

L'implantation de sources de données dans un système NoSQL : formalisation des règles de passage conceptuel/logique

Fatma Abdelhedi*,**, Amal Ait Brahim* et Gilles Zurfluh*

* IRIT - Université Toulouse Capitole - France -
<prenom.nom>@irit.fr

** CBI² - Sté TRIMANE Saint Germain en Laye - France -
<prenom.nom>@gmail.com

Résumé. La transformation digitale des entreprises et plus largement celle de la société, entraîne une évolution des bases de données vers le Big Data. Nos travaux s'inscrivent dans cette mutation et concernent plus particulièrement les mécanismes d'implantation d'une base de données sur une plateforme NoSQL. Pour automatiser ce processus, nous avons spécifié des algorithmes pour traduire un schéma conceptuel en un schéma logique NoSQL. A partir d'un diagramme de classes d'UML décrivant une base d'objets complexes, nous proposons des procédures de correspondance pour générer un schéma d'implantation destiné à une plateforme NoSQL orientée colonnes. Nous introduisons un schéma intermédiaire de niveau logique afin de limiter les impacts liés aux évolutions techniques des plateformes NoSQL. Une expérimentation du processus de correspondance a été réalisée sur une application médicale.

1 Introduction

Les applications Big Data, Chen et Zhang (2014), développées dans les domaines tels que le spatial, la santé ou la gestion commerciale, répondent à un double objectif : (1) assurer le passage à l'échelle (ou scalabilité), c'est-à-dire répartir les données et distribuer les traitements sur un nombre important de machines afin d'être en mesure de stocker de très grands volumes de données et d'absorber des charges très importantes ; (2) manipuler les données complexes avec des outils qui prennent en compte la répartition logique de ces données.

Notre problématique générale est de proposer des modèles et des outils décisionnels capables de localiser des sources de données pertinentes, d'alimenter des entrepôts et d'exploiter ces derniers à des fins d'analyse. Les présents travaux se situent en amont dans ce processus décisionnel, au niveau des sources de données utilisées pour alimenter les entrepôts. Avant l'avènement du Big Data, ces sources de données étaient principalement constituées de bases de données relationnelles, de fichiers informatiques et de documents formatés en HTML ou XML. Avec la diffusion des plateformes NoSQL, les systèmes de décision doivent intégrer de nouvelles sources de données pour alimenter les entrepôts. Ainsi, des systèmes d'alimentation