

Zwei neue Systeme am LRZ

Das Leibniz-Rechenzentrum hat zwei neue Systeme für Big Data und Machine Learning in Betrieb genommen: Teramem1 und DGX-1. Für Big-Data-Anwendungen stehen nun mit vier Broadwell-Prozessoren mit insgesamt 96 Rechenkernen 6 Terabyte Cache-kohärenter Speicher zur Verfügung. Das System ist besonders für Anwendungen wie Genom-Analysen, die Durchforstung großer Datenbanken oder die Auswertung riesiger Datenbestände – wie sie zum Beispiel auf den großen Supercomputern am LRZ anfallen – geeignet, die wenige Rechenkerne aber riesigen Hauptspeicher benötigen. Da Teramem1 in das Linux-Cluster des LRZ integriert ist, steht das gesamte dortige Software-Angebot zur Verfügung. Das Machine-Learning-System DGX-1 ist ein „Supercomputer in a box“ mit acht hochwertigen NVIDIA-Tesla-P100-Grafikprozessoren. Die 28.672 NVIDIA-CUDA-Recheneinheiten sind über einen NVLink Interconnect verbunden und können 170 TeraFlop/s Rechenleistung erbringen. Es verfügt über 16 Gigabyte Speicher und besitzt einen „Host-Prozessor“ auf Intel X86-Basis mit 40 Kernen. DGX-1 kann flexibel mit Docker-Images für die jeweiligen Anwendungen bestückt werden. Besonders oft nachgefragte Softwarepakete wie TensorFlow, Theano, CNTK, Torch, Digits und Caffe sind bereits vorinstalliert. Das Betriebsmodell ist so dynamisch wie das gesamte Forschungsgebiet. Einzelne Nutzer können das gesamte DGX-1-System exklusiv für sich reservieren, um ihre komplexen Machine-Learning-Aufgaben interaktiv zu bearbeiten zum Beispiel riesige neuronale Netze trainieren, die anschließend auf wesentlich weniger leistungsfähiger Hardware zur Mustererkennung eingesetzt werden können. (Kontakt: [Ludger Palm](#), LRZ)

Dresden als Schmiede digitaler Zukunftsindustrien

Nachdem auf dem Nationalen IT-Gipfel im November 2016 erste Digital Hubs vorgestellt und Dresden und Stuttgart als mögliche weitere Standorte genannt wurden, hat die Sächsische Staatsregierung – gemeinsam mit der TU Dresden und dem Branchenverband Silicon Saxony – am 10. Januar offiziell ihr Konzept präsentiert. Mit der Digital-Hub-Initiative will das BMWi die Entstehung von Zentren unterstützen, in denen sich deutsche und internationale Gründer, Wissenschaftler, Investoren und etablierte Unternehmen gegenseitig bei der digitalen Transformierung stärken. Dresden hofft, mit dem „Smart Systems Hub – Enabling IoT“ (Internet der Dinge) in den Kreis der bis zu zwölf geplanten Digital Hubs aufgenommen zu werden. Hier sollen digitale Zukunftsindustrien weiterentwickelt und die Digitalisierung der

Leitindustrien in Deutschland vorangetrieben werden. Insbesondere sollen Entwicklungen in Maschinen- und Fahrzeugbau, Halbleitertechnologie und Mikroelektronik durch den Einsatz von Software- und IT-Lösungen in die nächste Generation überführt und damit neue Märkte definiert und gestaltet werden. Gemeinsam mit dem Lehmann-Zentrum – als akademischem Nukleus und IT-Kompetenzzentrum der TU Dresden – und der Errichtung eines DLR-Software-Instituts unterstreicht die Planung des „Smart Systems Hub“ die Exzellenz des Standorts für das Thema Digitalisierung. An der TU Dresden gibt es seit Jahren erfolgreiche Aktivitäten in den Bereichen Softwareentwicklung, Hochleistungsrechnen, Big Data, 5G-Mobilfunk- und Netztechnologie, Internet der Dinge oder Sicherheit im Internet, auf denen das DLR-Institut und das Digital Hub aufbauen können. Weitere Informationen: <https://www.staatsregierung.sachsen.de/kompetenz-und-innovationszentrum-smart-systems-hub-4256.html> (Kontakt: [Wolfgang E. Nagel](#), ZIH)

