Multi-model Query Processing Meets Category Theory and Functional Programming

Valter Uotila Jiaheng Lu University of Helsinki first.last@helsinki.fi Dieter Gawlick Zhen Hua Liu Souripriya Das Oracle Corporation first.last@oracle.com

Gregory Pogossiants SATS Technologies gregp_21@yahoo.com

ABSTRACT

The current multi-model database management systems (MMDBS) are becoming more complex. We propose category theory as a foundation for a new query language design, query processing, and transformation frameworks for MMDBS. We describe the recent challenges of MMDBS and represent possible solutions to them. Finally, we propose a category theory-inspired prototype system.

Reference Format:

Valter Uotila, Jiaheng Lu, Dieter Gawlick, Zhen Hua Liu, Souripriya Das, and Gregory Pogossiants. Multi-model Query Processing Meets Category Theory and Functional Programming. In the 2nd Workshop on Search, Exploration, and Analysis in Heterogeneous Datastores (SEA Data 2021).

1 INTRODUCTION

The multi-model database management systems (MMDBS) [9, 10] are gradually becoming more complex, which creates an urgent need for a better theory to formalize the systems. We identify that the end-user's experience is often poorly addressed in the design and implementation of the systems. For example, NoSQL is mainly targeted at developers. Technology is supposed to evolve according to the business and end-user's needs. Higher-level abstraction can simplify the systems and enable a better user experience.

The theory should be a standard across different domains and it should be powerful enough to express a wide variety of concepts on a suitable abstraction level. We believe that a candidate to be such a theory is category theory. Liu et al. [8] proposed this role to category theory to reason about declarative constructions and transformations between various data models. The standard introduction to category theory is MacLane [6] and other good are [11, 12].

David Spivak [13] has applied category theory to model relational databases in order to category theoretically migrate relational data. The commercial application of this category theory-based relational database framework is implemented by Conexus [1, 2].

2 CHALLENGES IN MMDBS

MMDBS is characterized by the capability to handle multiple data models against a single, unified backend. The models can include relational, graph, hierarchical, text, images, audio, video, spatial, expressions, and other complex data structures. It is required that MMDBS implement a single declarative query language that enables users to execute cross-model queries. Another wanted feature is a unified indexing mechanism that can index multiple data instances across different models. MMDBS should have the capability to perform extensive data transformations which automatically create views and materialize data between different models. Oracle converged database [3] is an example of a commercial MMDBS.

3 TOWARDS MODERN MMDBS

Historically, we had hierarchical and network data models, and then the relational data model. Now, in addition to the relational model, we have re-invented the hierarchical models as JSON/XML, and the network models as RDF and property graphs. NoSQL system complicates the matter by forcing users to access data without declarative language in a very loose transactional system. All of these efforts have regressed the usability of DBMS.

The principle of DBMS is that there is no single data model that is the best or the worst. Therefore, it is time to introduce the concept of a virtual data model. Virtual data model design is similar to the concept of virtual memory in classical OS design and virtual machine in modern cloud computing environment design.

The modern DBMS needs to follow both schema-first or schemalater paradigms and also support temporal aspects of data [4]. The temporal dimension of data is often poorly implemented in DBMS. For example, a part of temporality is event detection which could be tackled by developing calculus logic on top of queries. The modern DBMS would benefit from the unification of meta-data and data to define schema-flexible storing, indexing, and querying features [7].

4 DEMO SYSTEM AND CONCLUSION

We have developed a demonstration system called MultiCategory [15, 16] to demonstrate our solutions. The system's backend is implemented with Haskell. It offers a fold function-based query processing mechanism which is a method to model queries from a category theoretical perspective [5]. A multi-model schema is represented as a category that is mapped to the multi-model instance. Formally our approach for modeling MMDBS and data transformations using category theory is represented in [14].

Our future work includes researching data integration, migration, transformation, temporal, and virtual data model challenges using category theory. Recent progress in applied category theory has shown that category theory is a very powerful framework to model and formally define complex systems.

