



Editorial

Uta Störl¹ · Stefanie Scherzinger²

Angenommen: 26. Februar 2024 / Online publiziert: 1. März 2024
© The Author(s) 2024

1 Schwerpunktthema: „Data Management on Quantum Hardware“

Mit der Verfügbarkeit von Quanten-Hardware als kommerzielle Cloud-Dienste wächst das Interesse in verschiedenen Teildisziplinen der Informatik, das Potenzial dieser Technologie zu erkunden.

Dabei stellt die Entwicklung von hybriden Algorithmen, die teilweise auf klassischer Hardware und teilweise auf Quanten-Hardware ausgeführt werden, bedeutende Herausforderungen dar. Die Programmierung erfordert eine radikal neue Denkweise, und die experimentelle Evaluierung auf Simulatoren oder realer Hardware eine sehr spezifische Herangehensweise. Eine weitere Hürde besteht darin, sich mit den Einschränkungen der derzeit verfügbaren Quantenprozessoren auseinanderzusetzen, da bisher nur kleine Eingabegrößen verarbeitet werden können.

Das wachsende Interesse seitens der Datenbank-Community an diesem Thema zeigt sich bereits in eigenen Tutorials (SIGMOD'23, ICDE'24) und ersten einschlägigen Workshops. Aktuelle Forschungsthemen umfassen das maschinelle Lernen auf Quantencomputern sowie Unterstützung bei grundlegenden Aufgaben der Datenbankmanagementsysteme, wie beispielsweise Anfrageoptimierung und das Scheduling von Transaktionen. Diese Aufzählung ist keineswegs abschließend, da die Forschung in diesem Bereich noch in einem sehr frühen Stadium ist und die potenzielle Verfügbarkeit zukünftiger Hardware, wie QRAM, vollkommen neue Möglichkeiten eröffnen könnte.

Mit dieser Ausgabe des Datenbank-Spektrums bieten wir ein dediziertes Forum, um aktuelle Arbeiten an der Schnittstelle zwischen Forschung im Datenmanagement, Quanten-

Softwareentwicklung und der Architektur von Datenbanksystemen vorzustellen.

Den Einstieg in die Thematik ermöglicht ein Beitrag von Sieglinde M.-L. Pfaendler (IBM Quantum), Konstantin Kosnon (IBM Quantum) und Franziska Greinert (TU Braunschweig). Ihr Artikel *Advancements in Quantum Computing – Viewpoint: Building Adoption and Competency in Industry* ebnet Organisationen und Technologieführern den Weg in das Gebiet des Quantencomputings. Die Autorengruppe präsentiert eine fundierte Einführung in Schlüsselkonzepte, gewährt Einblicke in die rasant fortschreitende Entwicklung, und gibt einen Überblick darüber, wie Unternehmen die Entwicklung von Kompetenzen und die Integration von Quantencomputern strategisch angehen können. Dies beinhaltet auch eine Übersicht über europäische Aktivitäten sowie verfügbare Lehrmaterialien in deutscher Sprache.

Die Autorengruppe Nitin Nayak, Tobias Winker, Umut Çalikylmaz, Sven Groppe und Jinghua Groppe von der Universität zu Lübeck adressieren mit ihrem Artikel zu *Quantum Join Ordering by Splitting the Search Space of QUBO Problems* ein wichtiges Optimierungsproblem der Datenbankforschung. In ihrem Artikel stellt die Autorengruppe den Ansatz vor, das Join Ordering Problem als Quadratic Binary Optimization (QUBO) Problem zu formulieren. Dieses wird dann auf Quanten-Hardware gelöst. Während auch frühere Ansätze einen QUBO-basierten Ansatz wählen, ist man in der vorgestellten Arbeit nicht auf linkstiefe Anfragebäume eingeschränkt. Die Integration der hybriden Algorithmen mit dem Kostenmodell des Anfrageoptimierers von PostgreSQL stellt dabei einen wertvollen Beitrag dar, da bisherige Ansätze das Optimierungsproblem lediglich isoliert für algebraische Ausdrücke betrachtet haben. Dadurch werden erstmalig Studien mit realen Abfragen möglich. Die Autoren führen ihre Experimente sowohl auf der Architektur des D-Wave Quantum Annealers als auch auf gatterbasierten Quantencomputern durch.

Eine bereits etablierte Forschungsrichtung ist das *Quantum Machine Learning*. In ihrem Beitrag *Satellite Image Representations for Quantum Classifiers* geben Johann Ma-

✉ Uta Störl
uta.stoerl@fernuni-hagen.de

Stefanie Scherzinger
Stefanie.Scherzinger@uni-passau.de

¹ FernUniversität in Hagen, Hagen, Deutschland

² Universität Passau, Passau, Deutschland

ximilian Zollner, Paul Walther und Martin Werner von der TU München einen Einblick in die Herausforderungen und Lösungsansätze im Bereich der Verarbeitung von Satellitenbildern. Da Satellitenbilder zu groß und hochdimensional sind, um die Daten in den begrenzten Eingabebereich von Quantencomputern einzupassen, sind Transformationen zur Verringerung der Dimensionalität notwendig. Der Artikel vergleicht somit verschiedene Techniken zur Dimensionsreduktion als Teil hybrider quanten-klassischer Systeme zur Darstellung von Satellitenbildern.

