Datenbank Spektrum (2014) 14:169–172
DOI 10.1007/s13222-014-0169-7

EDITORIAL

Editorial

Theo Härder · Jens Teubner

Online publiziert: 15. Oktober 2014

Schwerpunktthema: Data Management on New Hardware

Seit Jahren bieten innovative Prozessor-Architekturen, bahnbrechende Neuerungen bei den Speichertechnologien und stürmische Weiterentwicklungen bei den Infrastrukturen neue wissenschaftliche Fragestellungen und Forschungsfelder für die Datenbankgemeinde. Wegen der dramatischen Steigerungen bei Speicherkapazitäten und Ein-/Ausgabe-Raten – bei gleichzeitiger Reduktion der Kosten – haben diese Entwicklungen im Zeitalter von *Big Data* auch eine große wirtschaftliche Bedeutung.

Andererseits decken enorm verbesserte Möglichkeiten zur Nutzung von Parallelität durch Multi-Core-Prozessoren oder Cluster-Architekturen sowie größere Bandbreiten und höhere Transferraten bei der Datenübertragung ständig neue Engpässe und erhöhtes Blockierungspotential in existierenden Systemen auf. Deshalb erzwingen diese Entwicklungen auch eine ständige Anpassung und Optimierung der verfügbaren Methoden und Techniken bei den Software-Lösungen, insbesondere bei Datenbanksystemen. Während früher solche Anpassungen oft durch bloße E/A-Optimierung zu erreichen waren, erfordert heute die effiziente Ausnutzung der verschiedenartigen und komplexen Charakteristika der modernen Hardware eine abgestimmte Vorgehensweise, die mehrere oder alle Komponenten zugleich betrifft. Außerdem standen früher allein die verschiedenen Aspekte der Leistungsoptimierung bei Datenbanksystemen im Mittelpunkt, während heutzutage neben hoher Performanz zunehmend Energieeffizienz oder gar Energieproportionalität beim DBMS gefordert wird. Weiterhin ist ein wichtiges Ziel bei dieser DBMS-Evolution die automatische und für die Anwendung transparente Anpassung an die hochentwickelten Hardware-Komponenten.

Wichtige Bereiche für neue Lösungen betreffen beispielsweise Hardware-unterstützte Anfrageverarbeitung, Datenverwaltung bei Nutzung von Co-Prozessoren oder GPUs, neuartige Anwendungen von neuen und künftigen Speichertechnologien (Flash, PCM, NVRAM usw.), DBMS-Architekturen für Transactional Memory, Low-Power Computing, Embedded Devices usw. Außerdem ist es erforderlich, für solche neuen DBMS-Architekturen geeignete Tools zur Analyse und Optimierung sowie zur Leistungs-/Energiemessung von Komponenten und Gesamtsystem bereitzustellen. Damit die oben genannten Ziele für das *Data Management on New Hardware* auch überprüft und verschiedene Ansätze zu ihrer Optimierung verglichen werden können, sind letztlich auch geeignete Benchmarks zu entwickeln, die nicht nur Performanz-zentriert sind, sondern insbesondere auch wichtige Aspekte der Energieeffizienz berücksichtigen.

Vor einigen Jahren hatte ein Heft des Datenbank-Spektrums schon einmal dieses Schwerpunktthema gewählt. Damals wurden ausschließlich Beiträge zu Aspekten der Leistungssteigerung des Datenbanksystems bzw. seiner Anwendungen eingereicht, wobei Nutzung von Flash-Speichern (SSDs) und Optimierungsmöglichkeiten bei der hauptspeicherbasierten Datenbankverarbeitung im Vordergrund standen. In den vier Beiträgen dieses Heftes hat sich der

T. Härder (✉)
AG Datenbanken und Informationssysteme,
TU Kaiserslautern, 67663 Kaiserslautern,
Deutschland
E-Mail: haerder@cs.uni-kl.de

J. Teubner
Fakultät für Informatik, Lehrstuhl Datenbanken und
Informationssysteme, TU Dortmund, 44227 Dortmund,
Deutschland
E-Mail: jens.teubner@cs.tu-dortmund.de

Schwerpunkt deutlich verschoben. Zwei Beiträge konzentrieren sich unter Nutzung neuer Hardware-Architekturen auch auf Fragen der Energieeffizienz. In weiteren Beiträgen werden vor allem DB-bezogene Optimierungsmöglichkeiten bei Einsatz von GPUs, FPGA Cores, Many-Core NUMA-organized DB Servers usw. untersucht.

