



Kompetenz in der KI-Forschung – Das Helmholtz AI Local Energy Consultant Team
Competence in AI research – The Helmholtz AI Local Energy Consultant Team

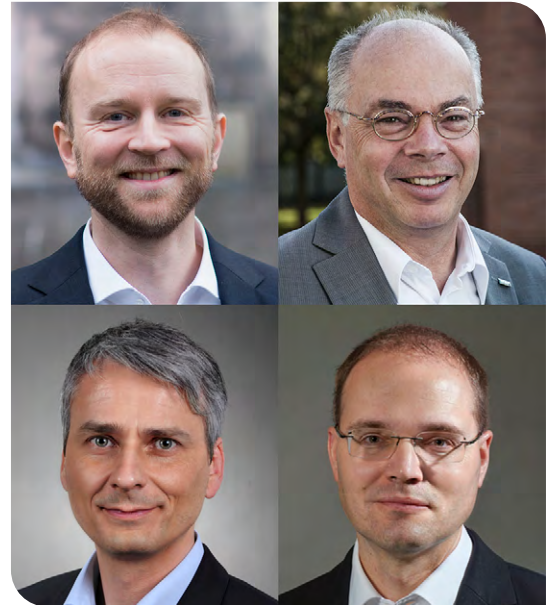
Auf zu neuen Herausforderungen: Let's go Quantum
Towards new challenges: Let's go Quantum

Anwendungs- und Systemüberwachung für SAP-Systeme
Application and System Monitoring for SAP Systems

Liebe Leserinnen und Leser,

effiziente und nachhaltige Lösungen für unsere Energiesysteme und -versorgung zu entwickeln, ist eine der großen Herausforderungen der Zukunft. Sie basieren zunehmend auf KI-Anwendungen und ermöglichen unter anderem neue Modelle der Energieverteilung, die den stark veränderten Voraussetzungen und Anforderungen gerecht werden. Praktische KI-Methoden müssen für diesen Anwendungsbereich entwickelt und angepasst werden. Hier kommt das seit Oktober am SCC aktive Helmholtz AI Consultant-Team ins Spiel (Titelbild), das Forschende in der Helmholtz-Gemeinschaft mit Fachwissen über KI-Methoden, Werkzeuge und Software-Engineering bei Energieforschungsprojekten unterstützt. Wir stellen Ihnen das Consultant-Team und dessen Mission ab Seite 20 vor.

„Auf zu neuen Herausforderungen“ heißt es, wenn es um ein anderes vielversprechendes Forschungsfeld geht: um das Quantencomputing. Es verheißt mit potentiell exponentiellen Rechenbeschleunigungen ein hohes Innovationspotenzial. Sollen Maschinen effizient Strukturen in Daten erkennen, können Quantencomputer anderen Rechnern weit überlegen sein. Ein Team des SCC beschäftigt sich daher mit dem Ziel, Algorithmen für Quantencomputer zu erforschen und in Machine Learning-Anwendungen zu übersetzen. Ab Seite 22 berichten wir darüber, wie junge Wissenschaftlerinnen und Wissenschaftler des SCC die Herausforderung „Let's go Quantum“ motiviert annehmen.



Als „Herz“ der Geschäftsprozesse, z.B. für das Personal- und Finanzmanagement, den Einkauf und die finanzielle Projektumsetzung sowie für diverse Self-Services, „schlägt“ das SAP-System im KIT. Erst durch das Zusammenspiel einer Vielzahl von Server-, Speicher-, und Applikationssystemen versorgt das SAP-System zahlreiche Dienstleistungseinheiten und Institute zuverlässig mit Diensten und Anwendungen. Besonders herausfordernd ist es, dieses Herz „gesund“ zu erhalten und frühzeitig zu erkennen, ob sich Unregelmäßigkeiten oder Probleme anbahnen. Das SAP-Betriebsteam hat hierfür ein Überwachungs- und Diagnosesystem der Meisterklasse aufgebaut (Seite 6).

Viel Freude beim Lesen

Martin Frank, Bernhard Neumair, Martin Nußbaumer, Achim Streit

Dear reader,

developing efficient and sustainable software for our energy supply systems is one of the grand challenges of the future. Increasingly based on artificial intelligence (AI) the new software among other things, enables energy distribution models that meet the changed prerequisites and requirements. For this, practical AI methods must be developed or adapted, which is where the Helmholtz AI consultant team, active since October, comes into play (cover picture). The team supports scientists doing energy research in the Helmholtz Association with specialists knowledge of AI methods, tools and software engineering. We introduce you the consultant team and their mission from page 20 onwards.

‘Off to new challenges’ is said of the another very promising field of research: quantum computing. With theoretically exponential computing acceleration it promises a high potential for innovation. If machines are to efficiently recognize structures in data, quantum computers can be far superior to today's computers. Therefore a team at SCC is researching algorithms for quantum computers and translating them into machine learning applications. Starting on page 22, we report on how young SCC scientists are motivated to take on the ‘Let's go Quantum’ challenge.

At the ‘heart’ of the business processes at KIT, e.g. for personnel and financial management, purchasing and financial project implementation as well as for various self-services, ‘beats’ the SAP system. Only through the interaction of a large number of servers, storage and application systems does it dependably supply services and applications to numerous organisational units and institutes. It is particularly challenging to keep this heart ‘healthy’ and to recognize early on whether irregularities or problems are looming. For this purpose the SAP operations team has set up a first class monitoring and diagnosis system (page 6).

Enjoy reading

Martin Frank, Bernhard Neumair, Martin Nußbaumer, Achim Streit

Inhaltsverzeichnis

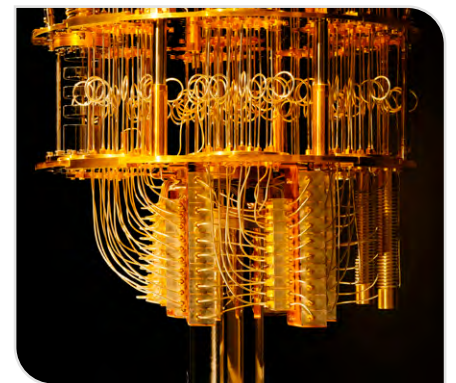
DIENSTE UND INNOVATION

- 04 RADAR4KIT – ein generisches Forschungsdatenrepositorium für das KIT
- 06 Anwendungs- und Systemüberwachung für SAP-Systeme am KIT
- 09 Das neue NETVS
- 12 Der GridKa-Rechencluster – Wissenschaft Non-Stop
- 14 Jupyter@SCC



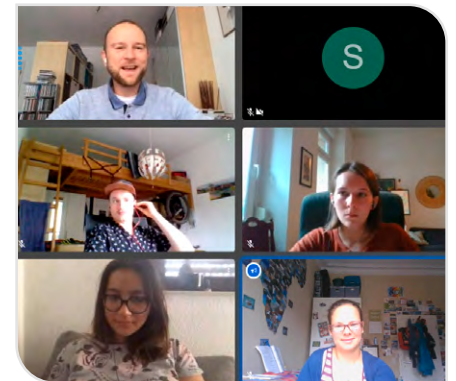
FORSCHUNG UND PROJEKTE

- 16 Lohnt sich das Upgrade auf die neue NVIDIA A100 GPU?
- 18 EU-Projekt DEEP-Hybrid-Datacloud bringt Werkzeuge für Benutzer und Entwickler von KI Anwendungen
- 20 Kompetenz in der KI-Forschung – Das Helmholtz AI Local Energy Consultant Team stellt sich vor
- 22 Auf zu neuen Herausforderungen: Let's go Quantum
- 25 NFDI – Nationale Forschungsdateninfrastruktur am SCC
- 28 Landesprojekt bwCard: Digitale Identitätskarte



