GA Gauß-Allianz

Infobrief
Nr. 113: Dezember 2023/Januar 2024

Top500 des Supercomputing Herbst 2023

Die 62. Top500-Liste der weltweit schnellsten Supercomputer zeigt im Vergleich zu der im Juni 2023 erschienenen Liste einige Veränderungen im Ranking. Das von HPE Cray am Oak Ridge National Laboratory in Tennessee, USA, installierte Frontier-System behält seinen Spitzenplatz mit einer unveränderten Linpack-Rechenleistung von 1,194 ExaFlop/s und ist nach wie vor der einzige Exascale-Rechner auf der Liste. Das neue Aurora-System der Argonne Leadership Computing Facility in Illinois, USA, erreichte mit einer Rechenleistung von 585,34 PetaFlop/s den 2. Platz, der zuvor von Fugaku gehalten wurde. Das schnellste europäische System auf dem 5. Platz der Top500-Liste ist das ebenfalls von HPE Cray gelieferte System LUMI mit einer um 23% gesteigerten Gesamtleistung von 379,7 PetaFlop/s am finnischen CSC – IT Center for Science. In Deutschland sind die Spitzenreiter JUWELS Booster (44,1 PetaFlop/s) am FZJ auf Platz 18, SuperMUC-NG (19,5 PetaFlop/s) am LRZ auf Platz 40 sowie Hawk am HLRS (19,3 PetaFlop/s) auf Platz 42. Insgesamt zählt die aktuelle TOP500-Liste 36 deutsche Systeme. Die GA-Mitglieder sind mit 30 Systemen in der Liste vertreten; 9 davon unter den Top 100. Unter den Top 10 der Green500-Liste, die die Energieeffizienz der Systeme vergleicht, besetzt das von der Universität Frankfurt betriebene System Goethe-NHR den 9. Platz. Weitere Informationen: top500.org/lists/top500/2023/11/.

Erweitertes KI-Angebot am LRZ

Der SuperMUC des LRZ wurde erweitert. Die neue Installation SuperMUC-NG Phase 2, kurz SNG-2, beschleunigt das HPC für KI-Anwendungen in Wissenschaft und Forschung. In der aktuellen Top500-Liste ist das System das viertschnellste in Deutschland. SNG-2 wurde entwickelt, um stärker nachgefragte KI-Methoden in etablierte HPC-Workflows zu integrieren. Der von Intel und Lenovo entwickelte und gelieferte Supercomputer enthält 240 Rechenknoten und verfügt über 480 Intel-CPUs sowie 960 Intel-GPUs, die durch Lenovos Neptune-Technologie direkt mit Heißwasser gekühlt werden. Darüber hinaus verfügt der Supercomputer über ein verteiltes, asynchrones Objektspeichersystem (DAOS), das Intel-Prozessoren des Typs Xeon Scalable und den persistenten Speicher Optane von Intel nutzt, um den Zugriff auf große Datenmengen zu beschleunigen. Um eine breite Nutzung sicherzustellen, setzt das LRZ auf ein zusätzliches Kursangebot in Kombination mit einem erweiterten Expert:innen-Team. Der allgemeine Zugang für die Nutzenden wird für Frühjahr 2024 erwartet. Rechenzeit wird über das Gauss Centre for Supercomputing (GCS) vergeben. Weitere Informationen: lrz.de/presse/ereignisse/2023-11-13-SNG-2-installation-de/. (Kontakt: Herbert Huber, LRZ@GCS)

Goethe-NHR: Platz 9 der Green500

Seit Mai 2023 belegt das HPC-System Goethe-NHR der Goethe-Universität Frankfurt mit 46,5 GigaFlops/Watt Platz 9 der Energieeffizienz-Weltrangliste. Damit ist er der energieeffizienteste Rechner Deutschlands. Mit 9 PetaFlop/s Rechenleistung belegt er Platz 84 der Top500. Der Goethe-NHR implementiert 880 AMD MI210 GPUs mit jeweils 8 GPUs in einem Server. Insgesamt verwenden mehr als 50 Nutzergruppen die Systeme der Goethe-Universität. Es wurden viele Anwendungen für GPUs optimiert, wobei erhebliche Geschwindigkeitssteigerungen und Verbesserungen der Energieeffizienz erreicht wurden. Beispiele sind die Ereignisrekonstruktion in der Teilchenphysik (10.000x), Gitter-Eichrechnungen (44x), Hybrid-Quantenmolekulardynamik (160x) und Berechnung der Bayes'schen Inferenz von elektronenmikroskopischen Bildern (100x). Weitere Informationen: nhrsw.de/. (Kontakt: Volker Lindenstruth , JWG)

