

Neuer Weltrekord im Sortieren riesiger Datenmengen

Informatiker am KIT entwickeln neues, robusteres Verfahren



Mit einem Cluster, der über 200 Rechenknoten verfügt, haben Karlsruher Informatiker einen Rekord im Sortieren sehr großer Datenmengen aufgestellt. (Foto: Rolf Mayer)

Wissenschaftler am Karlsruher Institut für Technologie (KIT) haben ein neues, robusteres Sortierverfahren für sehr große Datenmengen entwickelt. Damit übertreffen sie den Rekord des Massachusetts Institute of Technology (MIT) sogar bei geringem Hardwareaufwand.

Über das Internet vernetzte Rechner erzeugen immer größere Datenmengen. Um diese auswerten zu können, muss man sie zunächst nach einem bestimmten Kriterium ordnen. Das effiziente Sortieren von Daten ist von zentraler Bedeutung für Suchmaschinen oder Datenbanken - und damit ein wichtiges Forschungsthema in der theoretischen wie auch in der praktischen Informatik.

Dr. Elisabeth Zuber-Knost
Pressesprecherin

Kaiserstraße 12
76131 Karlsruhe
Tel.: +49 721 608-7414
Fax: +49 721 608-3658

Weiterer Kontakt:

Klaus Rümmele
Stabsabteilung Presse,
Kommunikation und Marketing
(PKM)

Tel.: +49 721 608-8153

Fax: +49 721 608-5681

E-Mail: klaus.ruemmele@kit.edu

Nähere Informationen zum Sortierbenchmark:

<http://www.sortbenchmark.org>

Der seit Jahren etablierte SortBenchmark, eine im Internet veröffentlichte Tabelle, über die Fachleute etwa von den Unternehmen Microsoft und Hewlett-Packard entscheiden, verzeichnet die jeweils aktuellen Rekorde im Sortieren. In der Königsdisziplin müssen mindestens 10^{12} Datensätze, also insgesamt 100 Terabyte sortiert werden.

Ein Forscherteam um Professor Peter Sanders am Institut für Theoretische Informatik hat sich nun in zwei Kategorien des SortBenchmark gegen die Konkurrenz durchgesetzt. So sortierten die Wissenschaftler, neben Sanders Dr. Mirko Rahn, Johannes Singler und Tim Kieritz, 100 Billionen Byte Daten in etwas weniger als drei Stunden, was einem Durchsatz von 564 GB pro Minute entspricht. Dafür nutzten sie einen Computerverbund mit 200 Rechenknoten, den Mitarbeiter des Steinbuch Centre for Computing (SCC) am KIT konfiguriert hatten. Ein Team des Internet-Giganten Yahoo schaffte zwar einen minimal besseren Wert, nutzte dafür aber mehr als 17mal so viele Rechenknoten.

Die KIT-Forscher erhöhten außerdem die Rekordzahl an Datensätzen, die in unter einer Minute sortiert werden können, auf 9,5 Milliarden (950 GB). Das ist mehr als das Dreifache des bisher vom MIT gehaltenen Rekords, der zudem auf einer größeren Maschine erzielt worden war. Auch der von Yahoo in dieser Kategorie neu eingereichte Wert lag um den Faktor zwei niedriger. Zudem verbesserten die Karlsruher Wissenschaftler einen von Google im November 2008 aufgestellten Rekord beim schnellen Sortieren von einer Billion Byte Daten Wert von 68 Sekunden auf 64 Sekunden, wiederum mit viel geringerem Hardware-Aufwand.

Der Vorsprung des Karlsruher Teams, so erläutert Peter Sanders, „ergibt sich vor allem aus einem neuen Verfahren, das sowohl die Anzahl der Festplattenzugriffe als auch die erforderliche Netzwerkkommunikation in die Nähe der minimal denkbaren Werte rückt.“ Der Sortieralgorithmus sei zudem robuster als die meisten Konkurrenzverfahren, weil er gute Leistung für beliebige Eingabedatensätze garantiere. Hinzu komme, so Sanders, „eine besonders effiziente Implementierung, welche die jeweils vier Festplatten und acht Prozessor-Kerne jedes Rechenknotens sehr gut auslasten kann.“ Ermöglicht hätten dies am Institut entwickelte Software-Bibliotheken.

Im Karlsruher Institut für Technologie (KIT) schließen sich das Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft und die Universität Karlsruhe zusammen. Damit wird eine Einrichtung international herausragender Forschung und Lehre in den Natur- und Ingenieurwissenschaften aufgebaut. Im KIT arbeiten insgesamt 8000 Beschäftigte mit einem jährlichen Budget von 700 Millionen Euro. Das KIT baut auf das Wissensdreieck Forschung – Lehre – Innovation.

Die Karlsruher Einrichtung ist ein führendes europäisches Energieforschungszentrum und spielt in den Nanowissenschaften eine weltweit sichtbare Rolle. KIT setzt neue Maßstäbe in der Lehre und Nachwuchsförderung und zieht Spitzenwissenschaftler aus aller Welt an. Zudem ist das KIT ein führender Innovationspartner für die Wirtschaft.

Diese Presseinformation ist im Internet abrufbar unter: www.kit.edu

Das Foto kann in druckfähiger Qualität angefordert werden unter: presse@verwaltung.uni-karlsruhe.de oder +49 721 608-7414.