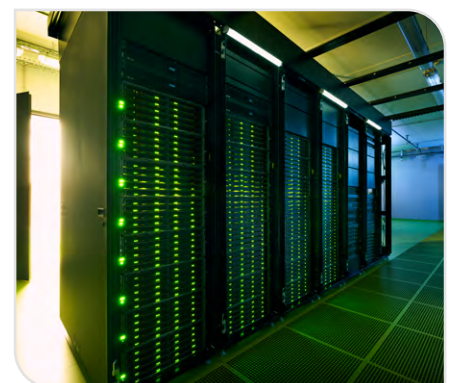
STUDIUM UND WISSENSVERMITTLUNG

- 32 MathSEE Modeling Week – Studierende präsentieren Ergebnisse



VERSCHIEDENES

- 34 SCC wird Zentrum für Nationales Hochleistungsrechnen
- 35 Impressum & Kontakt





RADAR4KIT – ein generisches Forschungsdatenrepositorium für das KIT

Das KIT stellt Wissenschaftlerinnen und Wissenschaftlern ein neues disziplinübergreifendes Repositorium für die Archivierung und Publikation ihrer Forschungsdaten zur Verfügung. Die Speicherung der Daten findet ausschließlich auf der IT-Infrastruktur des KIT am SCC statt.

Felix Bach

RADAR4KIT - Research Data Repository for KIT - ist ein disziplinübergreifendes Forschungsdatenrepositorium für die Archivierung und Publikation von Forschungsdaten - z.B. aus wissenschaftlichen Studien und Projekten des KIT. Unter Forschungsdaten werden digitale Daten verstanden, die im Forschungsprozess entstehen¹. Im Repositorium RADAR4KIT werden Forschungsdaten in Form von Datensätzen, die aus einer oder mehreren Dateien bestehen können, gespeichert. Diese Datenpakete enthalten die eigentlichen Forschungsdaten sowie beschreibende Metadaten. Durch die

Metadaten soll es Forschenden erleichtert werden, die Daten bei einer späteren Nachnutzung interpretieren zu können. Das Ziel ist es, Forschungsdaten „FAIR² zu machen“ – das heißt findable, accessible, interoperable und reusable.

RADAR4KIT wird vom Karlsruher Institut für Technologie als Dienstleistung für die Wissenschaft angeboten und setzt auf den von FIZ Karlsruhe betriebenen Dienst RADAR³ auf. Die Speicherung der Daten findet ausschließlich auf IT-Infrastruktur des KIT am SCC statt. Der Dienst richtet sich primär an Forschende des KIT

als „Datengeber“, welche ihre Daten archivieren oder publizieren wollen. Daten können für Dritte („Datennutzer“) zugänglich gemacht oder im Internet publiziert werden. RADAR4KIT steht als Online-Dienst über ein Web-Portal⁴ bereit, Forschende des KIT können sich mit ihrem KIT-Konto direkt anmelden - eine Registrierung ist zur Dienstonutzung nicht erforderlich. Lediglich spezielle Berechtigungen für Datengeber müssen in Absprache mit dem Serviceteam RDM@KIT⁵ zugeteilt werden. Es können auch Konten für KIT-Externe angelegt werden, die z.B. an Projekten des KIT beteiligt sind.

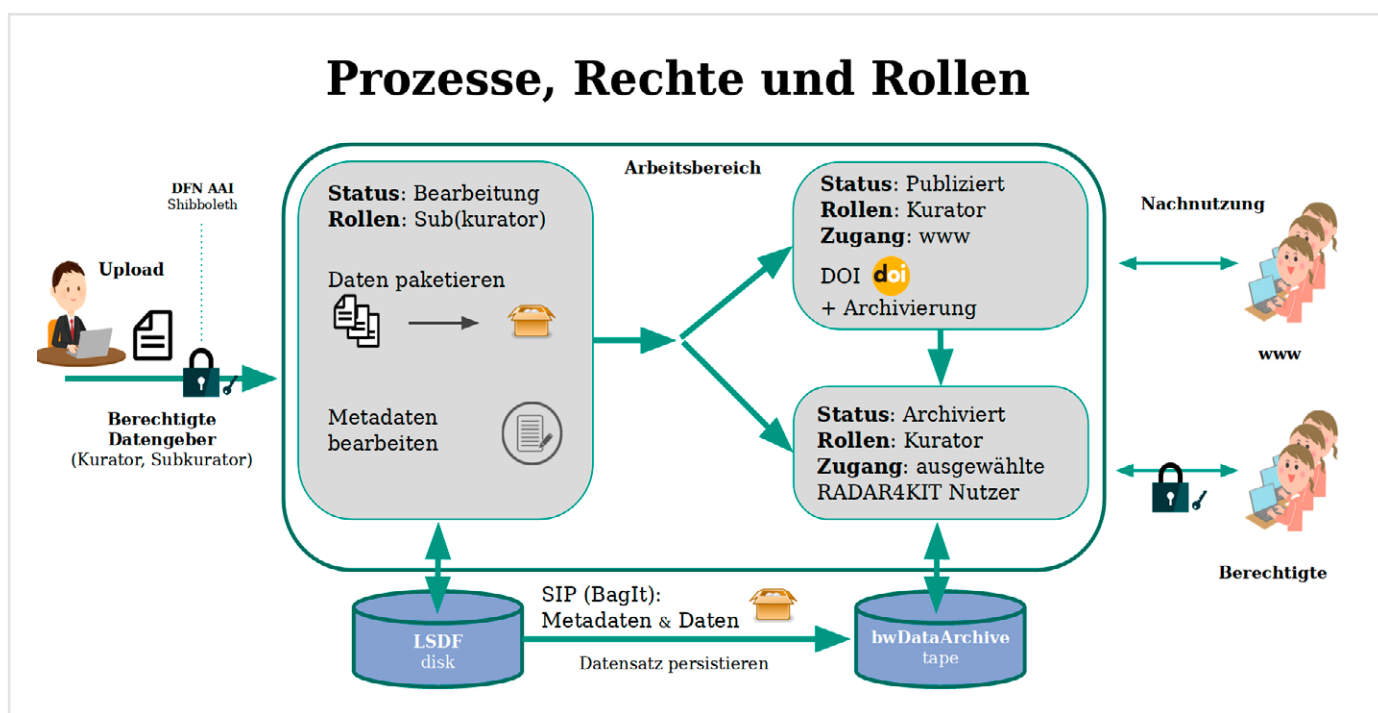


Abbildung 1: Ablauf der Erstellung von Datenpaketen und deren Archivierung und Publikation.

Für verschiedene Nutzergruppen (z.B. Forschungsgruppen, Projekte, Institute) können voneinander getrennte Arbeitsbereiche in Absprache mit dem Serviceteam RDM@KIT eingerichtet werden. Bei RADAR4KIT registrierte Nutzerinnen und Nutzer können darin unterschiedliche Rechte erhalten - etwa als Datengeber mit Publikationsrechten für einen oder mehrere Arbeitsbereiche (auch „Kurator“ genannt). Kuratoren können in die für sie bestimmten Arbeitsbereiche Forschungsdaten hochladen, diese bearbeiten, archivieren und gegebenenfalls publizieren. Sie können außerdem weitere registrierte Personen als sog. Subkuratoren für ihren Arbeitsbereich berechtigen. Subkuratoren sind ebenfalls Datengeber, haben jedoch nicht die Möglichkeit, Datenpakete eigenständig zu archivieren bzw. zu publizieren oder andere als Subkuratoren zu bestimmen. Durch diese Rollen und Rechte sind unterschiedliche Arbeitsprozesse umsetzbar (Abbildung 1), die gemeinsam mit den Wissenschaftlerinnen und Wissenschaftlern aus den diversen Disziplinen vom Serviceteam RDM@KIT nach und nach ausgestaltet werden.

