DBIS Database Day

Friday, April 27, 2018

Humboldt-Kabinett, Institut für Informatik, Rudower Chaussee 25, 12489 Berlin

Der Lehrstuhl DBIS möchte Sie zum DBIS Database Day am Freitag, den 27. April 2018 in das Humboldt-Kabinett, Institut für Informatik, Humboldt-Universität zu Berlin, Rudower Chaussee 25, 12489 Berlin einladen.
An diesem Tage werden drei Wissenschaftler Vorträge zu datenbankorientierten Themen halten:

- **10:00 Uhr - 11:00 Uhr**: Prof. Dr. Stefan Manegold, CWI, Amsterdam, NL & Leiden University, Leiden, NL, zum Thema *Progressive Indices: Adaptive Indexing Without Prejudice*
- **11:00 Uhr – 12:00 Uhr**: Prof. Dr. Wolfgang Lehner, Technische Universität Dresden, zum Thema *Challenges of Scalable Data Management beyond Performance*
- **14:15 – 15:00 Uhr**: Dissertationsverteidigung von Herrn M.Sc. Steffen Zeuch zum Thema *Query Execution on Modern Hardware*

Alle Mitglieder des Institutes sowie Studentinnen und Studenten sind herzlich zu diesen Vorträgen eingeladen. Im Anhang finden Sie weitere Informationen zu den Vorträgen und den Vortragenden.

Der DBIS Database Day wird großzügig vom Graduiertenkolleg SOAMED unterstützt.

Mit freundlichen Grüßen,
Professor Johann-Christoph Freytag, Ph.D.

Progressive Indices: Adaptive Indexing Without Prejudice

Prof. Dr. Stefan Manegold, Centrum Wiskunde & Informatica (CWI), Amsterdam, NL Leiden University, Leiden, NL

While database indexes are crucial to achieve good database query performance, the decision which indexes to build and maintain, i.e., balancing the trade-off between the benefits of indexes for query processing versus the extra storage and processing costs to create and maintain them, is known to be a very difficult problem. The problem become particularly difficult in exploratory big data analytics scenarios where the query workload is a priori not known and where there is virtually no time between data acquisition and data analysis to prepare indexes.

To solve the problem, workload-driven adaptive and partial indexing techniques have been proposed, pioneered by Database Cracking. Database cracking is a method to create partial indices as a side effect of processing queries. Cracking effectively distributes the cost of index creation over the sequence of incoming queries while focusing indexing efforts on highly queried subsets of data. However, these strengths are also serious weaknesses: Varying workload patterns will lead to unpredictable performance, cracking has a disproportionately negative effect on the first query's performance and convergence towards a full index might never occur. For a general-purpose data management system, these weaknesses make it difficult to rely on cracking as sole autonomous indexing technique.

In talk we given an overview of various existing adaptive indexing techniques and then introduce progressive indexing, a novel indexing technique. Progressive indexing strictly limits per-query indexing cost to a user-defined fraction of scan costs, allowing the first and subsequent queries to

complete without heavy penalties. At the same time, all indexing effort is spent towards predictable convergence towards a full index.

Queries are either able to exploit already-built portions on the index given a compatible predicate, or are at worst not slowed down beyond scan cost. We implemented progressive indexing as an extension of Quick-, Merge-, Bucket- and Radixsort, and perform an experimental evaluation using synthetic workload patterns and data. Results show that progressive indexing achieves its design goals without compromising the cumulative response time.

Challenges of Scalable Data Management beyond Performance

Prof. Dr. Wolfgang Lehner, Technische Universität Dresden

Novel application areas like interactive data science over large datasets, which go far beyond traditional transactional data processing techniques, are questioning many longstanding architectural assumptions found in many database systems. Since the typical combination of data-intensive as well as compute-intensive workloads require extreme performance, fully exploiting modern as well as non-standard hardware components reflects a well-established research direction. However, within this presentation, I am arguing that the design space of modern database systems is not only one-dimensional with respect to high performance but has to consider multiple factors to come up with a well-balanced solution beyond performance only. Within this presentation, I will touch two aspects and discuss problems as well as developed solutions in detail.

As a first example, I will address energy consumption as a growing limitation for scalable main-memory database systems, which make heavy use of the main power consumers - processors and main memory. While modern processors provide a rich set of energy-control features, they lack the capability of controlling them appropriately because of missing application-specific knowledge. I will therefore present the concept of an energy-control loop as a DBMS-integrated approach for adaptive energy-control on scale-up in-memory database systems that obeys a query latency limit as a soft constraint and actively optimizes energy efficiency and performance of the DBMS. Experiments in the context of the ERIS research database kernel show energy savings ranging from 20% to 40% for real-world workload profiles.

As a second example, I will focus on resilience mechanisms in main-memory database systems. While recent studies have shown that all future hardware components become less and less reliable in total and multi-bit flips are occurring regularly rather than exceptionally, hardware-based protection techniques will eventually become not affordable. An emerging research direction is employing application-specific protection techniques in higher software layers. I will present AHEAD, an adaptive, on-the-fly hardware error detection approach for main-memory column-store systems. AHEAD uses an arithmetic error coding technique, which allows on-the-fly detection of errors during query processing. Results of the evaluation clearly show the benefits of AHEAD with respect to traditional resilience mechanisms.

Short Bio:

Wolfgang Lehner is full professor and head of the Database Technology Group as well as director of the Institute for System Architecture at TU Dresden, Germany. His research focuses on database system architecture specifically looking at crosscutting aspects from algorithms down to hardware-related aspects in main-memory centric settings. He is part of TU Dresden's research cluster of excellence with topics in energy-aware computing, resilient data structures on unreliable hardware, and orchestration of widely heterogeneous systems. He is heading a Research Training Group on large-scale adaptive system software design and acts as a principal investigator in Germany's national "Competence Center for Scalable Data Services and Solutions" (ScaDS). Wolfgang also maintains a close research relationship with the international SAP HANA development team. He serves the

community in many PCs, is an elected member of the VLDB Endowment, is chairing the review board of Computer Science within the German Research Foundation (DFG), and is an appointed member of the Academy of Europe.