Energieeffizientes HPC mit Custom Computing

Zur Steigerung der Energieeffizienz von HPC-Systemen werden diese zunehmend mit Hardwarebeschleunigern ausgerüstet. Die Nutzung von Beschleunigern hinkt deren prinzipiellem Performance- und Effizienz-Potenzial jedoch hinterher. Im DFG-Projekt „PerficienCC – Performance und Effizienz in HPC mit Custom Computing“ wird das Paderborn Center for Parallel Computing (PC²) diese Lücke verkleinern. Der Fokus liegt auf anwendungsspezifischen Hardwarebeschleunigern, sogenannten FPGAs. In einer engen Kooperation von FPGA-Experten des PC² und Entwicklern wissenschaftlicher Codes an der Universität Paderborn werden das Potenzial zur Beschleunigung wichtiger Anwendungen untersucht und gezielte Beiträge zur Portierung von Open-source-Anwendungs-codes geleistet. Generalisierungen dieser Anwendungs-codes werden darüber hinaus in Bibliotheken zur Verfügung gestellt sowie für Schulungen und Beispielmateriale aufbereitet. Dadurch wird die Zugänglichkeit von FPGA-Technologie für Entwickler erhöht und der Nutzen von FPGAs empirisch bewertet. Weitere Informationen: <http://pc2.uni-paderborn.de/perficiencc> (Kontakt: [Christian Plessl](#), PC²)

Metaprogrammierung für heterogene verteilte Systeme

Jeder einzelne Rechenknoten moderner HPC-Systeme ist inzwischen derart komplex, dass verschiedene Speicher- und Parallelitätsebenen gezielt berücksichtigt werden müssen, um die Rechenleistung optimal zu nutzen. Die gängigen Hardware-Konzepte unterscheiden sich nicht nur deutlich voneinander, sondern erfordern auch

komplett unterschiedliche Programmiermodelle für die Anwendungssoftware. Dazu kommt, dass die Systeme aus immer mehr homogenen oder heterogenen Rechenknoten bestehen, deren Zusammenarbeit für die effiziente Lösung großer Probleme, die die gesamte Rechenleistung des Systems erfordern, entscheidend ist. Auch hier gibt es sehr unterschiedliche Konzepte, die je nach Anwendung und System als optimal gelten. Das im 5. HPC-Call des BMBF geförderte neue Verbundprojekt „Mephisto – Metaprogrammierung für heterogene verteilte Systeme“ wird es erlauben, Programme zu entwickeln, die auf allen gängigen Hochleistungsrechnern ohne Veränderung des Quellcodes eine hohe Performance erreichen. Darüber hinaus soll die Verteilung der Daten über den gesamten Hochleistungsrechner automatisiert werden. Mephisto lehnt sich dabei bewusst an die Datencontainer und Algorithmen des C++-Standards an, sodass auch Anwender ohne spezielle Kenntnisse der Parallelprogrammierung in der Lage sind, das Leistungspotenzial auszuschöpfen. Neben dem ZIH der TU Dresden als Projektleiter, sind die Ludwig-Maximilians-Universität München und das Helmholtz-Zentrum Dresden-Rossendorf Partner in diesem Vorhaben. Das Projekt ist zum Februar 2017 mit einer Laufzeit von drei Jahren gestartet. (Kontakt: [Andreas Knüpfer](#), ZIH)