Datenpakete, die noch nicht archiviert oder publiziert wurden, sich also im Bearbeitungszustand befinden, sind nur für berechnigte Datengeber sowie die Administratoren einsehbar. Ein Kurator kann im eigenen Arbeitsbereich jederzeit an weitere bei RADAR4KIT registrierte das Recht vergeben, als Datengeber (Kurator oder Subkurator) zu agieren.

Für die Publikation eines Datenpakets sind dessen valide Beschreibung in Form deskriptiver Metadaten sowie die Vergabe einer Lizenz durch den Datengeber erforderlich. Nach der Publikation sind die Metadaten und Daten öffentlich recher-

chier- und zugreifbar. Für die Daten kann der Datengeber optional einen Embargozeitraum bestimmen, innerhalb dessen nur die Metadaten öffentlich recherchierbar und verfügbar sind. Nach Ablauf der Embargofrist werden auch die zugehörigen Daten automatisch öffentlich zugreifbar. RADAR4KIT vergibt für jedes publizierte Datenpaket einen Persistenten Identifier (hier: Digital Object Identifier, kurz DOI) und registriert diesen bei DataCite. Über diesen DOI ist das publizierte Datenpaket nachhaltig identifizierbar, zitierfähig und kann mit einer herkömmlichen wissenschaftlichen Publikation z.B. im Repository KITopen der KIT-Bibliothek verknüpft werden. Publizierte Datensätze werden im Langzeitarchivdienst bwDataArchive⁶ für mindestens 10 Jahre archiviert.

Datensätze können auch lediglich archiviert werden. Diese können ebenfalls durch Metadaten beschrieben werden, erhalten jedoch keinen DOI und sind nur für einen ausgewählten, vom Datengeber berechtigten Personenkreis, zugreifbar.

RADAR4KIT bietet außerdem weitreichende Möglichkeiten zur Integration in die IT-Infrastruktur am KIT. Z.B. kann zusätzlich zur webbasierten Benutzungsoberfläche über eine REST-basierte Programmierschnittstelle (API) auf den Dienst zugegriffen werden. Weiterhin werden die deskriptiven Metadaten sowohl im RADAR-eigenen als auch im standardisierten DublinCore-Format über einen OAI-Provider öffentlich zum Harvesting angeboten, so dass Nachweissysteme und Suchportale diese Daten indizieren, aufbereiten und nachnutzen können.

RADAR4KIT unterstützt einen Peer Review-Prozess vor einer Datenpublikation. Dazu kann ein Datenpaket vor der

Publikation in den Status „in Begutachtung“ versetzt werden. In diesem Zustand ist das Datenpaket nicht mehr editierbar. Über einen eindeutigen Link, den der Datengeber an den zuständigen Verlag bzw. die Gutachter weitergeben kann, wird ein temporärer Zugriff auf das noch nicht veröffentlichte Datenpaket ermöglicht. Für die inhaltliche Qualitätssicherung der eingestellten Forschungsdaten sind die Datengeber ansonsten grundsätzlich selbst verantwortlich.

RADAR4KIT – a generic research data repository for the KIT

KIT provides scientists with a new interdisciplinary repository for archiving and publishing their research data. The data is stored exclusively on the KIT IT infrastructure at the SCC. Research data is digital data that is created during the research process. The RADAR4KIT repository stores research data in the form of data sets, which may consist of one or more files. These data packages contain the actual research data as well as descriptive metadata. The metadata is intended to make it easier for researchers to interpret the data in the event of subsequent use.

¹ www.rdm.kit.edu/fodaten.php

² www.force11.org/group/fairgroup/fairprinciples

³ www.radar-service.eu

⁴ radar.kit.edu

⁵ www.rdm.kit.edu

⁶ bwDataArchive Homepage: www.rda.kit.edu

Anwendungs- und Systemüberwachung für SAP-Systeme am KIT

Das SAP Enterprise Resource Planning (ERP)-System unterstützt mit diversen Anwendungen die Geschäftsprozesse am KIT. Es wurde das Ziel gesetzt, eine Überwachungs- und Störungsmanagement-Infrastruktur für aktuelle und zukünftige SAP-Systeme aufzubauen. Das Konzept zur Realisierung dieses Ziels besteht aus zwei Teilen: Erstens aus der Systemüberwachung der gesamten SAP-Systemumgebung und zweitens einer Geschäftsprozessüberwachung. Der erste Teil des Ziels konnte vom SAP-Systemteam im Juli umgesetzt werden und wird hier erläutert. Die Prozessüberwachung befindet sich in der Konzeption.

Dimitri Nilsen, Oleg Dulov, Gerald Helck

Systemüberwachung

In der Vergangenheit wurde die Systemüberwachung des SAP-Systems des KIT auf der Basis der bestehenden SAP-Anwendung innerhalb von SAP ohne zusätzliche Überwachungswerkzeuge implementiert. Somit bestand keine zentrale Anlaufstelle für die Problemerkennung. Dies entspricht nicht mehr den modernen Anforderungen im Betrieb einer solch umfangreichen SAP-Umgebung und bietet zudem keine Integration zwischen den einzelnen SAP-Instanzen.

Die aktuelle SAP-Systemlandschaft setzt sowohl Windows- als auch Linux-

Betriebssysteme ein. Sie besteht aus vielen verschiedenen Komponenten, wie Applikations-Server, Datenbanken, Nutzer-Frontends und diverse Server für Dokumentenablagen. Außerdem ist ein Umstieg auf die modernere SAP HANA-Technologie (Datenbank und die neuen SAP-Anwendungen) bereits im Gange.

Für die Metriken des Betriebssystems (z.B. CPU-Load, Disk Usage, Netzwerk-Traffic) wird ein Open-Source-Stack mit folgenden drei Komponenten eingesetzt: Telegraf, InfluxDB und Grafana (Abbildung 1). Telegraf ist ein Dienst, welcher auf den zu überwachenden Systemen läuft und die relevanten Metriken sammelt. Diese

Metriken werden anschließend an die InfluxDB Zeitreihen-Datenbank verschickt. Außerdem wird InfluxDB für das Sammeln von sowohl SAP-spezifischen Daten als auch für Daten von diversen Hilfssystemen der SAP-Systemlandschaft benutzt. Die Visualisierung und die Benachrichtigung (Alerting) erfolgt mittels Grafana-Dashboards.

