



## Editorial

Alexander Böhm<sup>1</sup> · Carsten Binnig<sup>2</sup> · Tilmann Rabl<sup>3</sup> · Theo Härder<sup>4</sup>

Accepted: 4 November 2022 / Published online: 9 December 2022

© Gesellschaft für Informatik e.V. and Springer-Verlag GmbH Germany, part of Springer Nature 2022

### 1 Schwerpunktthema “Testing and Benchmarking Database Management Systems”

Today’s database management systems are the result of several decades of research and engineering. These efforts have resulted in a multitude of both open-source and commercial systems that are widely deployed in production and provide the backbone of a vast range of mission-critical applications. To ensure a high software quality, development teams routinely have to apply sophisticated testing strategies for finding defects early and ensuring the robustness, performance, and scalability of their solutions.

This special issue of the *Datenbank-Spektrum* focuses on aspects that are related to techniques for testing and benchmarking database management systems. Moreover, it gives an overview of ongoing community efforts and recent workshops on the topic. As part of the special issue, we selected a set of interesting *research contributions* that we summarize in the following:

- In the first contribution *Testing Very Large Database Management Systems: The Case of SAP HANA*, the development team of SAP’s in-memory database system HANA gives a holistic overview of their challenges with developing and testing complex system software. The article discusses a multitude of dimensions, trade-offs and best practices that have evolved during the development of the system in the last decade.
- The second contribution *Benchmarking JSON Document Stores in Practice* by Stefano Belloni and Daniel Ritter (SAP SE) shares insights from using state-of-the-art benchmarks like DeepBench to analyze popular JSON document stores. Their discussion and interesting findings highlight the importance of developing new benchmarks and extending existing ones to uncover additional optimization potential in systems that are widely used today.
- The third contribution is the vision paper *Towards a Benchmark for Shared Databases* by Muhammad El-Hindi et al. (TU Darmstadt). This paper continues along the lines of extending existing benchmarks and discusses the specific challenges of benchmarking database systems that are shared among multiple organizations. Specifically, those systems have strong requirements about data and workload verification, as well as auditing tasks that are not covered by most of the benchmarks that are popular today.
- Next, Meikel Poes (Oracle Corp.) recaps the history of the Transaction Processing Performance Council (TPC) which created several seminal benchmarks like TPC-C, TPC-H, TPC-DS, and others in *TPC, where art thou?*. The article discusses key achievements, milestones and challenges of TPC, as well as the adoption of the TPC-family of benchmarks by various commercial systems as well as by the overall database community.
- While standard benchmarks are a key ingredient in database testing, they cannot reflect the specific interaction pattern of individual applications. Hence, in the next contribution *Comprehensive and Efficient Workload Com-*

---

✉ Alexander Böhm  
alexanderboehm@google.com

Carsten Binnig  
carsten.binnig@cs.tu-darmstadt.de

Tilmann Rabl  
tilmann.rabl@hpi.de

Theo Härder  
haerder@cs.uni-kl.de

<sup>1</sup> Google, ABC-Str. 19, 20354 Hamburg, Germany

<sup>2</sup> Daten- und AI-Systeme, TU Darmstadt, Hochschulstr. 10, 64289 Darmstadt, Germany

<sup>3</sup> Data Engineering Systems, HPI, Universität Potsdam, August-Bebel-Str. 88, 14482 Potsdam, Germany

<sup>4</sup> AG Datenbanken und Informationssysteme, TU Kaiserslautern, 67663 Kaiserslautern, Germany

pression, Shaleen Deep et al. (Microsoft Corp.) discuss the challenge of integrating these specific patterns into the database quality assurance process, with a particular focus on compressing the workload and thus finding a good trade-off between test runtime and statement coverage.

## 2 Community-Beiträge

In the community section of this special issue, we selected three contributions that summarize the past and ongoing efforts on *Testing and Benchmarking Database Management Systems*.

- In the first article *Reviving the Workshop Series on Testing Database System – DBTest*, Tilmann Rabl et al. discuss the history and recent instances of the DBTest workshop that was re-instantiated after several years of inactivity in a bi-annual format, again co-located with ACM SIGMOD. They summarize some of the challenges the workshop had over the last years, as well as highlights and important trends.
- In the second article, Manuel Rigger et al. report on a recent, interdisciplinary Dagstuhl seminar on *Ensuring the Reliability and Robustness of Database Management Systems*, highlight some key findings and outline future activities.
- In the third contribution, Wolfgang Lehner (TU Dresden) reports on a three-day workshop for PhD Students on *Reproducible Science in Data Management*. The workshop was collocated with this year's GI symposium on foundations of databases.

Overall, with this special issue we see the wide spectrum of different ongoing research and community efforts and hope to see even more activity in the future.

Die Rubrik „Datenbankgruppen vorgestellt“ enthält den Beitrag *The Big Data Analytics Group at Saarland University* von Jens Dittrich (Univ. des Saarlandes). Der Beitrag gibt einen Überblick über die Forschungs- und Lehraktivitäten dieser Arbeitsgruppe. Bereits seit 2008 hat diese Arbeitsgruppe mit Variationen der klassischen Lehre und verschiedenen Online-Lehrformen experimentiert, so dass ihre Erfahrungen sehr wertvoll für die künftige Gestaltung der Datenbanklehre in der DBIS-Gemeinde sein dürften.

