



Achim Streit studierte von 1994 bis 1999 Ingenieurinformatik mit Nebenfach Elektrotechnik an der TU Dortmund. 2003 promovierte er über Job Scheduling Strategien für verteilte und parallele Computing-Systeme an der Universität Paderborn; während dieser Zeit war er wissenschaftlicher Mitarbeiter am Paderborn Center for Parallel Computing (PC<sup>2</sup>). Mitte 2005 wurde er Leiter der Abteilung „Verteilte Systeme und Grid Computing“ im Jülich Supercomputing Centre (JSC) am Forschungszentrum Jülich. Mitte 2010 wechselte er an das Karlsruher Institut für Technologie (KIT) als Mitglied des Direktoriums des Steinbuch Centre for Computing (SCC) und wurde zum Informatikprofessor ernannt.

Er ist Sprecher des Topic 2 „Data Intensive Science and Federated Computing“ im Helmholtz-Program „Supercomputing & Big Data“ und dessen stellvertretender Programmsprecher. Als SCG-Direktor ist er verantwortlich für GridKa, dem deutschen Tier-1 Daten- und Analysezentrum im Worldwide LHC Computing Grid (WLCG). Er ist Koordinator der Helmholtz Data Federation (HDF), Co-Koordinator des Helmholtz Analytics Framework (HAF) sowie Initiator der Helmholtz Information & Data Science Academy (HIDA). Darüber hinaus ist er verantwortlich für die Beteiligung des KIT in der European Open Science Cloud (EOSC). Er ist Mitglied in den Graduiertenschulen KSETA und HIDSS4Health sowie den KIT-Zentren KCIST, KCETA, Klima und Umwelt sowie MathSEE.

ÜBERBLICK UND ALLGEMEINES

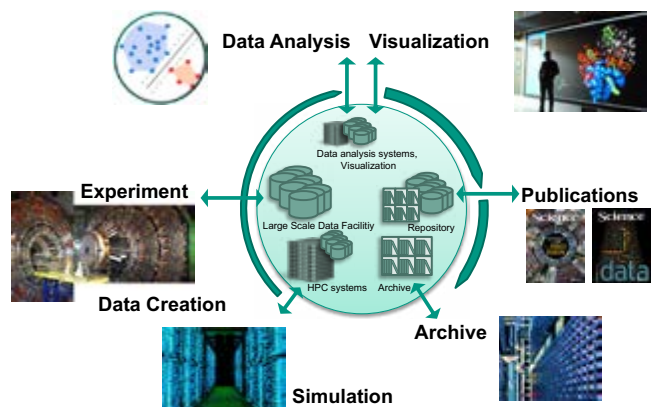
Die Professur Verteilte und Parallele Hochleistungssysteme ist eng in das SCC integriert und erlaubt so eine enge Verzahnung von Forschung, IT-Forschungsinfrastrukturen und praxisorientierter Lehre. Im Fokus stehen drei große Themenbereiche: **Data Management**, **Data Intensive Computing** und **Cloud Computing**.

Im **Data Management** werden Methoden und Technologien zum verteilten, effizienten und sicheren Umgang mit Forschungsdaten und zur Datenarchivierung (bit stream preservation) für den gesamten Datenlebenszyklus erforscht und entwickelt, die in Kooperation mit nationalen und internationalen Partnern in der Praxis und im engen Schulterschluss mit Anwendungswissenschaften eingesetzt werden. Besonderer Fokus liegt auf dem Umgang mit großen Datenmengen.

Im **Data Intensive Computing** werden u.a. effiziente Compute-Data-Schedulingverfahren sowie Datenanalyseverfahren erforscht. Ebenso werden parallele Simulations- und Daten-Analyse-Anwendungen auf heterogenen Parallelrechnern mit tausenden Prozessorkernen skaliert.

Im **Cloud Computing** werden ebenfalls modernste Managementmethoden zur effizienten Nutzung verteilter Ressourcen erforscht, z. B. mittels Scheduling und Auktionen.