Ein weiteres Open-Source-Tool, Ansible, wird für die automatisierte Konfiguration auf Betriebssystemebene verwendet. Die Ansible-Skripte werden im GitLab-Repositoryum des KIT gespeichert und mittels Continuous Deployment und Rolling Upgrades auf die Zielmaschinen ausge-

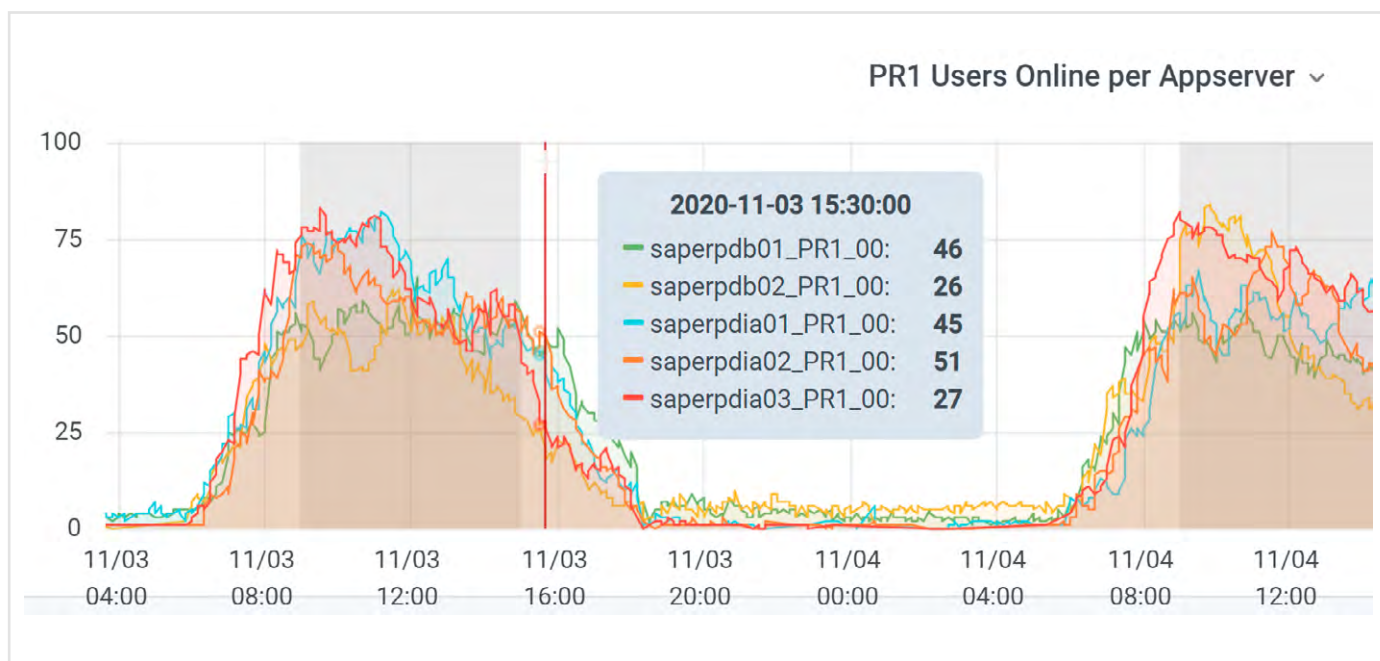


Abbildung 1: Nutzung des produktiven SAP-Systems dargestellt in Grafana

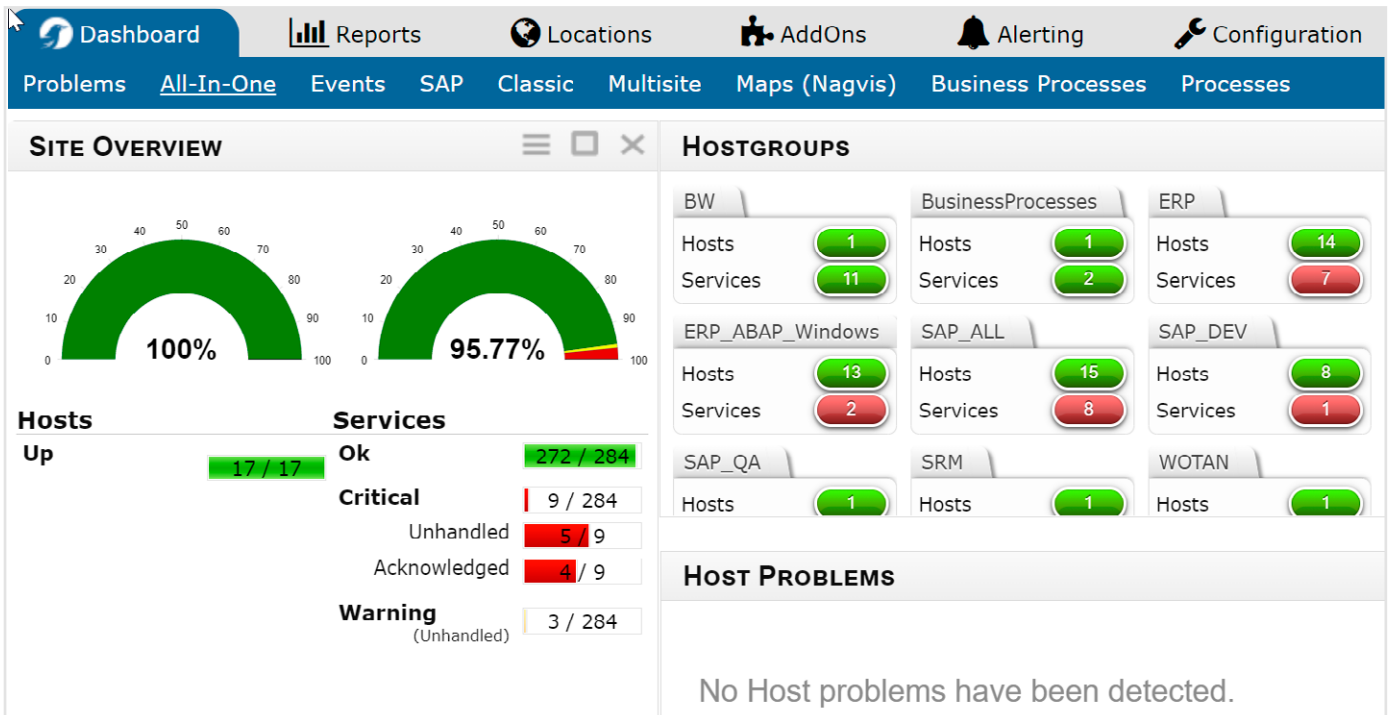


Abbildung 2: Wotan Monitoring Dashboard

rollt. Der für die Systemüberwachung in SAP verwendete Monitoring Stack orientiert sich an Verfahren, die bereits im Zusammenhang mit dem Grid Computing Centre Karlsruhe (GridKa) sowie den Landesdiensten bwCloud und bwSync&Share verwendet wurden. Allerdings wird er um SAP-spezifische Werkzeuge erweitert, um alle hier benötigten Daten sammeln zu können.

Die so implementierte Systemüberwachung bietet folgende Funktionen an:

- Statusübersicht technischer Systeme einschließlich einzelner Instanzen, Datenbanken und Hosts
- Visualisierung von Metriken und Ereignissen mit ihrer aktuellen Bewertung und den zuletzt gemeldeten Werten
- Drill-Down-Funktionen von Statusinformationen über technische Systeme bis hin zu einzelnen Metriken
- Visualisierung von Metriken und Ereignissen einschließlich Schwellenwerten
- Jump-in-Fähigkeit in den metrischen Monitor zur Anzeige historischer metrischer Werte einschließlich interaktiver Auswahl des anzuzeigenden Zeitraums

Wotan Monitoring

Für die Überwachung der weiteren SAP-spezifischen Metriken hat das SCC eine auf der Open-Source-Lösung Nagios/Icinga basierende Software Wotan Monitoring gekauft und eingesetzt. Die Wotan Monitoring-Software ermöglicht eine schnelle Problemerkennung und ein zuverlässiges Eskalationsmanagement bei Zwischenfällen und umfasst Überwachung, Alarmierung, Berichterstattung und Dokumentation. Die Konfiguration basiert auf einem Vorlagenkonzept.

Vordefinierte Konfigurationsvorlagen können als Ausgangspunkt für die Ableitung eigener kundenspezifischer Vorlagen für SAP- und Nicht-SAP-Metriken verwendet werden (Abbildung 2).

