On The Potential Integration of an Ontology-Based Data Access Approach in NoSQL Stores

Oliver Curé, LIGM, Université Paris-Est Marne-la-Vallée, Marne-la-Vallée, France
Fadhela Kerdjoudj, LIGM, Université Paris-Est Marne-la-Vallée, Marne-la-Vallée, France
David Faye, Computer Science Department, Faculty of Sciences and Applied Technology, Université Gaston Bérger de Saint Louis, Saint Louis, Sénégal
Chan Le Duc, LIASD, Université Paris 8, Paris, France
Myriam Lamolle, LIASD, Université Paris 8, Paris, France

ABSTRACT

NoSQL stores are emerging as an efficient alternative to relational database management systems in the context of big data. Many actors in this domain consider that to gain a wider adoption, several extensions have to be integrated. Some of them focus on the ways of proposing more schemas, supporting adapted declarative query languages and providing integrity constraints in order to control data consistency and enhance data quality. The authors consider that these issues can be dealt with in the context of Ontology Based Data Access (OBDA). OBDA is a new data management paradigm that exploits the semantic knowledge represented in ontologies when querying data stored in a database. They provide a proof of concept of OBDA’s ability to tackle these three issues in a social application related to the medical domain.

Keywords: Document Store, NoSQL, Ontology Based Data Access (OBDA), Relational Database Management System (RDBMS), Social Application, SPARQL Protocol and RDF Query Language (SPARQL)

INTRODUCTION

NoSQL (Sadalage & Fowler, 2012) covers a wide range of technologies and data architectures for managing web-scale data and having the following common features: persistent data, non-relational model, avoid join operations, data distribution, massive horizontal scaling, no fixed and flexible schemata, replication support, vendor-specific and usually procedural query systems rather than using a standard declarative query language, consistent within a node of the cluster and eventually consistent across the cluster and simple transactions, i.e., not
ACID (Atomicity, Consistency, Isolation and Durability properties expected from a Relational Data Base Management System - RDBMS).

According to their data model and replication/sharding strategy, we distinguish four NoSQL categories, each one having its own specificities and facilitating the management of some particular kind of data: view of a database as something for storing a value (Key-value Stores), more flexibility about stored data (Document Stores), management of use cases like relationships (Graph Databases) or aggregation of data (Column Databases).

Solutions in the NoSQL ecosystem are emerging in various domains such as social, scientific and even financial applications. Nevertheless, many actors consider that in order to increase its adoption rate, NoSQL systems need to integrate some new features. In fact, the desired features correspond to the ones found in Relational Database Management Systems (RDBMS). We can identify three important ones which are concerned with more schema, more declarative query languages and more data integrity to enhance data quality and business intelligence (BI) processing. In fact, excluding ACID transactions and consistency issues (Vogels, 2009), after these additions a NoSQL system may start resembling a RDBMS. We argue that the integration of these features needs to consider the semantics of the elements of the application domain. This could be a major breakthrough for both NoSQL stores and the Semantic community since RDBMS is not really reactive in integrating semantics.

Ontology Based Data Access (henceforth OBDA) may be a good fit in this direction since it aims to represent the concepts and properties of a domain with a formalized ontology. OBDA provides a semantic conceptual schema over a repository of data and, due to its logical formalism, it is likely to support formal analysis, optimization and reasoning. In this paper, we focus on the currently most popular form of OBDA systems: those based on Description Logics (DL) (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2003). DL-based OBDA is largely motivated by the Semantic Web and has mainly been studied for data repositories corresponding to RDBMS.

The main contribution of this work is to show that OBDA is even more needed in the NoSQL ecosystem. Moreover, we consider that a common OBDA approach can be designed for both RDBMS and NoSQL, hence supporting a form of data integration from both these data management systems.

This paper is an extended version of Curé, Kerdjoudj, Faye, Le Duc, and Lamolle (2012), i.e., it is self-contained by presenting DL-based OBDA and NoSQL in details, and is organized as follows. In Section 2, we present the background knowledge on OBDA and NoSQL. Section 3 introduces a social medical application that will serve as a running example. Section 4 tackles issues on the three features identified for NoSQL systems: schema modeling, declarative language and constraint violation detection. Section 5 contains a discussion and concludes the paper.

BACKGROUND KNOWLEDGE

In this section, we introduce the main notions needed to understand the concepts used in this paper.

Basically, we present the main characteristics of DLs and in particular the DLs that are used by OBDA in the context of the Semantic Web. Then, we present some of the most popular NoSQL data models, i.e. document and column family stores.

DL-Based OBDA

DLs correspond to a fragment of first order logic with sound and complete inference procedures. They are generally used to represent the knowledge of a particular application domain and are composed of a Terminological Box (TBox) and an Assertional Box (ABox) that respectively specify the general properties
Bio-Inspired Techniques for Resources State Prediction in Large Scale Distributed Systems
Andreea Visan, Mihai Istin, Florin Pop and Valentin Cristea (2013). Development of Distributed Systems from Design to Application and Maintenance (pp. 18-36).
www.igi-global.com/chapter/bio-inspired-techniques-resources-state/72244?camid=4v1a

Toward an Access Infrastructure for Mobile Cloud: A Channel Assignment Scheme for Wireless Mesh Networks
www.igi-global.com/article/toward-an-access-infrastructure-for-mobile-cloud/95115?camid=4v1a