Editorial

Kai-Uwe Sattler¹ · Alfons Kemper² · Theo Härder³

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2018

1 Schwerpunktthema: Data Management on New Hardware

Das Schwerpunktthema dieser Ausgabe wird vom DFG-SPP 2037 „Scalable Data Management for Future Hardware“ bestritten. In diesem seit Sommer 2017 laufenden Schwerpunktprogramm werden in 10 Einzelprojekten Methoden und Techniken für Datenmanagementsysteme erforscht, die neue und zukünftige Hardware-Entwicklungen effizient nutzen, um aktuelle Herausforderungen wie den Umgang mit großen Datenmengen und komplexen Datenstrukturen oder auch die Echtzeitverarbeitung besser adressieren zu können.

Zu diesem Thema liegen vier Beiträge vor, welche die große Bandbreite moderner Hardwaretechnologien im Datenbankkontext illustrieren und gleichzeitig einen Einblick in die Forschungsthemen des Schwerpunktprogramms geben. Im ersten Beitrag mit dem Titel *Integration of FPGAs in Database Management Systems: Challenges and Opportunities* geben die Autoren (von der Friedrich-Alexander-Universität Erlangen Nürnberg und der Otto-von-Guericke-Universität Magdeburg) einen Überblick zum Einsatz von FPGAs im Datenbankbereich. Basierend auf einer Diskussion der wesentlichen Herausforderungen wie Rekonfigurierbarkeit und die Unterstützung heterogener Architekturen werden anschließend zwei konkrete Ansätze für FPGA-basierte Anfrageverarbeitung vorgestellt: ADAMANT und

ReProVide. ADAMANT zielt durch Abstraktion von verschiedenen Beschleunigern (FPGAs, GPUs) und die Trennung in lokale und globale Optimierung auf eine adaptive DBMS-Architektur für heterogene Hardware ab. ReProVide basiert auf einer partiellen Rekonfiguration von FPGA-basierten SoCs, indem vorsynthetisierte Anfrageoperatoren auf dem FPGA dynamisch neu verknüpft werden können.

Der zweite Beitrag beschäftigt sich mit einer weiteren Art von Coprozessoren – den GPUs. Unter dem Titel *Efficient and Scalable k-Means on GPUs* stellen die Autoren (vom DFKI GmbH und der TU Berlin) einen Ansatz zur Implementierung des k-Means-Clustering-Verfahrens für GPUs vor. Ausgehend von einer Analyse der Performance-kritischen Teile des Algorithmus beschreiben sie Techniken zur Beschleunigung der Zentroid-Berechnung sowie zur Verschmelzung der Phasen zu Objektzuordnung und Zentroid-Berechnung, um die Berechnung in einem Durchlauf auf der GPU zu ermöglichen. Die Experimente zeigen dabei einen deutlichen Performance-Gewinn.

Neue Speichertechnologien für den Hauptspeicher sind Gegenstand des Beitrags *Data Management on Non-Volatile Memory: A Perspective*. Obwohl noch nicht allgemein verfügbar, stellt nicht-flüchtiger Hauptspeicher (NVM) durch seine Latenz nahe am DRAM sowie die prinzipielle Byte-Adressierbarkeit eine sehr vielversprechende Technologie für Datenbanken dar und ist daher ein hochaktueller Forschungsgegenstand. Die Autoren (von der TU Ilmenau, TU München, TU Dresden & SAP SE, Walldorf) beschreiben die wesentlichen Eigenschaften und Konsequenzen dieser Technologie, diskutieren die Rolle von NVM in der Speicherhierarchie für das Datenmanagement und geben einen Überblick zu existierenden Datenbankstrukturen und Datenbank-Engines, die für NVM ausgelegt sind.

Der vierte Beitrag zum Thema *Cooking DBMS Operations using Granular Primitives* beschäftigt sich mit Primitiven für Datenbankoperationen als Basis für die Unterstützung heterogener Hardware-Architekturen. Ausgehend von atomaren Operationen wie Map, Scan, Reduce und Scatter/Gather sowie deren optimierte Implementierung geben die Autoren von der Otto-von-Guericke-Universität Magdeburg einen Überblick zu darauf aufbauenden Realisierungen von

✉ Kai-Uwe Sattler
kus@tu-ilmenau.de

Alfons Kemper
alfons.kemper@in.tum.de

Theo Härder
haerder@cs.uni-kl.de

¹ FG Datenbanken und Informationssysteme, TU Ilmenau, 98684 Ilmenau, Deutschland

² Institut für Informatik, TU München, 85748 Garching, Deutschland

³ AG Datenbanken und Informationssysteme, TU Kaiserslautern, 67663 Kaiserslautern, Deutschland

Filtern, Hash-Operationen und Sortierung bis hin zu Standard-Datenbankoperatoren.

2 Community-Beiträge

Der Beitrag *Datenbanken und Information Retrieval an der Universität Bamberg* von Andreas Henrich und Daniela Nicklas in der Rubrik „Datenbankgruppen vorgestellt“ beschreibt zwei Arbeitsgruppen der dortigen Universität, die in den Themenfeldern Datenbanken und Information Retrieval forschen und lehren. Insbesondere wird dabei auf das spezielle Bamberger Umfeld eingegangen, das die Forschungsthemen und Lehranforderungen mitbestimmt. Weiterhin wird ein Überblick über die aktuellen Forschungsthemen und die Lehraktivitäten gegeben.