Die Wotan-Agenten sind sowohl auf dem Betriebssystem als auch im SAP-System als SAP-spezifische Anwendungen installiert. Außerdem bietet Wotan Monitoring ein Plug-in zur Visualisierung der Daten in Grafana. Wotan verfügt über zahlreiche Prüfungen zur Überwachung von SAP-Prozessen und -Aufträgen direkt

im SAP-System, einige davon sind unten aufgeführt:

- Laufzeitanalyse von abgebrochenen oder fehlerhaften SAP-Jobs
- Laufzeitanalyse der SAP BTC/DIA/UPD-Prozesse, Status von UPD (Background Updates)
- Datenbank-Metriken und Analysen der DB-Transaktionen
- SAP-System- und Kundeneinstellungen
- Überprüfung von externen Anbindungen, REST- und RFC-Schnittstellen
- Betriebssystemüberwachung

SAP Solution Manager

Der SAP Solution Manager (SolMan) ist ein Produkt von SAP und dient als zentrale Instanz für die Überwachung des Konfigurations- und Installationsstatus aller SAP-Systeme sowie für das Change Management. Der integrierte Maintenance Optimizer (MOPZ) beschleunigt und verbessert die Vorbereitung und Durchführung von Updates und Upgrades. Das Einspielen von Support Packages kann mit dem Maintenance Optimizer und den von SAP bereitgestellten Upgrade-

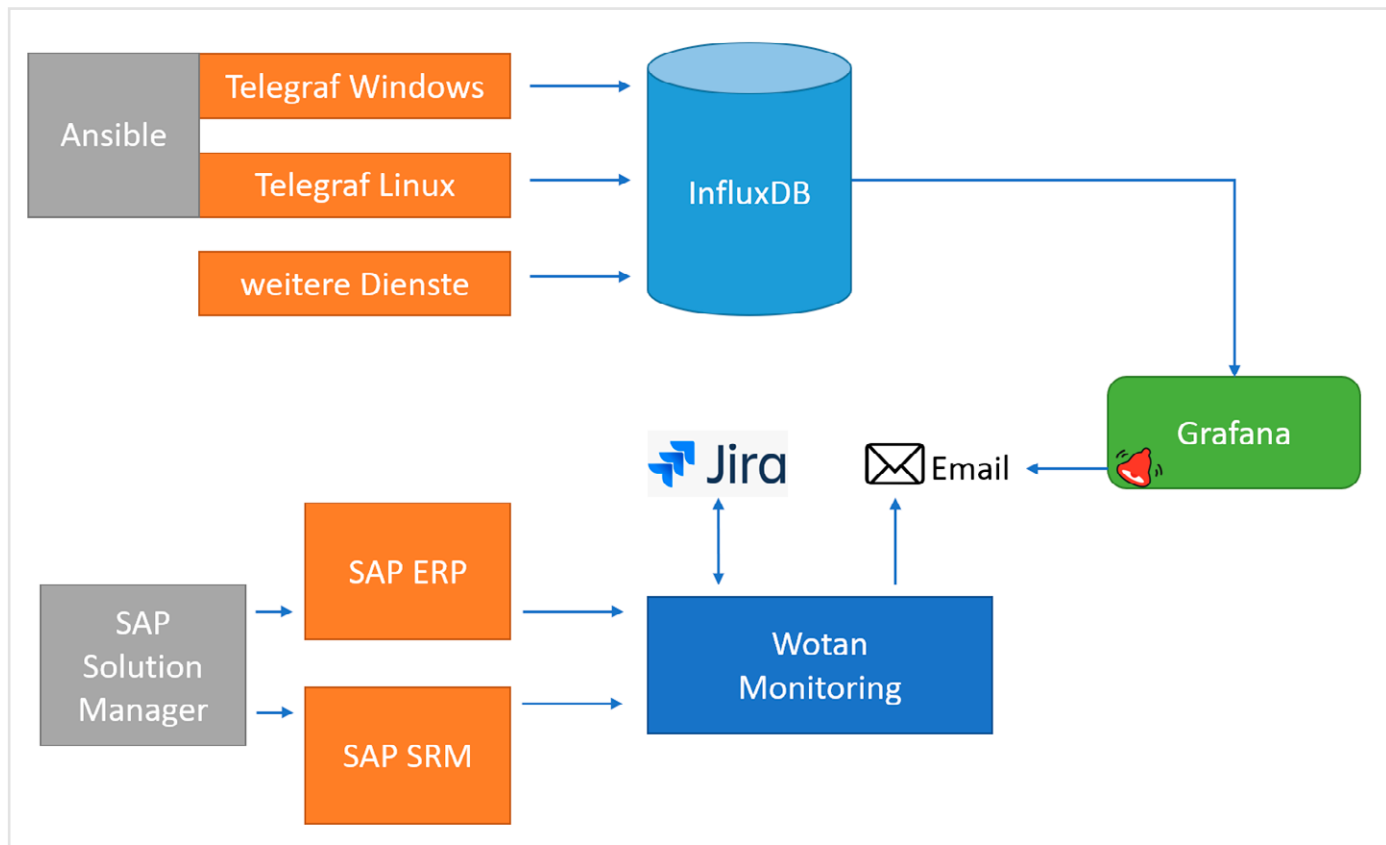


Abbildung 3: Die in der SAP-Systemlandschaft eingesetzte Benachrichtigungsinfrastruktur macht einen effizienten Incident Management-Prozess möglich.

Inhalten sicherer durchgeführt werden. Das SAP-Betriebsteam setzt den SolMan vor allem zur Überwachung von Schnittstellen und Geschäftsprozessen ein.

Alerting und IT-Incident Management

Um die schnelle und zuverlässige Reaktion auf die vom Monitoring erkannten Probleme zu gewährleisten, wird eine Reihe von Benachrichtigungsmedien eingesetzt (Abbildung 3). Standardmäßig wird eine Benachrichtigung per E-Mail versendet. Weitere Benachrichtigungskanäle sind MS Teams Chats und SMS (bei Wotan Monitoring).

Wotan bietet die Möglichkeit, verschiedene Arten von Wartungszeiträumen (Downtimes) zu definieren, z.B. periodische Downtimes, bei denen Zeiträume für sich wiederholende Wartungen eingeplant werden können. Durch die Verbindung der Wotan-Software mit dem

Servicemanagement-Tool Jira werden die erkannten Probleme aus Wotan in Jira-Tickets umgewandelt und stehen zur weiteren Bearbeitung zur Verfügung. Dies stellt einen transparenten Fehlerbehebungsprozess für die SAP-Administratoren und -Administratorinnen dar.

Ausblick

Durch den Einsatz von Top-Level-Monitoring-Lösungen wurde eine zentrale Überwachungs- und Benachrichtigungsinfrastruktur

sowohl für IT-System- als auch für SAP-spezifische Metriken geschaffen. Die nächsten Schritte sind unter anderem die kontinuierliche Verbesserung der Alarmierung und die Einrichtung der verschiedenen Überwachungsfunktionen des SAP Solution Managers für das IT-Service-Management sowie für die Geschäftsprozesse. Für die Zukunft ist eine stärkere Integration von Jira geplant.

Application and System Monitoring for SAP Systems at KIT

The SAP Enterprise Resource Planning (ERP) system supports business processes at KIT with various applications. The objective was to establish a monitoring and incident management infrastructure for current and future SAP systems. The concept to reach this goal consists of two parts: Firstly, system monitoring of the entire SAP system environment and secondly, business process monitoring. The first part of the objective was implemented by the SAP system team in July and is explained here. The process monitoring is currently in the conception phase.

Das neue NETVS

Das SCC stellt bereits seit vielen Jahren eine mandantenfähige Oberfläche für die Verwaltung von Netzdiensten wie beispielsweise DNS für die IT-Beauftragten und Verantwortlichen in der IT-Administration am KIT zur Verfügung. Nun geht das Netzdiensteverwaltungssystem (NETVS) in die nächste Runde – mit neuer Oberfläche, Architektur und Features. Dieser Artikel soll einen Einblick in die Architektur und Neuerungen des NETVS schaffen.

