
Document-oriented data warehouses: models and extended cuboids

Modèle orienté-document et cuboïdes étendus

**Max Chevalier¹, Mohammed El Malki^{1,2} Arlind Kopliku,
Olivier Teste, Ronan Tournier**

1. Université de Toulouse, IRIT (UMR 5505, www.irit.fr), Toulouse, France
prenom.nom@irit.fr

2. Capgemini (www.capgemini.com) Toulouse France
prenom.nom@capgemini.com

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

ABSTRACT. Within the Big Data trend, there is an increasing interest in Not-only-SQL systems (NoSQL). These systems are promising candidates for implementing data warehouses particularly due to the data structuration/storage possibilities they offer. In this paper, we investigate data warehouse instantiation using a document-oriented system (a special class of NoSQL systems). On the one hand, we analyze several issues including modeling, querying, loading data and OLAP cuboids. We compare document-oriented models (with and without normalization) to analogous relational database models. On the other hand, we suggest improvements in order to benefit from document-oriented features. We focus particularly on extended versions of OLAP cuboids that exploit nesting and arrays. They are shown to work better on workloads with drill-down queries.

KEYWORDS: NoSQL, document-oriented system, big data warehouse, OLAP cuboid.

RESUME: Parmi les nouvelles technologies du Big Data, il y a un intérêt croissant pour les systèmes Not-Only-SQL (NoSQL). Ces derniers sont des candidats prometteurs pour l'implantation d'entrepôts de données grâce aux possibilités qu'ils offrent en termes de stockage et de structuration des données. Dans ce papier, nous étudions l'instanciation d'entrepôts de données avec les systèmes orientés document, une des catégories les plus répandues des systèmes NoSQL. Dans un premier temps, nous étudions les enjeux primaires des entrepôts tels que la modélisation, l'interrogation, le chargement des données, et les cuboïdes OLAP. Dans un deuxième temps, nous proposons des améliorations qui sont spécifiques aux systèmes orientés document. En particulier, nous proposons des versions étendues des cuboïdes OLAP qui exploitent l'imbrication et les tableaux. Nous montrons que ces cuboïdes répondent plus rapidement à des charges de travail composées de requêtes OLAP de type "drill-down".

MOTS CLES: NoSQL, système orienté-document, entrepôts de données big data, cuboïde OLAP