adopted by MMDBs. Finally, we will discuss the open problems in designing query languages for MMDBs and provide insights on the research challenges and directions for future work. In addition, during this tutorial, we will invite the participants to write and run some multi-model queries by using ArangoDB to provide them hands-on experience.

The slides of this tutorial can be downloaded at this site[1]. To the best of our knowledge, this is the first tutorial to discuss state-of-the-art research works and industrial trends on multi-model data query languages. Multi-model data management and MMDBs have attracted a lot of attention during the past decades. Three existing tutorials are related to this topic. The tutorial [7] discussed the general challenges and issues in multi-model data management, and the tutorial [9] compared the two solutions for managing multi-model data, i.e., MMDBs and integrated polystores. The tutorial [6] investigated the query languages and processing paradigms for graph data. In this tutorial, we will not concentrate on the query languages for a single data model and their processing paradigms in this tutorial, which were not surveyed by previous tutorials.

2 TUTORIAL ORGANIZATION

The tutorial is planned for 6 hours and is divided into 6 parts:

Part I: Introduction(15 minutes)

We start the tutorial by introducing data variety and motivating the need of multi-model data management.

- Basics on data variety
- The need and essence of multi-model data management

Part II: Data models (45 minutes)

We will briefly discuss the major data models adopted by database systems and a benchmark for multi-model data.

- The relational model and its extensions
- The semi-structured data models, e.g. XML and JSON
- The graph data models

Part III: Multi-model data query languages (60 minutes)

We will discuss several well-known multi-model data query languages, which fall into three categories.

- The SQL-extensions
- The XML/JSON-extensions
- The graph-extensions

Part IV: Comparison of the query languages (60 minutes)

We will make a comparative study of the query languages from 4 perspectives.

- The semantic difference
- The expressive power
- The internal representation
- The manner of query evaluation

Part V: Open problem and challenges (30 minutes)

We will conclude with a discussion of open problems and challenges in designing multi-model data query languages.

- An algebra for a multi-model query language.
- General approaches for cross-model query processing.

Part VI: Hands-on experience (150 minutes)

We will invite the participants to write and run some multi-model queries by using ArangoDB.

- Generate an E-commence dataset with Unibench [15, 16]
- Hands-on experience for multi-model queries with ArangoDB.

[1] https://www.helsinki.fi/en/node/93817

3 SHORT BIBLIOGRAPHIES

Qingsong Guo is a Postdoctoral Researcher at the University of Helsinki, Finland. His current research interests include multi-model data management and automatic management of big data with deep learning algorithms.

Jiaheng Lu is an Associate Professor at the University of Helsinki. His main research interests lie in the Big Data management and database systems. He has published more than one hundred journal and conference papers. He has published several books on XML, Hadoop and NoSQL databases.

Chao Zhang is a Ph.D. candidate at the University of Helsinki. His research topic is on multi-model database benchmarking and cross-model query optimization.

Calvin Sun is the Chief Database Architect at Huawei Cloud. He has 20+ years working experience in the development of several database systems, ranging from embedded database, large-scale distributed database, to cloud-native database.

Steven Yuan is the director of Huawei Toronto Distributed Scheduling and Data Engine Lab. He leads an over 30 people research team in big data and cloud domain.

REFERENCES

[1] ECMA-404 The JSON Data Interchange Standard. https://www.json.org/json-en.html.

[2] Extensible Markup Language (XML). https://www.w3.org/XML/.

[3] R. Angles, M. Arenas, P. Barceló, A. Hogan, J. L. Reutter, and D. Vrgoc. Foundations of modern query languages for graph databases. *ACM Comput. Surv.*, 50(5):68:1–68:40, 2017.

[4] E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, 1970.

[5] E. F. Codd. Extending the database relational model to capture more meaning. *ACM Trans. Database Syst.*, 4(4):397–434, Dec. 1979.

[6] A. Deutsch and Y. Papakonstantinou. Graph data models, query languages and programming paradigms. *Proc. VLDB Endow.*, 11(12):2106–2109, 2018.

[7] J. Lu and I. Holubová. Multi-model data management: What's new and what's next? In *Proceedings of the 20th International Conference on Extending Database Technology, EDBT 2017, Venice, Italy, March 21-24, 2017*, pages 602–605. OpenProceedings.org, 2017.

[8] J. Lu and I. Holubová. Multi-model Databases: A new journey to handle the variety of data. *ACM Computing Surveys*, 52(3), 2019.

[9] J. Lu, I. Holubová, and B. Cautis. Multi-model databases and tightly integrated polystores: Current practices, comparisons, and open challenges. In *CIKM '18*, pages 2301–2302, New York, NY, USA, 2018. ACM.

[10] I. Robinson, J. Webber, and E. Eifrem. *Graph Databases: New Opportunities for Connected Data.* O'Reilly Media, Inc., 2nd edition, 2015.

[11] M. Saeed, M. Villarroel, A. Reisner, G. Clifford, L.-w. Lehman, G. Moody, T. Heldt, T. Kyaw, B. Moody, and R. Mark. Multiparameter Intelligent Monitoring in Intensive Care II (Mimic-II): A Public-Access Intensive Care Unit Database. *Critical care medicine*, 39:952–60, 05 2011.

[12] M. H. Scholl. Extensions to the Relational Data Model. In *Conceptual Modelling, Databases and CASE: An Integrated View of Information Systems Development.* Jon. Wiley & Sons, 1992.

[13] M. H. Scholl, H. Paul, and H. Schek. Supporting flat relations by a nested relational kernel. In *VLDB'87, September 1-4, 1987, Brighton, England*, pages 137–146. Morgan Kaufmann, 1987.

[14] P. T. Wood. Query languages for graph databases. *SIGMOD Rec.*, 41(1):50–60, 2012.

[15] C. Zhang and J. Lu. Holistic evaluation in multi-model databases benchmarking. *Distributed and Parallel Databases*, pages 1–33, 2019.

[16] C. Zhang, J. Lu, P. Xu, and Y. Chen. UniBench: A Benchmark for Multi-model Database Management Systems. In *TPCTC '18, Rio de Janeiro, Brazil, August 27-31, 2018, Revised Selected Papers*, volume 11135 of *Lecture Notes in Computer Science*, pages 7–23. Springer, 2018.