Im ersten Beitrag mit dem Titel *HyPer beyond Software: Exploiting Modern Hardware for Main-Memory Database Systems* überprüfen Florian Funke, Alfons Kemper, Tobias Mühlbauer, Thomas Neumann und Viktor Leis (TU München) die Nutzung neuartiger und vielfältiger Hardware-Möglichkeiten zur Optimierung von Hauptspeicher-Datenbanksystemen im Kontext des HyPer-Projektes. Insbesondere wird die Virtuelle Speicherverwaltung eingesetzt, um auf den DB-Daten OLAP-Anfragen von parallelen OLTP-Transaktionen zu separieren. Weiterhin untersuchen die Autoren Konzepte und Verfahren zur Trennung der DB-Daten in „heiße und kalte“ Partitionen, zur adaptiven Parallelisierung und Partitionierung, um eine erhöhte Datenlokalität bei Prozessorkernen zu erreichen, sowie zur Verbesserung der Synchronisation bei OLTP-Transaktionen. Schließlich berichten sie, wie heterogene Prozessoren von verbrauchsarmen Rechnern zur leistungsstarken und energieeffizienten Anfrageverarbeitung eingesetzt werden können.

Im folgenden Beitrag fassen Daniel Schall und Theo Härder (TU Kaiserslautern) die Arbeiten zu ihrem DFG-Projekt *Energieeffiziente Verarbeitung in Datenbanksystemen* zusammen. Unter den Titel *WattDB—A Journey towards Energy Efficiency* beschreiben sie die Entwicklung von WattDB, einen verteilten DBMS, das auf einen dynamischen Cluster von leistungsschwachen Rechnern abläuft. Das Projekt untersucht, wie und ob die Leistung einen zentralisierten DB-Servers durch ein Rechner-Cluster bereitgestellt werden kann, wobei Energieproportionalität bei der DB-Verarbeitung approximiert werden soll. WattDB nähert sich diesem Ziel durch automatisches Zu- und Abschalten von Rechnern in Abhängigkeit von der DB-Arbeitslast an. Ein wesentliches Problem ist die Erreichbarkeit aller DB-Daten von jedem aktiven Rechnerknoten, was flexible, dynamische Datenpartitionierung und Datenallokation impliziert. Mit einem Experiment auf großem Server und dynamischem Zehn-Knoten-Cluster – mit vergleichbaren Ressourcen hinsichtlich CPU-Leistung, Hauptspeicher- und Cache-Größe und Externspeicher-Ausstattung, wobei eine identische Version von WattDB mit derselben Arbeitslast eingesetzt wurde – konnten die genauen Abweichungen bei Transaktionsleistung und Energieverbrauch gemessen werden. Während das Cluster durchgehend bessere Werte für Energieeffizienz erreichte, konnte es nur bei mittleren oder geringen OLAP-Lasten hinsichtlich Transaktionsleistung mit dem großen Server mithalten.

Der dritte Beitrag *The Design and Implementation of CoGaDB: A Column-oriented GPU-accelerated DBMS* von Sebastian Breß (TU Dortmund) liefert einen Einblick in die Probleme und Techniken beim Entwurf und bei der Implementierung eines Hauptspeicher-Datenbanksystems, das zur Leistungssteigerung eine „eingebaute“ GPU als Co-Prozessor einsetzt, um OLAP-Arbeitslasten in optimierter Weise verarbeiten zu können. CoGaDB setzt das Optimierer-Framework HyPE zur Realisierung eines Hardware-unabhängigen Anfrageoptimierers ein, der in der Lage ist, Kostenmodelle für DB-Operatoren zu lernen und Arbeitslasten effizient auf verfügbare Prozessoren zu verteilen. CoGaDB implementiert weiterhin effiziente Algorithmen – insbesondere auch den Star Join – für den kombinierten Einsatz auf CPU und GPU. Der Beitrag macht deutlich, wie diese neuen Techniken in einem einzigen System zusammenspielen. Schließlich belegen empirische Experimente, dass sich CoGaDB zur Laufzeit schnell durch zunehmende Genauigkeit seiner Kostenmodelle an die konkret verfügbare Hardware anpasst.