Janis Streib

Der Netzbetrieb steht regelmäßig vor neuen Anforderungen und Herausforderungen. Um diesen gerecht zu werden, wird das hauseigene NETVS¹ mitsamt zugehöriger Datenbankstruktur und Anwendungslogik stetig weiter entwickelt. Seit dem Wochenende des 7./8.11.2020 ist nun das neue Webinterface mit neuer Entwicklungsschnittstelle und Datenbank in Betrieb. Kernfunktionen sind derzeit die Verwaltung von DNS-Einträgen sowie von MAC-Adressen-basierter Authentifikation. Dabei können Änderungen transaktional – also durch Ausführung mehrerer Aktionen in einem Durchlauf, ohne die gegenseitige Beeinflussung paralleler Ausführungen – geplant und durchgeführt werden.

Anwendungsarchitektur

Das NETVS besteht im Kern aus drei Komponenten: Der Datenbank samt integrierter Anwendungslogik (NETDB), dem API-Server zur Bereitstellung der Programmierschnittstelle (API) sowie dem Webinterface. In der bisherigen Architektur bestand das Webinterface aus genau einer Serverkomponente, die mittels des Templatesystems "Jinja2" und einer direkten Verbindung zur NETDB das Webinterface mitsamt den gewünschten Daten serverseitig erzeugt. Allerdings zeigten sich schnell einige Nachteile dieses Entwurfs:

1. Performance

Da bei jedem Seitenaufruf die Seite serverseitig neu und individuell erzeugt werden musste, hatte man mit einer bestimmten Mindestladezeit zu rechnen. Außerdem skalierte die Ausgabe großer Datenmengen nicht. Das hatte bei

bestimmten großen Ausgaben zu einer sehr hohen Serverlast und gegebenenfalls sogar zum Abbruch der Anfrage geführt.

2. Wartbarkeit

Die Serveranwendung hinter dem Webinterface kommunizierte direkt mit der Datenbank. So musste das gesamte Datenmodell nochmals erneut (parallel zum API-Server) vorgehalten werden. Das führte dazu, dass alle Änderungen am Datenmodell an mehreren Stellen gepflegt werden mussten.

3. Konsistenz

Die Datenein- und -ausgabe verhielten sich zwischen API und Webinterface nicht konsistent. So waren manche Ausgaben in der API nicht möglich, die das Webinterface jedoch darstellen konnte und umgekehrt. Genauso waren Feldbeschreibungen und die Semantik von Parametern nicht immer gleich.

Um diese Probleme anzugehen, wurde ein vollständiges Redesign der Architektur vorgenommen. Das Webinterface besteht nun aus zwei Komponenten: Der Middleware, die serverseitig ausgeführt wird, sowie einem JavaScript-Frontend, was beim Nutzer im Browser läuft. Die Middleware dient jetzt nur noch zur Bereitstellung von API-Diensten, die nicht durch die NETDB bereitgestellt werden können (beispielsweise MacFinder), und zur Erstellung von Session-Tokens. Alle anderen NETDB-bezogenen Anfragen werden vom Client nun direkt an den API-Server gerichtet. Die Darstellung erfolgt dann komplett im Client.

Die Anmeldung im NETVS geschieht auch nicht mehr über eine eigene Anmelde- maske, sondern über OpenID Connect². Die Anmeldung erfolgt somit zentral über den Shibboleth Identity Provider des KIT, der auch eine Zwei-Faktor-Authentifizierung erlaubt (Abbildung 1).

API 3.0

Die Entwicklerschnittstelle wurde ebenfalls grundsätzlich überarbeitet und löst API 2.x ab, die nun nur noch eingeschränkt zur Verfügung steht. Damit wurde erreicht, dass die Struktur über alle Ebenen hinweg konsistent ist und sich die Semantik und Dokumentation aus der API ableiten lässt – die API dokumentiert sich somit selbst. Durch diese Eigenschaften ist es erstmals möglich, einfach zu wartende, erweiterte Ressourcen für Entwickler bereitzustellen. So gibt es nun einen API-Browser³ sowie eine Python-Bibliothek⁴, die automatisch aus der API erzeugt werden.

Ein weiteres neues Werkzeug in der API ist die Join-Funktion im Transaktionskontext. Diese erlaubt die Formulierung komplexerer API-Abfragen, vergleichbar zu Joins in relationalen Datenbanken. Beispielsweise kann man nun in genau einer Abfrage zu allen Domain-Namen, die einem bestimmten Muster entsprechen, die zugehörigen DNS-Einträge transaktional abfragen, ohne dafür separate Anfragen tätigen zu müssen.

¹ netvs.scc.kit.edu

² www.scc.kit.edu/dienste/openid-connect.php

³ git.scc.kit.edu/scc-net/netvs/netdb-client-lib

⁴ netvs.scc.kit.edu/swagger

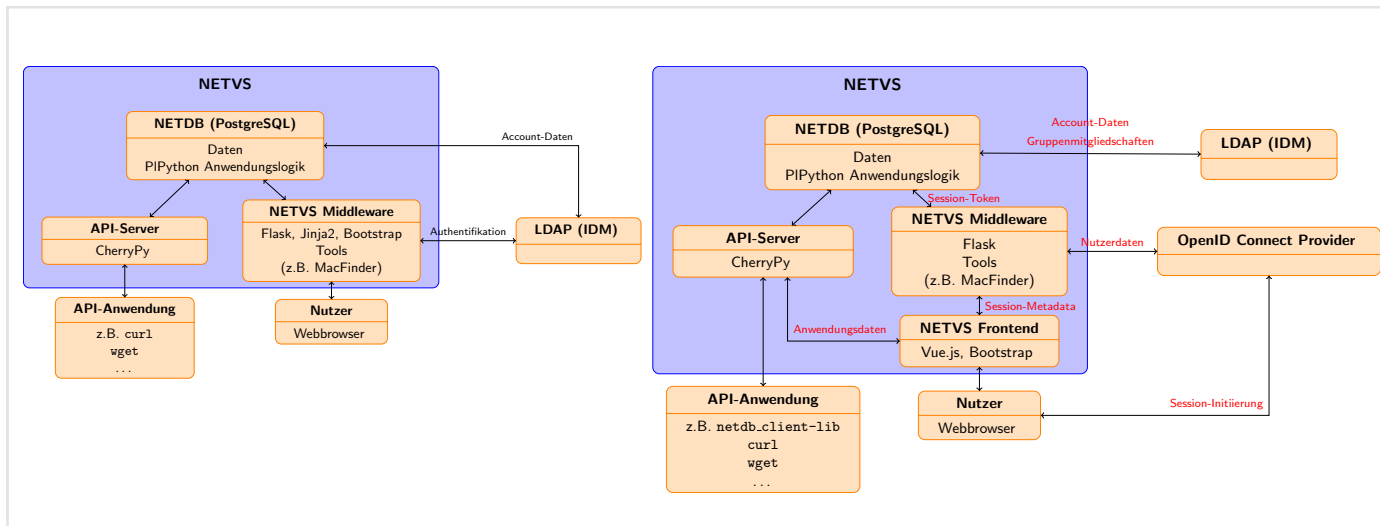


Abbildung 1: Übersicht über die bisherige (links) und die neue Architektur (rechts) des NETVS

