GA Gauß-Allianz

Infobrief
Nr. 115: März 2024

Neuer Supercomputer im PC² Paderborn

Die Universität Paderborn hat mit pro-com Datensysteme GmbH einen Vertrag über die Installation und Wartung eines HPC-Systems auf Basis der nächsten AMD-Prozessorgeneration abgeschlossen. Mit dem vom PC² betriebenen Rechner werden über 740 Knoten der Firma Lenovo, aufgeteilt in CPU-, GPU- und FPGA-Partitionen, zur Verfügung stehen. Jeder Rechenknoten ist mit zwei Prozessoren und leistungsstarken Prozessorkernen ausgestattet und verfügt durch DRAM-Module über eine sehr hohe Speicherbandbreite. Die GPU-Partition wird 27 Knoten mit jeweils 4 GPUs des Typs H100 von NVIDIA umfassen; 32 Knoten sind für FPGA-Karten vorbehalten. Das über NDR-InfiniBand verbundene System bietet 5 Petabyte Speicher des Typs IBM Spectrum Scale. Die Abwärme des HPC-Systems wird zu 97% über Warmwasserkühlung abgeführt. Die darüber erreichten Rücklauftemperaturen von über 45°C ermöglichen eine vollständige Freikühlung und eine Abwärmenutzung. Ab 2025 wird ein Großteil des Rechnersystems im Rahmen des NHR-Verbundes bereitgestellt und unterstützt u.a. den Forschungsschwerpunkt Atomistische Simulation. Weitere Informationen: pc2.uni-paderborn.de/about-pc2/announcements/news-events/article/contract-next-supercomputer-2025. (Kontakt: Jens Simon, PC²)

CLAIX-2023: Neues HPC-System an der RWTH Aachen

Das IT Center der RWTH Aachen hat sein neues HPC-Cluster Aix-la-Chapelle (CLAIX) erfolgreich in Betrieb genommen. Unter dem Namen CLAIX-2023 bietet es leistungsstarke CPUs von Intel (Xeon 8468 Sapphire Rapids) mit insgesamt 96 Kernen sowie DDR5-Speicher mit 256 bis 1024 Gigabytes pro Rechenknoten. Mit den zusätzlich integrierten 52 speziellen Servern für Machine-Learning- (ML) und KI- Anwendungen mit jeweils 4 GPUs des Typs H100 von NVIDIA ergibt sich allein im ML-Segment eine Gesamtleistung von mehr als 14 PetaFlop/s. Die 632 direktwassergekühlten Rechenknoten bewirken eine deutliche Leistungssteigerung und sind technologisch auf dem neuesten Stand in Bezug auf Nachhaltigkeit und Energieeffizienz. Darüber hinaus ist CLAIX-2023 mit einem schnellen NDR-Infiniband-Netzwerk ausgestattet, um große Modelle parallel simulieren zu können. Das zum Großteil über NHR4CES bereitgestellte Cluster unterstützt insbesondere die Ingenieurwissenschaften und die angewandten Naturwissenschaften. Weitere Informationen: itc.rwth-aachen.de/go/id/bgjoql/. (Kontakt: Christian Terboven, IT Center RWTH Aachen)

Wetter- und Klimamodell ICON ist Open Source

Die Wissenschaftsgemeinschaft in Deutschland und der Schweiz setzt einen Meilenstein für die Klima- und Wetterforschung: Seit Januar steht das verbreitete Modell ICON (ICOsahedral Non-hydrostatic modeling framework) unter einer Open-Source-Lizenz zur Verfügung. Dies markiert einen bedeutenden Fortschritt in der wissenschaftlichen Transparenz und fördert Innovationen. Ursprünglich für meteorologische Anwendungen konzipiert, wurde das ICON-Modell gemeinsam vom DKRZ, dem Deutschen Wetterdienst, dem KIT, dem Max-Planck-Institut für Meteorologie und dem Schweizerischen Zentrum für Klimasystemmodellierung zu einem umfassenden Erdsystemmodell weiterentwickelt. Durch die quelloffene Bereitstellung des Modellcodes ist das Modell frei nutzbar und trägt dazu bei, die Zusammenarbeit zwischen Forschung, nationalen Wetterdiensten und HPC-Herstellern zu intensivieren. Letztgenannte können etwa die Leistungsfähigkeit ihrer Hardware mithilfe von Wetter- und Klimamodellen testen und verbessern. Weitere Informationen: dkrz.de/de/icon-opensource. (Kontakt: Hendryk Bockelmann, DKRZ)