Der vierte Beitrag *Heterogeneity-aware Operator Placement in Column-Store DBMS* kommt von der TU Dresden mit den Autoren Thomas Karnagel, Dirk Habich, Benjamin Schlegel und Wolfgang Lehner. Unter der Annahme einer Multi-Core-CPU als homogene Ablaufplattform bestimmen existierende Anfrageoptimierer für eine SQL-Anfrage die effizienteste Auswertungsreihenfolge der erforderlichen physischen Operatoren. Jedoch nimmt heutzutage die Heterogenität bei der Hardware zu, so dass eine Multi-Core-CPU mehr und mehr durch verschiedene Recheneinheiten, wie z. B. GPU oder FPGA-Kernen, ergänzt wird. Wegen dieser Heterogenität wird die Optimierung der Zuordnung physischer Operatoren immer wichtiger. In ihrem Beitrag schlagen die Autoren eine entsprechende Strategie, HOP (Heterogeneity-aware physical Operator Placement) genannt, für hauptspeicherbasierte, spaltenorientierte Datenbanksysteme vor. Um Zuordnungsentscheidungen zu Laufzeit in optimaler Weise zu ermöglichen, wertet das Kostenmodell Merkmale der beteiligten Recheneinheiten, Ausführungseigenschaften der Operatoren sowie Ablaufdaten für jede Recheneinheit aus. Die experimentelle Auswertung des HOP-Strategie mit TPC-H-Anfragen zeigte beträchtliche Antwortzeitgewinne, die sich allein durch die optimierte Zuordnung nach dem HOP-Modell ergeben.

Die vier Beiträge zum Schwerpunktthema dieses Heftes werden durch einen Fachbeitrag *Eine Erweiterung des Relationalen Modells zur Repräsentation räumlichen Wissens* ergänzt. Norbert Paul und Patrick E. Bradley (KIT Karlsruhe) beschreiben darin, wie sich die enge Verwandtschaft von Topologie und Relationalem Datenmodell nutzen lässt, um topologische Konzepte in das Relationale Datenmodell einzuführen. Sie zeigen, dass der relationalen Abgeschlossenheit der Relationalen Algebra eine Art „räumlicher Abgeschlossenheit“ in der Topologie entspricht. Mit einer proto-

typischen Implementierung dieser topologisch-Relationalen Algebra illustrieren sie, wie Relationen zu topologischen Räumen werden können und wie eine entsprechend erweiterte Relationale Algebra auf diesen Räumen operiert. An einem Beispiel aus der räumlichen Wissensverarbeitung, dem Region-Connection-Calculus (RCC-8), zeigen die Autoren schließlich den Nutzen dieses generischen Ansatzes.

Unter der Rubrik „Datenbankgruppen vorgestellt“ finden Sie einen Beitrag von H.-Jürgen Appelrath und Marco Grawunder über *Die Abteilung Informationssysteme der Universität Oldenburg*. Dieser Beitrag skizziert nach einem Blick auf die geschichtliche Entwicklung des Abteilung in Universität und An-Institut OFFIS größere Projekte auf den Gebiet des intelligenten Datenmanagements mit Anwendungen in der Energiewirtschaft und im Gesundheitswesen sowie ein Framework zur Erstellung von Datenstrommanagementsystemen. Weiterhin geben die Autoren einen Überblick über eine Vielzahl weiterer aktueller Forschungsthemen ihrer Abteilung.

In diesem Heft bietet die Rubrik „Dissertationen“ sechs Kurzfassungen von Dissertationen aus der deutschen DBIS-Community.

Die Rubrik „Community“ enthält schließlich unter *News* weitere aktuellen Informationen aus der DBIS-Gemeinde.

Künftige Schwerpunktthemen

1 Informationsmanagement für Digital Humanities

In den Geisteswissenschaften fallen in immer größerer Menge digitale Forschungsdaten an. Dabei ergeben sich durch die spezifischen Rahmenbedingungen zahlreiche Herausforderungen für Datenbanken und IR-Systeme: Die Daten und Dokumente sind heterogen in Sprache, Struktur und Qualität. Es gibt zwar eine Vielzahl von Standards und Methoden, eine übergreifende Sicht existiert aber kaum. Relevante Kollektionen mit elektronischen Texten, Metadaten, Bildern und anderen multimedialen Ressourcen liegen in verschiedenen Disziplinen und Institutionen vor und bilden eine hochgradig verteilte und heterogene Informationslandschaft, deren Verarbeitung oft im Rahmen spezifischer, geisteswissenschaftlicher Forschungsfragen erfolgt. Von besonderer Bedeutung sind die Erschließung, Veröffentlichung und Verwaltung digitaler Ressourcen im Rahmen spezifischer Anwendungen z. B. in der Archäologie, den Geschichts-, Sprach- oder Religionswissenschaften, aber insbesondere auch im Kontext interdisziplinärer Forschung. Im Themenheft sollen einführende und überblicksartige Artikel sowie aktuelle Forschungsergebnisse zu ausgewählten Themen ein breites Bild zum aktuellen Stand des Informationsmanagements für Digital Humanities geben.