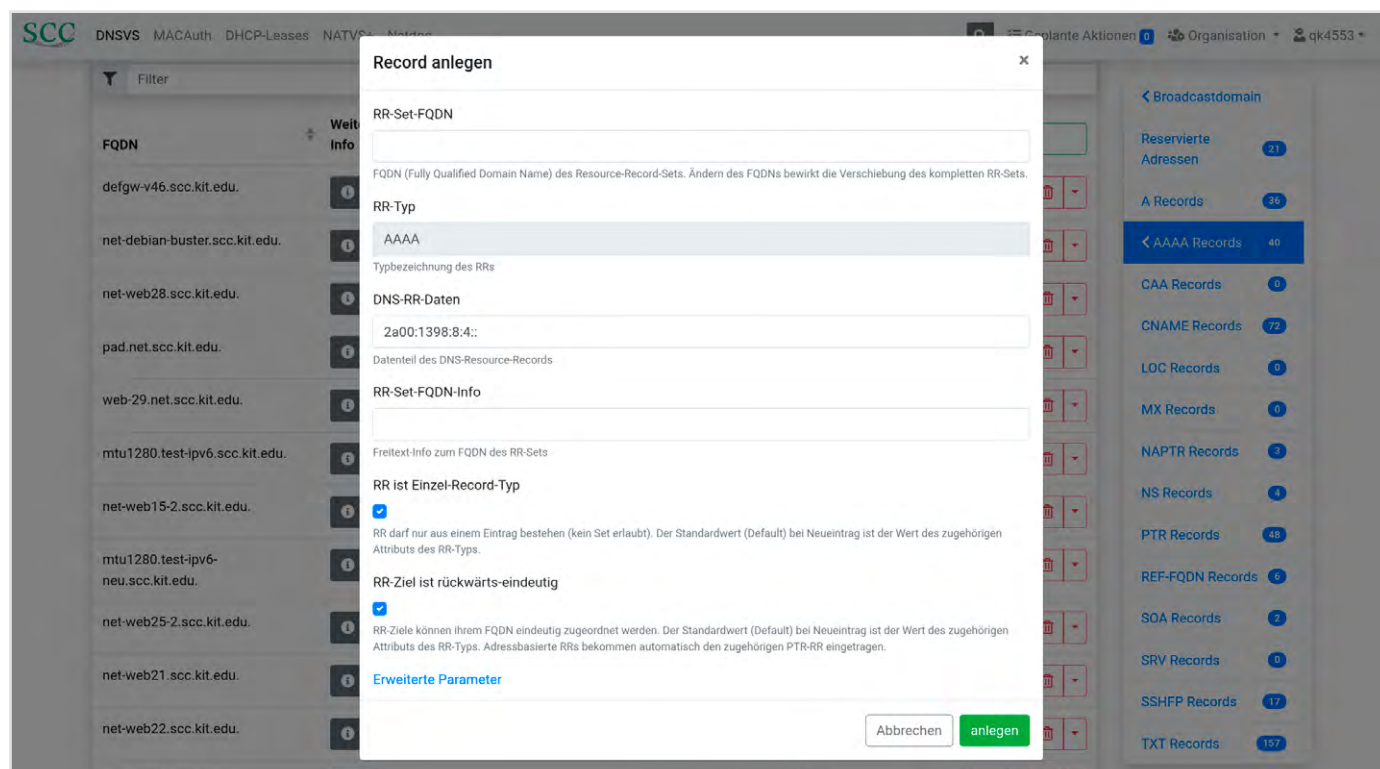


Abbildung 2: Die Eingabefunktionen werden nun einheitlich und größtenteils konsistent mit der API dargestellt.

Dieses "traditionelle" Vorgehen hat einen größeren lokalen Verarbeitungsaufwand zur Folge und kann – je nach Anwendung – keine Konsistenz zwischen den Anfragen garantieren.

Die Selbstdokumentation sowie die Join-Funktion bilden die Grundpfeiler des JavaScript-Clients. Ein Beispiel hierfür ist die Generalisierung der Eingabefunktionen (Anlegen und Bearbeiten) (Abbildung 2).

Hier werden die nötigen Parameter, Feldnamen und Feldbeschreibungen vollständig zur Laufzeit aus der API generiert. Bei Anpassungen der API sind keine Anpassungen im Client mehr nötig.

Mit der API 3.0 ändert sich auch die Authentifikation: Anstatt wie bisher mit Client-Zertifikaten erfolgt die Authentifikation nun über Tokens, die über das NETVS erstellt werden können.

Rechteverwaltung

Große Änderungen ergeben sich auch bei der Rechteverwaltung. IT-Beauftragte und Verantwortliche in der IT-Administration haben jetzt feinere Kontrolle über die Berechtigungen auf die ihnen zugeteilten Ressourcen.

Hierzu wurden mehrere neue Konzepte eingeführt:

1. Unterkonten

Unterkonten dienen zur Einschränkung der API-Tokens. So kann man jetzt ein Unterkonto seines eigenen Kontos anlegen, welches beispielsweise keine datenmodifizierenden Funktionen erlaubt. Reguläre Tokens zur Verwendung in eigenen Anwendungen können ausschließlich für Unterkonten angelegt werden.

2. Gruppen und Untergruppen

Gruppen bestimmen im neuen NETVS nun, welche Konten Rechte auf Broadcast-Domains (also den Adressraum, kurz BCDs) und Domains (Namensraum) haben. Mitglieder einer Gruppe können dann Operationen auf den durch die Gruppe vorgegebenen Namens- und Adressraum ausführen (z.B. DNS-Records anlegen). Gruppen können optional mit den gleichnamigen Gruppen aus dem zentralen Identity Management (IDM) synchronisiert werden (Abbildung 3).

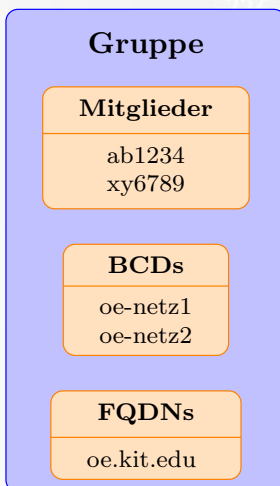


Abbildung 3: Betreuer-, Broadcastdomain- und Domainzuordnungen werden nun über Gruppen abgebildet.

Untergruppen bilden hier das Gegenstück zu den Unterkonten und erlauben die Einschränkung des Namens- und Adressraums der Unterkonten. Untergruppen können von den Gruppenmitgliedern selbst angelegt und administriert werden.

3. Organisationseinheiten und Unter-Organisationseinheiten

Organisationseinheiten (OEs) sind nun expliziter in die Rechteverwaltung eingebunden. IT-Beauftragte (ITBs) nehmen hierbei die Rolle der OE-Betreuer ein und können alle Gruppen und Broadcast-Domains, die ihrer OE zugeteilt sind, administrieren. Auch können durch OE-Betreuer weitere OE-Betreuer hinzugefügt werden. Des Weiteren verfügen OEs über eine Domainliste, die als möglicher Namensraum für die Gruppen zugeteilt werden können.

OEs können beliebig viele Unter-OEs besitzen. Die Rechte der OE-Betreuer vererben sich entsprechend durch die Hierarchie.

Die bisherigen Adressbereiche im DNS-Verwaltungssystem (DNSVS), die zuvor Einstiegspunkt und Rechteanker waren, existieren nun nicht mehr. Stattdessen ergeben sich die Adressbereiche aus den BCDs, denen der jeweils vorherige DNSVS-Adressbereich zugeordnet war. Hierdurch sind auch Dualstack-Bereiche (also Bereiche mit einem IPv4- und IPv6-Subnetz) in einer Ansicht einsehbar und verwaltbar. Zuvor waren hierfür zwei separate DNSVS-Bereiche nötig.