Die Rubrik „Community“ berichtet unter „News“ über aktuelle Informationen, welche die DBIS-Gemeinde betreffen.

## 3 Künftige Schwerpunktthemen

### 3.1 Trends in Social Media Analysis to Address Fake News, Hate Speech, or Bias

Social media has many benefits: from staying in contact with close and not-so-close friends, over exercising the right to voice one's opinion, to communicating with many like-minded people all over the world and providing an additional channel for information exchange.

Unfortunately, social media has also been abused and misused ever since its inception. Hate speech is prevalent on many sites alienating trusting users and hindering fruitful discussions. Fake news are distributed through social media platforms with dangerous effects. But even without malicious intention, social media can be misleading due to various biases in the system.

In this special issue of *Datenbank-Spektrum*, we will explore and present current trends in the field of automatically detecting and managing hate speech, fake news, bias and other toxic content in the context of social media.

We welcome original contributions including technical papers, application-oriented papers, case studies, survey papers and position papers. Topics of interest include, but are not limited to:

- Automatic detection of hate speech
- Methods to improve online discussions
- Trust and reputation of social media actors
- Identification of fake news
- Countermeasures to fight fake news
- Detection and/or mitigation of bias
- Dealing with bias in training data
- Content analysis and NLP
- Opinion mining and sentiment analysis on social media
- Information extraction and retrieval on social media
- Information diffusion within social networks
- Ethical and legal aspects

Paper format: 8–10 pages, double-column (cf. author guidelines at <http://www.springer.com/13222>). We welcome contributions in both German and English.

Deadline for submissions: Oct. 1st, 2022; Publication of special issue: DASP-1-2023 (March 2023)

Guest editors:

Ralf Krestel, ZBW & CAU Kiel

[r.krestel@zbw.eu](mailto:r.krestel@zbw.eu)

Udo Kruschwitz, Universität Regensburg

[Udo.Kruschwitz@ur.de](mailto:Udo.Kruschwitz@ur.de)

Michael Wiegand, Universität Klagenfurt

[michael.wiegand@aau.at](mailto:michael.wiegand@aau.at)

### 3.2 Managing Data and Metadata in Complex Enterprise Landscapes

The digital transformation generates huge amounts of heterogeneous data, across the entire lifecycle of all kinds of products and services and across all kinds of businesses. Extracting insights from these data by applying data analytics and AI constitutes a critical success factor for enterprises, e.g., to optimize processes and reinvent business models. Comprehensive analytics efforts and vast amounts of data have made enterprise data landscapes far more complex revealing globally distributed, federated and hybrid deployed structures of analytical and operational data systems. This poses new challenges to both data management and metadata management: new kinds of data platforms have emerged, e.g., data lakes, data catalogs and data marketplaces, semantic techniques for managing data and metadata are increasingly becoming popular in industry practice, data governance and data strategy concepts are developed to ensure the compliant and economically beneficial use of data.

In this special issue of *Datenbank-Spektrum*, we call for contributions on technical and organizational aspects of data management and metadata management in complex enterprise landscapes, interpreted broadly. We welcome original contributions – including technical papers, interdisciplinary and application-oriented papers, case studies and survey papers – relating to the following areas, but not limited to:

- Data platform architectures and technologies, e.g., data lakes, data catalogs, data marketplaces, feature stores
- Architecting and modeling data and metadata in data platforms, e.g., semantic data modeling for data lakes and data catalogs, reference data models, data model management, data model evolution
- Data engineering and metadata management for analytics and AI, e.g., for data pipelines and MLOps
- Data integration and data quality in complex enterprise landscapes, e.g., federated data integration, semantic data integration, distributed data quality assessments
- Enterprise data architecture: organizing data and metadata across the enterprise landscape, e.g., across several data lakes, data catalogs and operational systems

- Data governance and data strategy, e.g., data ownership and data stewardship across operational and analytical systems, organizational roles for data governance and data analytics, data offense and data defense concepts

Paper format: 8–10 pages, double-column (cf. author guidelines at [www.springer.com/13222](http://www.springer.com/13222)). Contributions either in German or in English are welcome.

Deadline for submissions: February 1st, 2023 Issue delivery: DASP-2-2023 (July 2023)

Guest editors:

Christoph Gröger, Robert Bosch GmbH, Stuttgart  
christoph.groeger@de.bosch.com

Holger Schwarz, University of Stuttgart  
holger.schwarz@ipvs.uni-stuttgart.de

### 3.3 Best Workshop Papers of BTW 2023

This special issue of the “*Datenbank-Spektrum*” is dedicated to the Best Papers of the Workshops, Demos and Data Science Challenge running at the BTW 2023 at the TU Dresden. The selected Workshop contributions should be extended to match the format of regular DASP papers.

Paper format: 8–10 pages, double-column

Selection of the Best Papers by the Workshop chairs and the guest editor: April 1st, 2023

Deadline for submissions: June 1st, 2023  
Issue delivery: DASP-1-2022 (Nov. 2023)

Guest editor:

Uta Störl, FernUni Hagen  
uta.stoerl@fernuni-hagen.de