Wir danken den Gutachtern für ihre Unterstützung bei diesem Themenheft. Aufgrund des noch sehr jungen Forschungsgebiets war es eine besondere Herausforderung, entsprechende Expertisen einzuholen.

2 Fachbeiträge

In der Ausgabe 3-2023 des Datenbank-Spektrums wurden die „Best Workshop Papers of BTW 2023“ mit erweiterten Beiträgen präsentiert. Nachdem im letzten Heft fünf dieser sechs eingeladenen Beiträge erschienen sind, veröffentlichen wir in diesem Heft nun den sechsten Beitrag: Aus dem Workshop Data Engineering for Data Science (DE4DS) wurde der Beitrag *A Provenance Management Framework for Knowledge Graph Generation in a Web Portal* von Erik Kleinsteuber, Tarek Al Mustafa, Franziska Zander, Birgitta König-Ries und Samira Babalou von der Uni Jena bzw. dem iDiv Halle-Jena-Leipzig nominiert. Die Autorengruppe beschäftigt sich mit der Verwaltung von Provenance-Daten in Knowledge Graph Management-Plattformen. Die Autorinnen und Autoren stellen ein Rahmenwerk vor und zeigen, wie in diesem Provenance-Informationen erfasst, gespeichert und abgerufen werden. Die Darstellung ist dabei an der standardisierten W3C Provenance Ontology ausgerichtet. Damit können wichtige Anforderungen aus dem Bereich Reproduzierbarkeit und Provenance erfüllt werden. Dieser Artikel kann auch als thematischer Vorgriff auf den Themenschwerpunkt des nächsten Datenbank-Spektrums gesehen werden: *Data Management for Research Data Infrastructures* (DASP-2-2024)

Auch der zweite Fachbeitrag *PyPads – Transparent Machine Learning Experiment Tracking* beschäftigt sich mit der Herausforderung der Reproduzierbarkeit, und zwar im Bereich des maschinellen Lernens. In diesem Bereich werden experimentelle Details oft aus Platz-, Zeit- oder Komplexitätsgründen weggelassen. Thomas Weißgerber, Mehdi Ben Amor, Christofer Fellicious und Michael Granitzer von der Universität Passau stellen in ihrem Beitrag *PyPads* vor – ein Rahmenwerk, welches eine Infrastruktur zur Erweiterung von Experimenten und Versuchsaufbauten um Protokollierungs-, Kommunikations- und Analysefunktionen darstellt.

3 Community-Beiträge

Die Rubrik „Dissertationen“ enthält in diesem Heft sieben Kurzfassungen von Dissertationen aus der deutschsprachigen DBIS-Community, welche im letzten Quartal erfolgreich verteidigt wurden.

Die Rubrik „Community“ enthält unter News den Call for Papers für die nächste BTW, welche im März 2025 in Bamberg stattfinden wird.

Außerdem findet sich in dieser Rubrik eine Meldung über die Verleihung des VDI-Preis 2023 für die Masterabsolventin Kristin Fritsch (Universität Passau) durch den Verein Deutscher Ingenieure (VDI) Bezirksverein München, Ober- und Niederbayern. Die Masterarbeit wurde – passend zum Schwerpunktthema dieses Heftes – im Schnittbereich von Datenbankforschung und Quanteninformatik angefertigt.

In der Rubrik „Produkt-News“ werden Neuerungen in verschiedenen Datenbankmanagementsystemen und Analyse-Plattformen vorgestellt.

4 Künftige Schwerpunktthemen

4.1 Data Management for Research Data Infrastructures

Research data plays a crucial role in advancing scientific knowledge and driving innovation across various fields. Effective data management approaches are essential for ensuring the integrity, accessibility, and usability of research data. The FAIR data principles (Findability, Accessibility, Interoperability, and Reuse) are generally accepted as a goal in this context, but their comprehensive implementation remains a challenge. Hence, innovative data management solutions for Research Data Infrastructures (RDIs) are needed to unlock the full potential of research data and enable collaboration, reproducibility, and knowledge exchange.

In Germany, the German National Research Data Infrastructure (NFDI) started in 2020 and comprises 26 consortia addressing a broad variety of disciplines. Currently, research data infrastructure projects inside and outside the NFDI develop data management solutions. In this special issue of Datenbank-Spektrum, we will provide a forum for exploring and presenting current trends and solutions in the field of data management for research data infrastructures.

We welcome original contributions, including technical papers, application-oriented papers, case studies, survey papers, and position papers addressing for example the following challenges in data management for research data infrastructures:

- Data Volume and Variety
- Data Integration and Interoperability
- Data Quality and Integrity
- Data Security and Privacy
- Long-Term Preservation
- Ethical and Legal Considerations
- Data Discoverability and Access

Paper format: 8–10 pages, double-column. We welcome contributions in both, German, and English.