Neues Klimamodell für präzisere Prognosen

Wie wirkt sich die globale Erwärmung im kommenden Jahrzehnt in Deutschland aus? Sechs Institutionen, darunter das DKRZ, entwickeln in dem Projekt „Coming Decade“ ein neues Klimaprognosesystem, um genauere Aussagen für Deutschland und Europa treffen zu können. Die Folgen des Klimawandels zeigen sich regional und jahreszeitlich sehr unterschiedlich: Neben der allgemeinen Erwärmung können komplexe Prozesse und Wechselwirkungen auch für kühlere Klimaphasen sorgen. Ein kalter Winter oder Sommer bildet dazu keinen Widerspruch, sondern ist eine normale Auswirkung der Klimavariabilität. Präzise und zuverlässige Prognosen für die kommenden Jahre sind entscheidend, um sich zukünftig besser auf die Klimaveränderungen und -variabilität anzupassen und vorausschauende Maßnahmen einleiten zu können. „Coming Decade“ schließt an das von 2011 bis 2019 in zwei Phasen durch das BMBF geförderte Vorhaben „MiKlip – Mittelfristige Klimaprognosen“ zur Entwicklung eines dekadischen Klimavorhersagesystems an, in dem das DKRZ ebenfalls beteiligt war. In „Coming Decade“ ist das DKRZ verantwortlich für den Aufbau einer gemeinsamen Analyseplattform und der Datenaufbereitung des neuen Klimamodells ICON-Seamless. Das für 3 Jahre vom BMBF geförderte Projekt

startete im Dezember 2023. Weitere Informationen: dkrz.de/de/kommunikation/aktuelles/projektstart-coming-decade. (Kontakt: Christopher Kadow, DKRZ)

Schnellere Löser für Differentialgleichungen

In den Natur- und Ingenieurwissenschaften werden Problemstellungen mathematisch häufig durch partielle Differentialgleichungen beschrieben. Da für die technisch wichtigen Probleme keine expliziten Lösungsformeln bereitstehen, sind Approximationen mit numerischen Verfahren nötig. Das Projekt „PDExa – Optimierte Softwaremethoden und -technologien für partielle Differentialgleichungen" adressiert die Entwicklung neuartiger Algorithmen für zukünftige Exascale-Hardware. Der Schwerpunkt des Projekts liegt auf der Implementierung von Finite-Elemente-Verfahren, bei denen auf jedem Element eines Rechengitters polynomielle Ansatzfunktionen zur Approximation der Problemgeometrie und des Lösungsfeldes benutzt werden. Herkömmliche Implementierungen stoßen hier oft auf Grenzen von Speicherbandbreite und Kommunikationskosten. Das Projekt PDExa arbeitet an innovativen Lösungen unter Verwendung von sogenannten matrixfreien Implementierungen und verbesserten iterativen Lösungs- und Vorkonditionierungsverfahren, bei denen teure Speicherzugriffe reduziert und durch lokale Berechnungen ersetzt werden. Das Projekt wird durch das BMBF im Förderprogramm SCALEXA für 3 Jahre finanziert. Weitere Informationen: gauss-allianz.de/en/project/title/PDExa. (Kontakt: Martin Kronbichler, RUB)

