Query Processing on modern CPUs

[Abstract]

Johann-Christoph Freytag
Humboldt-Universität zu Berlin,
Germany
freytag@informatik.hu-berlin.de

.

ABSTRACT

This talk consists of two parts both relating to the challenge how to better take advantage of fine-grain parallelism that comes with today's modern CPUs. After a general introduction about the changes in hardware over the last two decades, we show in the first part of the talk how to accelerate the processing of tree-based index structures by using SIMD instructions. We adapt the B+-Tree and prefix B-Tree (trie) by changing the search algorithm on inner nodes from binary search to k-ary search. We develop adaptations of tree structures that satisfy the specific constraints of SIMD instructions. We present algorithms for transforming the original tree layout into a SIMD-friendly layout. Our adapted B+-Tree speeds up search processes by a factor of up to eight for small data types compared to the original B+-Tree using binary search. Furthermore, our adapted prefix B-Tree enables a high search performance even for larger data types. The second part of this talk focuses on the problem of how to find the best partitioning of a given query execution plans and data into tasks to perform an optimal (query) execution on multiple CPU cores. Therefore, we first present a classification of existing approaches in various DBMSs as a basis to develop a generic Query Task Model (QTM) for query execution. This model opens up a design space for scheduling parallel task execution thus making different existing approaches comparable. Based on QTM we show that existing execution schedules do not guarantee the fastest execution on multiple cores – at the same time it allows us to characterize those best (fastest) execution schedules based on (the ratio of) data locality and instruction locality. This work was done together with Steffen Zeuch and Frank Huber.

About the Author

Johann-Christoph Freytag is currently full professor for Databases and Information Systems (DBIS) at the Computer Science Department of the Humboldt-Universität zu Berlin, Germany. Before joining the department in 1994, he was a research staff member at the IBM Almaden Research Center (1985-1987), a researcher at the European Computer-Industry-Research Centre (ECRC, in Munich, Germany, 1987-1989), and the head of Digital's Database Technology Center (also in Munich, 1990-1993). He holds a Ph.D. in Applied Mathematics/Computer Science from Harvard University, MA. Freytag's research interests include all aspects of query processing and query optimization in object-relational database systems, new developments in the database area (such as semi-structured data, data quality, databases and security), privacy in database systems, and data quality as well as applying database technology to applications such as GIS, genomics, and bioinformatics/life science. In the last years he received the IBM Faculty Award four times for collaborative work in the areas of databases, middleware, and bioinformatics/life science, as well as the HP Innovation Award of excellent cooperation in the area of databases and workflow systems. He organized the VLDB conference in Berlin in 2003 and was a member of the VLDB Endowment (2001-2007). From 2009 to 2015 Freytag headed the German database interest group of the (Fachbereich DBIS, Gesellschaft für Informatik). Since 2015 he is a member of the Extended Executive Board of the GI.

.

29th GI-Workshop on Foundations of Databases (Grundlagen von Datenbanken), 29.05.2017 - 02.06.2017, Blankenburg/Harz, Germany.