Data schema does matter, even in NoSQL Systems!

De l'importance d'un schéma de données dans les systèmes NoSQL

Paola Gómez[1], Rubby Casallas[2], Claudia Roncancio[1]

1. LIG, Université Grenoble Alpes
{paola.gomez-barreto,claudia.roncancio}@imag.fr

2. TICSw, Universidad de los Andes, Bogotá, Colombia
rcasalla@uniandes.edu.co

Article accepté et présenté à la conférence internationale RCIS, co-localisée avec INFORSID 2016 à Grenoble. La version longue de l'article, en anglais, est disponible dans les actes de RCIS.

RESUME. *La plupart de systèmes « NoSQL » n'utilisent pas de schéma de bases de données. Ces systèmes n'offrent donc pas de fonctions de gestion de schéma et la définition de la gestion des structures de données utilisées reviennent à l'application. Les données peuvent être semi-structurées et l'utilisateur a de nombreuses possibilités de structuration.*
L'absence de schéma prédéfinie présente certainement des avantages de flexibilité. Cependant, ceci a aussi un coût en termes de volume des données stockées, de performances des requêtes, de clarté du code applicatif et, a fortiori, la mise au point et la maintenance des applications. Ceci est démontré dans cet article où nous présentons une étude de l'impact de la structuration des données dans la base de documents MongoDB.
Nous avons expérimenté avec une variété d'alternatives de structuration et une série de requêtes avec complexité croissante. Cet article présente notre analyse et conclusions suite à cette expérimentation.

ABSTRACT. *A Schema-less NoSQL system refers to solutions where users do not declare a database schema and, in fact, its management is moved to the application code. This paper presents a study that allows us to evaluate, to some extent, the data structuring impact.*

The decision of how to structure data in semi-structured databases has an enormous impact on data size, query performance and readability of the code, which influences software debugging and maintainability. This paper presents an experiment performed using MongoDB along with several alternatives of data structuring and a set of queries having

increasing complexity. This paper introduces an analysis regarding the findings of such an experiment.

MOTS-CLES : NoSQL, systèmes orientés document, MongoDB.

KEYWORDS: NoSQL, schema-less systems, document-oriented systems, MongoDB.

DOI:10.3199/JESA.45.1-n © Lavoisier 2012 AR_DOI