

Approche dirigée par les modèles pour l'extraction automatique du modèle NoSQL

Amal Ait Brahim *, Rabah Tighilt Ferhat * et Gilles Zurfluh *

*Institut de Recherche en Informatique de Toulouse (IRIT)
Université Toulouse Capitole - 2 Rue du Doyen-Gabriel-Marty - 31042 Toulouse - France
<prenom.nom>@gmail.com
<http://www.ut-capitole.fr/>

Résumé. Les systèmes NoSQL permettent de gérer des bases de données (BD) massives vérifiant les 3V : Volume, Variété et Vitesse. La plupart de ces systèmes sont caractérisés par la propriété « schema less » qui signifie absence du modèle de données lors de la création d'une BD. Cette propriété apporte une souplesse indéniable en permettant l'évolution du modèle pendant l'exploitation de la base. Cependant, l'expression des requêtes d'interrogation (type SQL) exige une connaissance précise de ce modèle. Dans cet article, nous proposons un processus pour l'extraction automatique du modèle physique d'une BD NoSQL de type documents. Pour ce faire, nous utilisons l'architecture MDA (Model Driven Architecture) qui fournit un cadre formel pour la transformation automatique de modèles. À partir d'une BD NoSQL, nous proposons des règles de transformation formalisées avec QVT pour générer le modèle physique. Une expérimentation du processus d'extraction a été réalisée sur une application médicale.

1 Introduction

Ces dernières années ont vu l'explosion des données générées et accumulées par les dispositifs informatiques de plus en plus nombreux et diversifiés. Les BD ainsi constituées sont désignées par l'expression « Big Data » et sont caractérisées par la règle dite des « 3V » Chen et Zhang (2014). Celle-ci est due au volume des données qui peut dépasser plusieurs téraoctets et à la variété qui correspond au stockage de données de types et de formats diversifiés tels que des textes, des tableaux, des documents fortement structurés. De plus, ces données sont souvent saisies à très haute fréquence et doivent donc être filtrées et agrégées en temps réel pour éviter une saturation inutile de l'espace de stockage.

Les techniques d'implantation classiques, basées principalement sur le paradigme relationnel, connaissent des limites pour gérer les BD massives Angadi et al. (2013). Ainsi, de nouveaux systèmes de stockage et de manipulation des données ont été développés. Regroupés sous le terme NoSQL Han et al. (2011), ces systèmes sont bien adaptés pour gérer de gros volumes de données dont les modèles sont flexibles. Ils apportent aussi de grandes capacités de passage à l'échelle et de bonnes performances en temps de réponse Angadi et al. (2013).

La plupart des SGBD NoSQL sont caractérisés par la propriété « schema less » qui correspond à l'absence du schéma de données lors de la création d'une BD. Cette propriété apparaît