Energieeffiziente HPC-Systeme Made in Germany

Sei es in den Ingenieurwissenschaften, in Chemie, Medizin oder im Bereich Cloud-Systeme und KI: Um die Wirklichkeit möglichst genau digital abzubilden, werden leistungsfähige Supercomputer benötigt. Bei immer kleiner werdenden Technologien in der Chipentwicklung erhöht sich die Leistungsaufnahme dieser Systeme, wobei die schnellsten bereits bei über 25 MW liegen. Beschleunigerhardware bietet hier Leistungs- und Effizienzvorteile für parallele Anwendungen. Mit der Entwicklung eines Stencil- und Tensorbeschleunigers (STX) arbeitet das Fraunhofer ITWM zusammen mit den Instituten IIS und IZM an einem Architekturansatz, der hohe Leistung mit sehr guter Energieeffizienz und realistischen Kosten für die Entwicklung verbindet. Das System wird damit 3- bis 4-mal effizienter als GPU-basierte Systeme sein und soll demnächst in Pilotsystemen zur Verfügung stehen. Im Projekt STXDemo kooperieren die Fraunhofer Institute mit der Zollner AG und den Rechenzentren der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau (RPTU) sowie dem FZJ, um eine auf PCI Gen5 basierende Beschleunigerkarte und das Systemkonzept umzusetzen. Ihr Ziel: Erste Pilotsysteme mit Schlüsselanwendungen aus der Quantenchromodynamik sowie der Strömungsdynamik in 2025. Weitere Informationen: itwm.fraunhofer.de/stx. (Kontakt: Jens Krüger, ITWM)

Heterogener Speicher für optimierte Datenplatzierung

Im Projekt „Heuristics for Heterogeneous Memory“ (H2M) erforschen die RWTH Aachen und der französische Partner Inria die Integration neuer Speichertechnologien wie High Bandwidth Memory (HBM) und Non Volatile Memory (NVM) in HPC-Systeme. Im Vergleich zum herkömmlichen Dynamic Random Access Memory (DRAM) bieten diese Technologien unterschiedliche Eigenschaften und Kapazitäten. Die Kooperation zielt darauf ab, portable Schnittstellen zu entwickeln, um verfügbare Speicher zu identifizieren und den Zugriff darauf zu ermöglichen. Durch die Entwicklung von Zuweisungsabstraktionen und Heuristiken sollen Anwendungsentwickler:innen mehr Kontrolle über die Datenplatzierung erhalten und die Möglichkeit, geeignete Speicher für Datenobjekte zur Laufzeit auszuwählen. Ein automatisierter Workflow verbessert darüber hinaus die optimale Positionierung von Datenobjekten in bestehenden Anwendungen. Aktuelle Forschung untersucht, wie statische Platzierung bzw. dynamische Verschiebung von Daten zwischen Speichertypen zur Laufzeit die Leistung beeinflussen. Weitere Informationen: h2m.gitlabpages.inria.fr/. (Kontakt: Jannis Klinkenberg, IT Center)