Mögliche Themen aus diesem Bereich könnten z. B. sein:

- Integrierte Analyse, Verarbeitung und Visualisierung verteilter bzw. heterogener Kollektionen
- Nutzung, Entwicklung und Auswertung von Vokabularen, Thesauri und Ontologien
- Langzeitarchivierung und Datenprovenienz
- Katalogisierung, Annotation und Dokumentation von Ressourcen (Data Curation)
- Erkennung, Analyse und Visualisierung kollektionsinterner oder -übergreifender Zusammenhänge z. B. durch Analyse von Ort und Zeit, Themen, Named Entities
- Aspekte der Usability im Umgang mit verteilten und heterogenen Ressourcen
- Anwendungen zum Datenmanagement, zur Suche und zur Analyse in speziellen Anwendungsfeldern aus den Geisteswissenschaften
- *Big Data*-Technologien für die Digital Humanities
- Forschungsinfrastrukturen für die Digital Humanities

Gastherausgeber:
Andreas Henrich, Otto-Friedrich-Universität Bamberg
andreas.henrich@uni-bamberg.de

Gerhard Heyer, Universität Leipzig
heyer@informatik.uni-leipzig.de

Christoph Schlieder, Otto-Friedrich-Universität Bamberg
christoph.schlieder@uni-bamberg.de

2 Data Management for Mobility

Mobility is a major factor in our society and daily life. Thus, approaches for data management need to address the resulting dynamics, geospatial and temporal relationships, and distribution of resources. In Web design, the methodology of „mobile first“ – developing new Web applications for mobile usage first and adapt it later for the desktop case – is widely embraced by industry. However, it often only considers the user interface and not the data management. This special issue addresses novel approaches and solutions for mobile data management. We invite submissions on original research as well as overview articles covering topics from the following non-exclusive list:

- Data management for mobile applications
- Context awareness in mobile applications
- Analytic techniques in mobile applications
- Management of moving objects
- Data-intensive mobile computing and cloud computing
- Data stream management
- Complex event processing
- Case studies and applications
- Foundations of data-intensive mobile computing

Expected size of the paper: 8 – 10 pages (double-column)
Important dates:

- Notice of intent for a contribution: December 15th, 2014
- Deadline for submissions: February 1st, 2015
- Issue delivery: DASP-2-2015 (July 2015)

Guest editors:
Bernhard Mitschang, University of Stuttgart
bernhard.mitschang@ipvs.uni-stuttgart.de

Daniela Nicklas, University of Bamberg
daniela.nicklas@uni-bamberg.de

3 Best Workshop Papers of BTW 2015

This special issue of the „Datenbank-Spektrum“ is dedicated to the Best Papers of the Workshops running at the BTW 2015 at the University of Hamburg. 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 15th, 2015

Guest editor:
Theo Härder, University of Kaiserslautern,
haerder@cs.uni-kl.de

Deadline for submissions: June 1st, 2015

4 Big Data & IR

The term *Big Data* refers to data and respective processing strategies, which, due to their sheer size, require a data center for the processing, and which become available through the ubiquitous computer and sensor technology in many facets of everyday life. Interesting scientific questions in this regard are the organization and management of Big Data, but also the identification of problems that now can be studied and better understood through the collection and analysis of Big Data. In the context of information retrieval as the purposeful search for relevant content, there are two main challenges: 1) retrieval in Big Data and 2) improved retrieval because of Big Data.

Retrieval in Big Data focuses on the organization, the management, and the quick access to Big Data, but also addresses the creative process of identifying interesting research questions that can only be understood and answered in Big Data. Besides the development of powerful frameworks for the maintenance and analysis of text, multimedia, sensor, and simulation data, an important research direction is the question of what kind of insights Big Data may give us today and in the future.

The second challenge in the context of Big Data & IR is the improvement of retrieval approaches through Big Data. Examples include the classic question of improved Web or eCommerce search via machine learning on user behavior data, the usage of user context for retrieval, or the exploitation of semantic data like Linked Open Data or knowledge graphs.

We are looking for contributions from researchers and practitioners in the above described context. The contributions may be submitted in German or in English and should observe a length of 8–10 pages in the Datenbank-Spektrum format (cf. the author guidelines at www.datenbank-spektrum.de).

Important dates:

- Notice of intent for a contribution: August 15th, 2015
- Deadline for submissions: October 1st, 2015
- Issue delivery: DASP-1-2016 (March 2016)

Guest editors:
Matthias Hagen, Universität Weimar
matthias.hagen@uni-weimar.de

Benno Stein, Universität Weimar
benno.stein@uni-weimar.de