Copyright © 2021 for the individual papers by the papers' authors. Copyright © 2021 for the volume as a collection by its editors. This volume and its papers are published under the Creative Commons License Attribution 4.0 International (CC BY 4.0). Published in the Proceedings of the 2nd Workshop on Search, Exploration, and Analysis in Heterogeneous Datastores, co-located with VLDB 2021 (August 16-20, 2021, Copenhagen, Denmark) on CEUR-WS.org.

ACKNOWLEDGMENTS

This paper is partially supported by Finnish Academy Project 310321 and Oracle ERO gift funding.

REFERENCES

- [1] 2021. Categorical Databases. https://www.categoricaldata.net/
- [2] 2021. Conexus. https://conexus.com/
- [3] Arvind Bhope. 2021. Building a modern app with Oracle's Converged Database. https://blogs.oracle.com/database/post/building-a-modern-app-withoracles-converged-database
- [4] Dieter Gawlick. 2004. Querying the Past, the Present, and the Future. In Proceedings of the 20th International Conference on Data Engineering, ICDE 2004, 30 March 2 April 2004, Boston, MA, USA, Z. Meral Özsoyoglu and Stanley B. Zdonik (Eds.). IEEE Computer Society, 867. https://doi.org/10.1109/ICDE.2004.1320094
- [5] Torsten Grust. 2004. Monad Comprehensions: A Versatile Representation for Queries. Springer Berlin Heidelberg, Berlin, Heidelberg, 288–311. https://doi.org/10.1007/ 978-3-662-05372-0
- [6] S.M. Lane. 1998. Categories for the Working Mathematician. Springer New York, 233 Spring St, New York, NY 10013, USA.
- [7] Zhen Hua Liu and D. Gawlick. 2015. Management of Flexible Schema Data in RDBMSs - Opportunities and Limitations for NoSQL -. In CIDR.
- [8] Zhen Hua Liu, Jiaheng Lu, Dieter Gawlick, Heli Helskyaho, Gregory Pogossiants, and Zhe Wu. 2018. Multi-model Database Management Systems - A Look Forward.

In Polystores VLDB 2018 Workshops. 16-29.

- [9] Jiaheng Lu and Irena Holubová. 2019. Multi-model Databases: A New Journey to Handle the Variety of Data. ACM Comput. Surv. 52, 3 (2019), 55:1–55:38.
- [10] Jiaheng Lu, Irena Holubová, and Bogdan Cautis. 2018. Multi-Model Databases and Tightly Integrated Polystores: Current Practices, Comparisons, and Open Challenges. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (Torino, Italy) (CIKM '18). Association for Computing Machinery, New York, NY, USA, 2301–2302. https://doi.org/10.1145/3269206. 3274269
- [11] E. Riehl. 2017. Category Theory in Context. Dover Publications, 31 2nd St, Mineola, NY 11501, USA. www.math.jhu.edu/~eriehl/context.pdf
- [12] David Spivak. 2014. Category Theory for the Sciences. (2014).
 [13] David I. Spivak. 2010. Functorial Data Migration. *CoRR* abs/1009.1166 (2010). arXiv:1009.1166 http://arxiv.org/abs/1009.1166
- [14] Valter Uotila and Jiaheng Lu. 2021. A Formal Categorical Theoretical Framework for Multi-Model Data Transformation, In Poly: VLDB Workshop on Polystore Systems for Heterogeneous Data in Multiple Databases with Privacy and Security Assurances. *Poly 2021.*
- [15] Valter Uotila and Jiaheng Lu. 2021. MultiCategory demo video. https://youtu.be/ uceIi91AGsg.
- [16] Valter Uotila, Jiaheng Lu, Dieter Gawlick, Zhen Hua Liu, Souripriya Das, and Gregory Pogossiants. 2021. MultiCategory: Multi-model Query Processing Meets Category Theory and Functional Programming. Proc. VLDB Endow. 14, 2663 – 2666. Issue 12. https://doi.org/10.14778/3476311.3476314