In der Lehre werden die genannten Themen u.a. in den Vorlesungen „Parallelrechner und Parallelprogrammierung“ sowie „Verteiltes Rechnen“ und Seminaren und Praktika praxisnah vermittelt.



ERGEBNISSE UND ERFOLGE

Die Forschungsarbeiten werden durch die erfolgreiche Einwerbung und Durchführung zahlreicher Drittmittelprojekte flankiert.

Es begann der Aufbau der European Open Science Cloud (EOSC): im EOSCpilot Projekt arbeitet das SCC u.a. im Bereich Skills Development mit. Im EOSC-hub Projekt leitet das

SCC das Arbeitspaket zum föderierten Dienstmanagement. Im EOSCsecretariat.eu Projekt unterstützt das SCC den Aufbau der EOSC-Governance-Struktur.

Im DEEP-HybridDataCloud EU-Projekt werden Data Analytics Werkzeuge für den Einsatz in hybriden Cloud-Umgebungen angepasst und in die Cloud-Infrastrukturen integriert.

Das SCC koordiniert gemeinsam mit dem Jülich Supercomputing Centre (JSC) das Helmholtz Analytics Framework (HAF), in dem eine systematische Entwicklung von Datenanalyse-Techniken zusammen mit Domänenwissenschaftlern u.a. aus Klima, Bio- und Neurowissenschaften durchgeführt wird. Mit dem Helmholtz Analytics Toolkit (HeAT) wird eine leistungsfähige Open-Source Bibliothek für Scientific Big Data Analytics auf CPU/GPU-Systemen implementiert, die mit Tensor-Datenobjekten arbeitet, auf PyTorch aufbaut und MPI unterstützt.

Zahlreiche Vorträge und Veröffentlichungen der wissenschaftlichen Mitarbeiterinnen und Mitarbeiter sowie Doktorandinnen und Doktoranden ergänzten die Erfolge. Das Paper „Adaptive Lossy Compression of Complex Environmental Indices Using Seasonal Auto-Regressive Integrated Moving Average Models“ wurde auf der IEEE eScience Ende 2017 mit einem Best Paper Award ausgezeichnet.

## AUSGEWÄHLTE PUBLIKATIONEN

D. Gudu, M. Hardt, A. Streit: Combinatorial Auction Algorithm Selection for Cloud Resource Allocation Using Machine Learning. In: *Proceedings of the 24th International Conference on Parallel and Distributed Computing (Euro-Par)*, Springer LNCS 11014, S. 378-391, 2018. DOI: 10.1007/978-3-319-96983-1\_27.

U. Cayoglu, P. Braesicke, T. Kerzenmacher, J. Meyer, A. Streit: Adaptive Lossy Compression of Complex Environmental Indices Using Seasonal Auto-Regressive Integrated Moving Average Models. In: *Proceedings of 13th IEEE International Conference on eScience 2017, IEEE*. S. 315-324, 2017. DOI: 10.1109/eScience.2017.45.

E. Kühn, A. Streit: Online Distance Measurement for Tree Data Event Streams. In: *Proceedings of the 2nd IEEE International Conference on Big Data Intelligence and Computing (DataCom)*, IEEE. S. 681-688, 2016. DOI: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.122.

C. Jung, M. Gasthuber, A. Giesler, M. Hardt, J. Meyer, A. Prabhune, F. Rigoll, K. Schwarz, A. Streit: Progress in Multi-Disciplinary Data Life Cycle Management. In: *Journal of Physics: Conference Series*. volume 664 (3) 032018, 2015. DOI: 10.1088/1742-6596/664/3/032018.

## MITARBEITERINNEN UND MITARBEITER

### Verwaltungspersonal

Anja Müller

### Wissenschaftliches Personal

Dr. Markus Götz

Dr. Marcus Hardt

Dr. Eileen Kühn

Dr. Jörg Meyer

Elnaz Azmi

Uğur Çayoglu

Benjamin Ertl

Diana Gudu

Peter Krauß

Mehmet Soysal

Marco Strutz (extern)

Oskar Taubert

Michael Witt (extern)