Kerncraft: Performance-Modellierung von Schleifen

Die analytische Modellierung der Performance von Schleifenkonstrukten ermöglicht wichtige Einsichten in Engpässe bei der Programmausführung. Analytisch bedeutet hier, dass ein vereinfachtes Modell der Hardware Instruktionen unter idealisierenden Annahmen ausführt. Im besten Fall können damit die Laufzeit sowie die Skalierbarkeit vorhergesagt werden. Am Regionalen Rechenzentrum Erlangen (RRZE) wird seit 2015 das Softwarewerkzeug „Kerncraft“ entwickelt, das für bestimmte Schleifenkonstrukte eine automatisierte Erstellung analytischer Performancemodelle erlaubt. Unterstützt werden das bekannte Roofline-Modell für Engpässe bei der Ausführung von Instruktionen bzw. bei der Datenübertragung sowie das „Execution-Cache-Memory“ (ECM)-Modell, das die Wechselwirkung zwischen Daten und Code beschreibt. Die notwendigen Hardware-Informationen wie Bandbreiten und Cache-Größen werden automatisch erfasst oder manuell bereitgestellt. Kerncraft verfügt über verschiedene Module, die bestimmte Aspekte der Programmausführung modellieren. Hervorzuheben sind das Layer-Condition-Modul zur analytischen Vorhersage optimaler Blockgrößen bei Stencil-Algorithmen und der Cache-Simulator, auf den in komplexeren Fällen zurückgegriffen werden kann. Kerncraft wird seit Anfang 2017 im Rahmen des vom BMBF geförderten Projekts „Metacca – Metaprogramming for Accelerators“ mit dem Ziel weiterentwickelt, sowohl die unterstützten Architekturen zu erweitern als auch eine Schnittstelle zur „Low-Level Virtual Machine“ (LLVM)-Infrastruktur zu schaffen, die modernen Compilern zugrunde liegt. Zum Konsortium gehören neben der Friedrich-Alexander-Universität Erlangen-Nürnberg die Universität

des Saarlandes, die Universität Mainz und das Deutsche Forschungszentrum für Künstliche Intelligenz. Weitere Informationen: <https://github.com/RRZE-HPC/kerncraft> (Kontakt: [Gerhard Wellein](#), FAU)

Energieeffizienz im Rechenzentrum der TU Dresden

Im Rechenzentrumsneubau der TU Dresden wurde im Januar 2017 erstmals für einen Monat eine Energy Reuse Efficiency (ERE) von dauerhaft unter 1,0 gemessen (Durchschnitt: 0,89, Spitzenwert: 0,79). Die ERE-Metrik erweitert die etablierte PUE-Kennzahl (Power Usage Efficiency) zur Bewertung der Energieeffizienz von Rechenzentren um die Berücksichtigung der Abwärmenutzung. Ein ERE-Wert von unter 1,0 bedeutet, dass die zu Heizzwecken nachgenutzte Wärmeenergiemenge aus dem Rechenzentrum größer ist als die für den Betrieb von Infrastruktur-Komponenten wie USV-Anlagen und Kältemaschinen erforderliche Elektroenergie. Maßgeblich für diesen Erfolg waren die hohe Auslastung des warmwasser-gekühlten Hochleistungsrechners Taurus sowie die hohe Wärmeabnahme eines erst kürzlich an die TU Dresden übergebenen Gebäudes, das die Rechnerabwärme zur Gebäudeheizung nachnutzt. (Kontakt: [Daniel Hackenberg](#), ZIH)

Hessisches HPC trifft sich in Frankfurt

Vom 13. bis 16. März findet der 7. High Performance Computing Hessen Workshop (HiPerCH) in Frankfurt statt. In zwei Modulen werden Grundlagen zur Nutzung von Hochleistungsrechnern sowie die Nutzung des Werkzeugs zur Fehlererkennung TotalView vermittelt. HiPerCH ist insbesondere für die Nutzer/-innen hessischer Hochleistungsrechner sowie für die im Rahmen der DFG-Kooperation EPE (Enabling Performance Engineering) assoziierten Mitglieder konzipiert. Restplätze werden auch an auswärtige Bewerber/-innen vergeben. Weitere Informationen: <https://www.hpc-hessen.de/beratung/regelmaessige-veranstaltungen/hiperch-7/> (Kontakt: [Alexandra Feith](#), HKHLR)

Veranstaltungen

- 09.-10.03.2017: [Introduction to Parallel Programming with HPX](#), GCS@HLRS, Stuttgart
- 13.-15.03.2017: [Parallel I/O and Portable Data Formats](#), GCS@JSC, Jülich
- 16.-17.03.2017: [Vectorisation and portable programming using OpenCL](#), GCS@JSC, Jülich
- 20.-24.03.2017: [Parallel Programming in Computational Science 2017](#), IT Center, RWTH Aachen University
- 20.-24.03.2017: [Fortran for Scientific Computing](#), GCS@HLRS, Stuttgart
- 27.-31.03.2017: [25th VI-HPS Tuning Workshop](#), IT Center RWTH Aachen University
- 27.-31.03.2017: [Iterative Gleichungssystemlöser und Parallelisierung](#), GCS@HLRS, Stuttgart