

# OCL Constraints Checking on NoSQL Systems Through an MDA-Based Approach

Fatma Abdelhedi, CBI2 – TRIMANE, France

Amal Ait Brahim, Toulouse Institute of Computer Science Research (IRIT), Toulouse Capitole University, Toulouse, France

Gilles Zurfluh, Toulouse Institute of Computer Science Research (IRIT), Toulouse Capitole University, Toulouse, France

## ABSTRACT

Big data have received a great deal of attention in recent years. Not only is the amount of data on a completely different level than before, but also the authors have different type of data including factors such as format, structure, and sources. This has definitely changed the tools one needs to handle big data, giving rise to NoSQL systems. While NoSQL systems have proven their efficiency to handle big data, it's still an unsolved problem how the automatic storage of big data in NoSQL systems could be done. This paper proposes an automatic approach for implementing UML conceptual models in NoSQL systems, including the mapping of the associated OCL constraints to the code required for checking them. In order to demonstrate the practical applicability of the work, this paper has realized it in a tool supporting four fundamental OCL expressions: iterate-based expressions, OCL predefined operations, If expression, and Let expression.

## KEYWORDS

Big Data, Conceptual Model, Data Base, Integrity Constraints, MDA, NoSQL, OCL, QVT, Schema Transformations, Schema-Less, UML

## INTRODUCTION

Big data have received a great deal of attention in recent years. Not only the amount of data is on a completely different level than before, but also we have different type of data including factors such as format, structure, and sources. In addition, the speed at which these data must be collected and analyzed is increasing. This has definitely impacted the tools required to store Big Data, and new kinds of data management tools i.e. NoSQL systems have arisen (Cattell, 2011). Compared to existing systems, NoSQL systems are commonly accepted to support larger volume of data, provide faster data access, better scalability and higher flexibility (Angadi et al., 2013).

One of the NoSQL key features is that databases can be schema-less. This means, in a table, meanwhile the row is inserted, the attributes names and types are specified. Unlike relational systems where first the user defines the schema and creates the tables, second he inserts data, the schema-less property offers undeniable flexibility that facilitates the physical schema evolution. End-users are able to add information without the need of database administrator. For instance, in the medical program

DOI: 10.4018/IJDWM.2021010101

that follows-up patients suffering from a chronic pathology – case of study detailed in ILLUSTRATIVE EXAMPLE Section – one of the benefits of using NoSQL databases is that the evolution of the data (and schema) is fluent. In order to follow the evolution of the pathology, information is entered regularly for a cohort of patients. But the situation of a patient can evolve rapidly which needs the recording of new information. Thus, few months later, each patient will have his own information, and that's how data will evolve over time. Therefore, the data model (i) differs from one patient to another and (ii) evolves in unpredictable way over time. We should highlight that this flexibility concerns the physical level i.e. the stored database exclusively (Herrero et al., 2016).

In information systems, the importance and the necessity of conceptual models are widely recognized. The conceptual model provides a high level of abstraction and a semantic knowledge element close to human comprehension, which guarantees efficient data management (Abelló, 2015). Furthermore, this model is a document of interchange between end-users and designers, and between designers and developers. Also, the conceptual model is used for system maintenance and evolution that can affect business needs and/or deployment platform. The Unified Modeling Language (UML) is widely accepted as the standard of information system modeling.

On the one hand, NoSQL systems have proven their efficiency to handle Big Data. On the other hand, the needs of a conceptual modeling and design approach remain up-to-date. Therefore, we are convinced that it's important to provide a precise and automatic approach that guides and facilitates the Big Database implementation task within NoSQL systems. This approach will assist the developers to map Big Database UML conceptual model into NoSQL physical models. It's also required to have a tool to maintain data consistency since most of the NoSQL systems lacks of constraints checking and enforcement mechanism.

For this, we propose the “Object2NoSQL” MDA-based approach. The Model Driven Architecture (MDA) is well-known as a framework for models automatic transformations. The Object2NoSQL approach starts from a conceptual model (PIM) (UML class diagram and OCL constraints) and transforms it into a unified logical model compatible with the four types of NoSQL database (column, document, graph and key-value). The conceptual model is automatically transformed into a logical model using QVT rules. Then, logical model is transformed into a physical model (PSM after choosing one of the four platforms: Cassandra, MongoDB, Neo4J or Redis). In this paper, we focus on how to automatically transform UML/OCL conceptual model into NoSQL physical level. As discussed in the related work, few solutions have dealt with the NoSQL databases conceptual modeling. To the best of our knowledge, none of the existing contribution has treated the OCL constraints and their implementation into NoSQL databases.

The remainder of the paper is structured as follows. ILLUSTRATIVE EXAMPLE Section motivates our work using a case of study in the healthcare field. OBJECT2NOSQL APPROACH Section defines our models transformation approach. OCL2JAVA APPROACH Section introduces our OCL constraints transformation approach and shows two transformation processes. The first one creates a logical model starting from an OCL conceptual model, and the second one generates the java code required to check the constraints within NoSQL database. RELATED WORK Section reviews previous work. Finally, CONCLUSION Section summarizes the objectives achieved and announces future work.

## ILLUSTRATIVE EXAMPLE

To motivate and to illustrate our work, we present a case study in the healthcare filed. This case study concerns international scientific programs for monitoring patients suffering from serious diseases. The main goal of this program is (1) to collect data about diseases development over time, (2) to study interactions between different diseases and (3) to evaluate the short and medium-term effects of their treatments. The medical program can last up to 3 years. Data collected from establishments involved in this kind of program have the features of Big Data (the 3 V):

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/ocl-constraints-checking-on-nosql-systems-through-an-mda-based-approach/272015](http://www.igi-global.com/article/ocl-constraints-checking-on-nosql-systems-through-an-mda-based-approach/272015)

## Related Content

---

### Large-Scale System for Social Media Data Warehousing: The Case of Twitter-Related Drug Abuse Events Integration

Jenhani Ferdaous and Mohamed Salah Gouider (2022). *International Journal of Data Warehousing and Mining* (pp. 1-18).

[www.irma-international.org/article/large-scale-system-for-social-media-data-warehousing/290890](http://www.irma-international.org/article/large-scale-system-for-social-media-data-warehousing/290890)

### The Impact of Missing Data on Data Mining

Marvin L. Brown and John F. Kros (2003). *Data Mining: Opportunities and Challenges* (pp. 174-198).

[www.irma-international.org/chapter/impact-missing-data-data-mining/7600](http://www.irma-international.org/chapter/impact-missing-data-data-mining/7600)

### Parallel Computing for Mining Association Rules in Distributed P2P Networks

Huiwei Guan (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 107-124).

[www.irma-international.org/chapter/parallel-computing-mining-association-rules/73436](http://www.irma-international.org/chapter/parallel-computing-mining-association-rules/73436)

### A Survey on Fuzzy Association Rule Mining

Harihar Kalia, Satchidananda Dehuri and Ashish Ghosh (2013). *International Journal of Data Warehousing and Mining* (pp. 1-27).

[www.irma-international.org/article/survey-fuzzy-association-rule-mining/75613](http://www.irma-international.org/article/survey-fuzzy-association-rule-mining/75613)

### A Brief Study of Approaches to Text Feature Selection

Ravindra Babu Tallamaraju and Manas Kirti (2018). *Modern Technologies for Big Data Classification and Clustering* (pp. 216-243).

[www.irma-international.org/chapter/a-brief-study-of-approaches-to-text-feature-selection/185985](http://www.irma-international.org/chapter/a-brief-study-of-approaches-to-text-feature-selection/185985)