Deadline for submissions: February 1st, 2024
Publication of special issue: DASP-2-2024 (July 2024)

Guest editors:

Andreas Henrich, University of Bamberg
andreas.henrich@uni-bamberg.de

Wolfgang Müller, Heidelberg Institute for Theoretical Studies

Wolfgang.Mueller@h-its.org

Bernhard Seeger, Philipps-Universität Marburg
seeger@mathematik.uni-marburg.de

4.2 ML for Systems and Systems for ML

Machine learning (ML) methods are data-intensive, and data engineering workflows are an essential part of all machine learning workflows. They provide data for the specific machine learning tasks and usually combine multiple data engineering tasks. Within the lifecycle of a machine learning pipeline, collecting, storing, and selecting the data for (re)training the model and deploying the model (in a distributed environment) are performance-intensive parts. In this way, the data engineering tasks enable machine learning solutions.

It is also possible to use machine learning methods for individual data engineering tasks (like data exploration, discovery, certain data cleaning tasks, and data integration) or to optimize the data engineering processes. Machine learning for systems aims to replace components of systems with learned models. Database systems benefit from learned components such as cardinality estimation or query optimizer hinting.

Another current research topic is the development of systems for machine learning to support design of end-to-end machine learning solutions. For data engineering, systems are fundamental for processing data and their enhancement is crucial for efficient model training and inference. Thus, systems for machine learning enhance the application of models by integrating machine learning functionality into data management systems.

In this way, the fields of databases and machine learning are growing closer together and are mutually dependent.

We want this special issue to highlight the resulting new approaches. We welcome original contributions or extensions of previous work on the following topics (but not limited to):

- Learned system components or algorithms (such as index structures, query optimization and optimizer hinting)
- Self-tuning of database systems
- Data cleaning and monitoring techniques for ML workflows
- Data quality management and evaluation for ML workflows
- ML methods for data exploration, discovery, and integration
- Data, metadata and model management in ML applications and complex ML pipelines
- Novel use of natural language models and interfaces in data management
- Novel use of ML techniques in big data applications (e.g., digital biology, video analytics)
- New datasets, benchmarks, and evaluation methods for evaluating ML approaches
- Data integration, alignment, and preparation of multi-modal training datasets
- New/Extended Systems for ML (Database Systems extended for ML, ML systems extended by compression/indexing/partitioning)
- Physical/Logical optimization of ML pipelines (data flow by operator rewriting/fusion, data-/task-parallel execution strategies)
- Further topics combining database and machine learning technologies

Paper format: 8–10 pages, double-column. We welcome contributions in both, German, and English.

Deadline for submissions: June 1st, 2024

Publication of special issue: DASP-3-2024 (November 2024)

Guest editors:

Maximilian E. Schüle, University of Bamberg
maximilian.schuele@uni-bamberg.de

Meike Klettke, University of Regensburg
meike.klettke@ur.de

4.3 Cloud-Native Database Management Systems

Within the last decade, cloud-based database management systems (DBMS) and data management solutions have gained an ever increasing popularity. Companies and organizations have been migrating to the cloud in order to modernize and consolidate their infrastructure, reduce

costs, and improve on the overall scalability and elasticity of their data processing stacks.

Besides hosting open-source and commercial database management offerings, various cloud-native database systems have emerged. They offer unique characteristics that go beyond the capabilities of traditional DBMS architectures. This leads to new design choices for the core database system as well as the infrastructure to deploy and operate cloud-based DBMS to achieve unprecedented availability, scalability, elasticity, fault tolerance, and reliability offering various price-performance options.

This special issue of the Datenbank-Spektrum focuses on discussing all aspects of these cloud-native database management systems, ranging from the specific challenges of designing, implementing and continuously delivering these systems, to their successful consumption by customers. The topics of interest for this issue include, but are not limited to:

- Architectural aspects of cloud-native data management systems (componentization, service composition, disaggregation) and their implications
- Big data systems
- Streaming systems
- OLTP and HTAP processing in cloud environments
- High availability and fault tolerance
- Use of Machine Learning in the context of cloud-native DBMS
- Systems for „serverless“ data processing
- Load balancing and autoscaling
- (Geo-Distributed) Transaction processing
- Frameworks for the definition and monitoring of Service Level Indicators (SLI) and Objectives (SLO)
- Benchmarking and performance analysis
- Backup and restore
- Quality assurance and regression detection
- Site Reliability Engineering (SRE) for database management systems
- Troubleshooting and supportability aspects
- Testing strategies (i.e. stress-testing, fault-injection, etc)
- Data tiering and different storage strategies
- Security aspects
- Data encryption

- War stories about application migration, system architecture, and others
- Data migration between on-premise and cloud systems, or multi-cloud setups
- Data-intensive application leveraging the unique characteristics of cloud DBMS

We welcome traditional research articles, experience and application reports, as well as system overviews, surveys, and experimental studies. These can either be in the form of full submissions (8–10 pages) as well as short papers/extended abstracts (not more than 4 pages) for this issue.

Submission deadline is October 1st, 2024

Publication of special issue: DASP-1-2025 (March 2025)

Guest editors:

Alexander Boehm, SAP SE

alexander.boehm@sap.com

Norman May, SAP SE

norman.may@sap.com

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.