Kölner HPC-Zentrum unterstützt Krebsforschung

Mit fast 800 neu diagnostizierten Fällen pro Jahr in Deutschland ist das kleinzellige Bronchialkarzinom (SCLC) der aggressivste Krebs-Subtyp mit starker Therapieresistenz und geringer Überlebenswahrscheinlichkeit. Der interdisziplinäre SFB1399 versammelt Expertise aus Kliniken, Immunologie, Biochemie, Bioinformatik und der IT, um die Medikamentenresistenz zu verstehen und die Erkenntnisse in klinische Anwendungen zu überführen, die das Überleben von SCLS-Patient:innen verbessern. DNA-Schäden durch äußere Einflüsse wie UV-Strahlung oder Tabakkonsum führen zu Mutationen und können in Tumorbildung und Krebs münden. Durch die HPC-Infrastruktur und Expertise des Regionalen Rechenzentrums zu Köln (RRZK) zur Optimierung der Programme für die Charakterisierung von Mutationen, die Analyse von Abfolgen der Mutationsereignisse wie auch für die Untersuchung von Medikamentenempfindlichkeiten werden wesentliche Beiträge erbracht. Die gewonnenen Erkenntnisse sind zudem in das Design des nächsten Supercomputers RAMSES eingeflossen, der 2024 am RRZK in Betrieb gehen und hohe Standards in Energieeffizienz und Datenschutz erfüllen wird. Weitere Informationen: rrzk.uni-koeln.de/hpc-projekte/projekte/sfb-1399. (Kontakt: Viktor Achter, RRZK)

Dateneffizientes maschinelles Lernen für RNA

In Zusammenarbeit mit dem FZJ und DLR hat das SCC am KIT in der Zeitschrift Communications Biology eine Studie veröffentlicht, die moderne und klassische Ansätze des tiefen maschinellen Lernens (Deep-Learning) kombiniert, um auch für datenarme Szenarien Vorhersagemodelle entwerfen zu können. Das vorgeschlagene Modell, BARNACLE, verwendet einen Deep-Learning-Ansatz in der biologischen Strukturvorhersage: Für die Vorhersage räumlicher Nachbarschaften zwischen RNA-Bausteinen, nutzen die Forschenden die Tatsache, dass Nukleotide, die in der RNA räumlich nahe beieinander liegen, während der Evolution auch mit größerer Wahrscheinlichkeit gemeinsam mutieren. Genau nach diesen Mutationsmustern sucht das Modell, dessen Trainingsgrundlage eine Kombination aus selbstüberwachtem Vortraining auf vielen Sequenzdaten und die effiziente Nutzung der wenigen Strukturdaten ist. Mit diesem Ansatz konnte eine signifikante Verbesserung gegenüber klassischen statistischen Ansätzen sowie anderen neuronalen Netzen gezeigt werden. Die Methode ist zudem auf verwandte Aufgaben mit ähnlichen Datenbeschränkungen übertragbar. Weitere Informationen: scc.kit.edu/ueberuns/17136. (Kontakt: Markus Goetz, KIT)

NHR-Computational Physics Symposium

Am 2. und 3. November fand das zweite Computational Physics Symposium des NHR statt. Organisiert wurde die Veranstaltung vom Computational Physics Center des NHR-Verbundes. Das Symposium zielte darauf ab, Forschungsziele und numerische Herausforderungen in der Computerphysik zu präsentieren. Am ersten Tag standen Sitzungen zur Physik der kondensierten Materie und Materialwissenschaft auf dem Programm, während der zweite Tag der Elementarteilchenphysik, der Astroteilchenphysik und der Quantenchromodynamik gewidmet war. Expert:innen aus verschiedenen Bereichen der Computerphysik stellten ihre wissenschaftlichen Arbeiten vor und nutzten das Symposium als Plattform für den Austausch und zur Zusammenarbeit auf ihrem Gebiet. Insgesamt hat das Symposium gezeigt, dass das NHR-Zentrum eine führende Rolle in der Computational Physics Community einnimmt. Geplante Aktivitäten wie Benchmarking, Optimierung, Portierung von Anwendungscodes sowie die Identifizierung, Sammlung und Analyse wichtiger HPC-Workloads unterstreichen seine zukunftsorientierte Ausrichtung. Weitere Informationen: pc2.de/go/nhr-computational-physics-2023. (Kontakt: Lukas Mazur, NHR@PC²)

Veranstaltungen

- 13.12.2023: Introduction to LRZ HPC Systems with Focus on CFD Workflows, LRZ, online
- 23.–25.01.2024: Hybrid Programming in HPC - MPI+X, HLRS

HPC-Kalender der Gauß-Allianz:

hpc-calendar.gauss-allianz.de