

New World Record in Sorting Huge Amounts of Data

Computer Scientists of the KIT Develop a New, More Scalable Method of Sorting Data



With a cluster of 200 compute nodes, computer scientists from Karlsruhe have broken a world record in sorting huge amounts of data. (Photo by: Rolf Mayer)

Scientists of the Karlsruhe Institute of Technology (KIT) have developed a new, more scalable method of sorting huge amounts of data. With this, they beat the sorting record of the Massachusetts Institute of Technology (MIT) with even lower hardware effort.

Computers connected over the internet produce ever-growing amounts of data. For further processing, they usually need to be aligned first, i.e. sorted by a certain criterion. Efficient sorting of data is essential to search engines or databases and therefore a very important topic in theoretical as well as practical computer science.

Dr. Elisabeth Zuber-Knost
Press Officer

Kaiserstraße 12
76131 Karlsruhe, Germany
Phone: +49 721 608-7414
Fax: +49 721 608-3658

For further information, please contact:

Klaus Rümmele
Public Relations and Marketing
(PKM)
Phone: +49 721 608-8153
Fax: +49 721 608-5681
E-mail: klaus.ruemmele@kit.edu

Detailed information on the
SortBenchmark:
<http://www.sortbenchmark.org>

The well-established SortBenchmark, a table compiled by experts from companies like Microsoft and Hewlett-Packard and published on the internet, documents the current records in sorting. Participating in the most prestigious category requires the sorting of at least 10^{12} data records, i.e. 100 Terabyte in total.

A team of researchers lead by Professor Peter Sanders from the Institute for Theoretical Computer Science has now taken the lead in two categories of SortBenchmark. The scientists, apart from Sanders, Dr. Mirko Rahn, Johannes Singler, and Tim Kieritz, sorted 100 Terabyte of data in a little less than three hours, which corresponds to 564 GB per minute. For this purpose, they used a cluster of 200 compute nodes configured by staff of the KIT Steinbuch Centre for Computing (SCC). A team of the internet giant Yahoo reached a marginally better result, but used more than 17 times the amount of compute nodes.

The KIT researchers also raised the amount of data sorted per minute to 950 Gigabyte. This is more than three times the last record held by MIT with a much larger machine. Also the new value reported by Yahoo in this category was smaller by a factor of two. In addition, the scientists from Karlsruhe beat the record of 68 seconds reached by Google in November 2008 for the rapid sorting of a billion bytes of data by 64 seconds, again using much less hardware.

This lead of the team from Karlsruhe, explains Peter Sanders, “is mostly due to a new technique which brings the amount of hard disk accesses and the network communication volume required close to the smallest possible value.” This sorting algorithm also is more robust than most competing techniques, since it guarantees high efficiency for any kind of input. According to Sanders, “the implementation is also highly efficient. Each compute node’s four hard disks and eight processor cores can be exploited in an efficient way.” This is enabled by software libraries developed by the institute.

The Karlsruhe Institute of Technology (KIT) is the merger of the Forschungszentrum Karlsruhe, member of the Helmholtz Association, and the Universität Karlsruhe. This merger will give rise to an institution of internationally excellent research and teaching in natural and engineering sciences. In total, the KIT has 8000 employees and an annual budget of 700 million Euros. The KIT focuses on the knowledge triangle of research – teaching – innovation.

The Karlsruhe institution is a leading European energy research center and plays a visible role in nanosciences worldwide. KIT sets new standards in teaching and promotion of young scientists and attracts top scientists from all over the world. Moreover, KIT is a leading innovation partner of industry.

This press release is available on the internet under www.kit.edu.

The photo of printing quality may be requested by e-mail presse@verwaltung.uni-karlsruhe.de or phone +49 721 608-7414.