In den letzten beiden Heften des Datenbank-Spektrums wurden in der Rubrik „Kurz erklärt“ die Entwicklungen bei den Prozessoren und den Speichertechnologien erörtert, die auch einen großen Einfluss auf den Entwurf von Datenbanksystemen haben. In diesem Heft führen wir diese Betrachtungen mit den Netzwerktechnologien fort. Im Beitrag *Scalable Data Management on Modern Networks* beschreibt Carsten Binnig (TU Darmstadt) neue Kommunikationstechnologien in Hochgeschwindigkeitsnetzwerken, die eine deutliche Verbesserung von Performance und Skalierbarkeit in verteilten Datenverarbeitungssystemen erlauben.

Die Rubrik „Dissertationen“ enthält in diesem Heft 7 Kurzfassungen von Dissertationen aus der deutschsprachigen DBIS-Community.

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

3 Künftige Schwerpunktthemen

3.1 Scalable and intelligent Data Services and Solutions – Results from the Big Data Competence Center ScaDS Dresden/Leipzig

Im Oktober 2014 nahm das *Competence Center for Scalable Data Services and Solutions Dresden/Leipzig* offiziell seine Arbeit auf. In einem interdisziplinären Konzept bündelt ScaDS Dresden/Leipzig die in den beiden sächsischen Wissenschaftsregionen Dresden und Leipzig vorhandenen Kompetenzen und Ressourcen, um die Big-Data-Herausforderungen in unterschiedlichsten wissenschaftlichen und betrieblichen Anwendungsfeldern zu adressieren. Profilbestimmende Forschungsschwerpunkte liegen dabei in den Gebieten der Datenintegration, der Wissensextraktion sowie der visuellen Analyse von Daten.

Fünf Anwenderbereiche mit Big-Data-Herausforderungen sind direkt in das Zentrum eingebunden: Lebenswissenschaften, Werkstoff- und Ingenieurwissenschaften, Umwelt- und Verkehrswissenschaften, Digital Humanities und Business Data. Darüber hinaus wird ScaDS Dresden/Leipzig über ein Servicezentrum auch anderen Fachdisziplinen bzw. Branchen als zentraler Anlaufpunkt zur Nutzung von Big-Data-Technologien zur Verfügung stehen. Durch den Aufbau des Servicezentrums und die Entwicklung von Big-Data-Diensten soll eine langfristige Sichtbarkeit und Nachhaltigkeit des Kompetenzzentrums erreicht werden.

Nach vier Jahren Forschungs- und Entwicklungsarbeit sollen in diesem Heft des Datenbank-Spektrums ausgewählte Ergebnisse des ScaDS dargestellt werden.

Expected size of the paper: 8–10 pages, double-column (cf. the author guidelines at www.springer.com/13222). Contributions either in German or in English are welcome.

Deadline for submissions: Oct. 1st, 2018

Issue delivery: DASP-1-2019 (March 2019)

Guest editor:

Erhard Rahm, Universität Leipzig
rahm@informatik.uni-leipzig.de

3.2 Data and Repeatability

What is common practice in most natural sciences has only recently entered the database field: Ensuring repeatability (or reproducibility) of experiments, in order to validate scientific results and enable experimental comparisons of methods. In our field, the ability to reproduce and repeat experiments has two main ingredients: First, the software or sufficient method description. Second, the data to run the experiments on. This special issue on data and repeatability places its focus on this second, arguably more challenging part.

Providing data for experimentation must overcome many obstacles. For instance, the data must be non-private and non-proprietary; for many types of experiments, data must be properly labeled or accompanied by a gold-standard; in many cases, data is “massaged” before entering experiments; special properties of the data, such as distributions or size, must be known or even adaptable. Sometimes data is varied to fit specific needs of an experiment, i.e., through upsampling and augmentation. In addition, “input data” for experiments may refer to many different things: from raw data to cleaned data, from sets of non-integrated CSV-files to a fully integrated relational database, etc.

We are calling for non-typical database contributions that report on

- Experiences in handling data for scientific and industrial purposes
- Experiences in handling data in data science/ML/AI workflows
- Efforts to create, evaluate or use data for benchmarking
- Data preparation/cleaning, and data quality war stories
- Data life cycle management
- Long-term data preservation and curation
- Data hubs and repositories
- Description of datasets of general interest and open data
- Data and the law – legally managing data
- Possible impacts on our publishing culture

Submissions can range from single pages, for instance to introduce a dataset, to full-fledged scientific contributions, for instance an experimental analysis of data cleaning methods or war stories.

Expected size of the paper: 8–10 pages, double-column (cf. the author guidelines at www.springer.com/13222). Contributions either in German or in English are welcome.

Deadline for submissions: Feb. 1st, 2019

Issue delivery: DASP-2-2019 (July 2019)

Guest editors:

Jens Dittrich, Universität des Saarlandes

jens.dittrich@cs.uni-saarland.de

Felix Naumann, Hasso Plattner Institut, Universität Potsdam

Felix.Naumann@hpi.de

Norbert Ritter, Universität Hamburg

ritter@informatik.uni-hamburg.de

3.3 Best Workshop Papers of BTW 2019

This special issue of the “Datenbank-Spektrum” is dedicated to the Best Papers of the Workshops running at the BTW 2019 at the University of Rostock. 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, 2019

Deadline for submissions: June 1st, 2019

Issue delivery: DASP-3-2019 (November 2019)

Guest editor:

Theo Härder, University of Kaiserslautern

haerder@cs.uni-kl.de