Exascale-Berechnungen mit der ELPA-Bibliothek

Die seit 2008 quelloffen entwickelte ELPA-Bibliothek mit skalierbaren Eigenwertlösern für symmetrische Matrizen wird von allen namhaften Codes aus der Elektronenstrukturtheorie verwendet (vgl. gauss-allianz.de/en/info_letter/7). Ein Schwerpunkt des EU-Projektes „Novel Materials Discovery" (NOMAD CoE) war die Optimierung von ELPA für Exascale-Systeme und die neuesten GPUs der Hersteller AMD, Intel und NVIDIA. Mit den letzten Releases wurde für die nahezu auf allen großen HPC-Installationen installierte Bibliothek kontinuierlich eine immer umfangreichere Unterstützung von GPUs implementiert. Der vollständige Support von AMD-GPUs wurde durch Extreme-Scale-Berechnungen auf mehr als 1000 GPU-Knoten des finnischen HPC-Systems LUMI demonstriert. Mit dem ELPA-Release vom November 2023 steht allen Nutzenden eine Version mit performanter Unterstützung von AMD, Intel und NVIDA GPUs zur Verfügung. Weitere Informationen: elpa.mpcdf.mpg.de. (Kontakt: elpa-library@mpcdf.mpg.de, MPCDF)

KI-Sprachmodell als Service für die Wissenschaft

Die GWDG bietet seit Februar einen eigenen ChatGPT-Dienst an, der auf einer hauseigenen Hardware läuft. Damit können alle Nutzenden mit einem AcademicCloud-Account mit einer generativen KI chatten und sich bei ihrer Arbeit unterstützen lassen. Der Service basiert zurzeit auf Intel's NeuralChat-7b v3.1 und wird über die HPC-Infrastruktur der GWDG ausgeführt. Durch ein selbstentwickeltes System passt sich der Webdienst dynamisch an verschiedene Anfragevolumen an, um möglichst viele Anfragen gleichzeitig schnell zu bearbeiten. Besondere Sorgfalt lag bei der Konzipierung auf dem Datenschutz: Eingaben und Antworten werden nur lokal bei den Nutzenden innerhalb einer Sitzung gespeichert. Beim Schließen der Webseite werden alle Eingaben und Antworten gelöscht. Einzig die Anzahl der Anfragen pro Nutzer:in wird für Bedarfsauswertungen erfasst. Der Service wird aktiv weiterentwickelt, sodass in Zukunft komplexere Modelle angeboten und Integrationen mit anderen Plattformen ermöglicht werden können. Weitere Informationen: info.gwdg.de/news/der-gwdg-llm-service-generative-ki-fuer-die-wissenschaft/. (Kontakt: Julian Kunkel, GWDG)

Ausbau der Ressourcen für KI und Deep Learning

Bayern unterstützt die erwarteten KI- und Deep-Learning-Nachfragen der Forschung durch erhebliche finanzielle Mittel für neue Investitionen: Das NHR@FAU in Erlangen und das LRZ in Garching erhalten im Doppelhaushalt 2024/25 insgesamt 55 Mio. Euro. Das Budget wird v.a. für maßgeschneiderte KI-Hardware und den Betrieb verschiedener KI-Beschleuniger eingesetzt, die in eine anwendungsorientierte Speicherlandschaft eingebettet sind. Zusätzlich werden beide Zentren gezielt ihre Teams erweitern, um den Forschenden die bestmögliche Unterstützung zu bieten. Mit der Investition sollen sowohl die Leistung als auch der Energieverbrauch bei KI-Anwendungen optimiert werden. Darüber hinaus wird die neue Hardware mit Warmwasser gekühlt, um den Stromverbrauch zu reduzieren. Mit der Erweiterung der Hardware-Ressourcen etablieren sich NHR@FAU und LRZ als KI-Rechenzentren und versorgen die Hochschulen und wissenschaftlichen Einrichtungen im Freistaat. Weitere Informationen: go-nhr.de/ai-resources. (Kontakt: Gerhard Wellein, NHR@FAU)

Veranstaltungen

- 19.03.2024: Performance analysis on GPUs with NVIDIA tools, NHR@FAU, online
- 26.03.2024: HPC Tuning and Analysis, PC², hybrid
- 03.–05.04.2024: EUROCC2: N-Ways to GPU Programming Bootcamp, HLRS@GCS, online
- 10.04.2024: Introduction to LRZ HPC Systems with Focus on CFD Workflows, LRZ@GCS, online

HPC-Kalender der Gauß-Allianz:

hpc-calendar.gauss-allianz.de