligne. Les résultats de nos expérimentations montrent clairement l'intérêt de l'informatique proactive pour la conception des SSDPs autonomes.

1 Introduction

La Business Intelligence (BI) traditionnelle a été souvent construite pour offrir des mécanismes d'exploitation des données historisées d'entreprises à l'aide des requêtes supposées connues à l'avance. La connaissance préalable des requêtes a été largement utilisée pour déployer (Stöhr et al., 2000), optimiser (Bellatreche, 2018) et exploiter (Gallinucci et al., 2018) des entrepôts de données (\mathcal{ED}) de la première génération. La large considération de cette hypothèse par les chercheurs est souvent motivée par la complexité des problèmes associés au