Fazit

Das neue NETVS bringt viele technische und strukturelle Neuerungen mit sich. Neben weitreichenden Umstrukturierungen – begonnen bei den Änderungen der Architektur bis hin zu der neuen Gruppenstruktur – sind auch neue, mächtige Werkzeuge hinzugekommen.

Die Umstellung lief im Großen und Ganzen erfolgreich, und kleinere Probleme konnten schnell behoben werden. Sollten dennoch Probleme auftreten, hilft das NETVS-Team gerne. E-Mail: netvs@sc.kit.edu

The new NETVS

For many years, the SCC has provided a multi-client capable interface for the administration of network services such as DNS for the IT appointees and administrators at KIT. Now the network services management system (NETVS) is entering the next round – with a new interface, architecture and features. This article is intended to provide an insight into the architecture and innovations of NETVS.

Der GridKa-Rechencluster – Wissenschaft Non-Stop

Seit fast zwei Jahrzehnten bietet das Grid Computing Centre Karlsruhe (GridKa) Ressourcen und Services für daten- und rechenintensive Experimente der Teilchen- und Astroteilchenphysik. Neben Tape- und Platten-Speicher betreibt das GridKa einen der größten Rechencluster des weltweiten LHC Computing Grids. Für eine hohe Verfügbarkeit ist das GridKa praktisch durchgehend in Betrieb und durchläuft daher bereits seit Jahren eine andauernde Evolution.

Max Fischer

Als eines von 13 Tier-1-Zentren des Weltweiten LHC Computing Grids (WLCG) ist das Grid Computing Centre Karlsruhe (GridKa) speziell auf die Bedürfnisse der großen Kollaborationen aus Teilchen- und Astroteilchenphysik ausgelegt. Eines der Alleinstellungsmerkmale eines Tier-1-Zentrums im WLCG ist die Verfügbarkeit eines Tape-Speichers: Dieser ermöglicht die sichere und langfristige Speicherung von Rohdaten aus den Detektoren der einzigartigen Experimente. Zusätzlich bietet der Platten-Speicher sowohl Volumen als auch Bandbreite, um Daten kurz- und mittelfristig zu speichern und diese auf der ganzen Welt verfügbar zu machen. Für die effiziente Vorverarbeitung und Analyse dieser Daten dient der GridKa-Rechencluster.

Bunter Mix an Arbeit

Der Rechencluster am GridKa ist ausschließlich ausgelegt auf High-Throughput Computing (HTC), also einen möglichst hohen Gesamtdurchsatz an Berechnungen und Daten. Jeder einzelne Rechenauftrag (Job) ist „klein“, mit einer Nutzung von 1 oder 8 CPU-Kernen, 2 bis 4 GB RAM pro Kern und typischen Laufzeiten bis zu einem Tag, maximal einer Woche. Den Ausgleich macht die Masse: Ungefähr 30.000 Jobs laufen gleichzeitig auf den fast 50.000 CPU-Kernen des Clusters. So werden weit mehr als zwei Millionen Jobs im Monat abgearbeitet.

Dabei könnten die Rechenaufgaben nicht unterschiedlicher sein. Von Simulationen, die hauptsächlich physikalische Modelle berechnen, über Datenrekonstruktionen, welche Detektor-Rohdaten

für die Auswertung aufbereiten, bis hin zu Physikanalysen, die unentwegt Daten durchsuchen, ist alles dabei. Natürlich hat dabei jede Forschungskollaboration unterschiedliche Schwerpunkte, die sich auch über die Zeit ändern. Wer diese Woche noch CPUs mit Simulationen an ihre Grenzen bringt, liest vielleicht nächste Woche schon Daten aus aller Welt über das Netzwerk.

Apropos Kollaborationen: Einzelne Nutzerinnen oder Nutzer sucht man im GridKa-Rechencluster fast vergeblich. Stattdessen schickt praktisch jede Nutzergruppe sogenannte Pilot Jobs, welche als Platzhalter agieren und sich nur für Ressourcen in der Warteschlange des Rechenclusters anstellen. Erst wenn eine Maschine des Clusters den Pilot Job ausführt, ruft dieser die aktuellsten Rechenaufträge der Nutzergruppe ab und führt sie aus. Nur auf diese Art können die fast 200 Rechencluster im WLCG durch tausende Nutzer von dutzenden Institutionen verwendet werden, ohne im Verwaltungschaos zu enden – sowohl für Administrierende als auch Nutzende.

Keine Zeit für Pausen

Eine solche Menge und Vielfalt an Jobs abzuarbeiten, bringt seine ganz eigenen Herausforderungen mit sich. Während HPC-Rechencluster wie HoreKa oder ForHLR im wahrsten Sinne des Wortes einen Scheduler – also einen Terminplaner für die Ausführung von Jobs – betreiben, ist dies in einem HTC-Rechencluster wie dem von GridKa heutzutage nicht realistisch: Die Vielzahl kleiner Jobs ohne vorab bekannte Laufzeit bedeutet, dass

Blick in den Rechnerraum des Grid Computing Centre Karlsruhe (GridKa) am Campus Nord des KIT. (© KIT)

Mit CPUs gegen COVID-19

Das Zusammenspiel des GridKa-Clusters und der Pilot Jobs der Nutzerkollaborationen ist technisch angelehnt am Volunteer-Computing, dem Verfügbarmachen von Rechenressourcen als Beitrag zu einem größeren Rechenvorhaben aus kleinen Einzelaufgaben. Dies macht es möglich, schnell und zuverlässig Ressourcen für wichtige Rechenaufgaben zur Verfügung zu stellen: Seit April spendet das GridKa dauerhaft zwischen 5.000 und 10.000 CPU-Kerne an die Volunteer-Computing-Projekte Folding@Home und Rosetta@home sowie die Kollaboration WeNMR, um damit die Forschung über die Viruskrankheit COVID-19 zu unterstützen und zu beschleunigen.^{3,4} (doi.org/10.1051/epjconf/201921403053)

andauernd irgendwo im Cluster Ressourcen freierwerden. Um eine möglichst hohe Auslastung zu erreichen, müssen diese Ressourcen sowohl schnell, als auch für alle Nutzergruppen fair neu vergeben werden.

Da die Hardware des GridKa-Clusters regelmäßig erweitert wird, muss auch seine Steuerungssoftware – das Batch System – fortlaufend angepasst und verbessert werden. Um dadurch nicht die angestrebte Auslastung von 95% bis 98% zunichte zu machen, muss dies im laufenden Betrieb geschehen. Bereits Anfang 2018, als der GridKa-Cluster nur etwa 20.000 CPU-Kerne umfasste, war die damalige Batch System Software jedoch bis an die Grenze ihrer Möglichkeiten ausgereizt. Ein kompletter Austausch gegen eine weitaus besser skalierende Software war unumgänglich.

Zu Gute kam in dieser Situation die hohe Flexibilität der neu gewählten Batch System Software HTCondor². Ausgelegt auf hochgradig verteilte, unabhängige Komponenten, skaliert diese fast automatisch mit der Größe eines Rechenclusters und nicht mit der Rechenleistung einer einzigen zentralen Maschine. Über einen Zeitraum von mehreren Wochen konnten somit schrittweise einzelne Rechenknoten aus der Kontrolle des alten Batch-Systems in das neue überführt werden.

Die Tore zur Welt

Obwohl der GridKa-HTCondor-Rechencluster immer noch seine ursprünglichen Kernkomponenten nutzt, ist er bis heute bereits auf mehr als die doppelte Größe angewachsen. Über 1.000 Rechner werden von dem System verwaltet und ihre Ressourcen stetig neu aufgeteilt, um den aktuellen Anforderungen der unterstützten Kollaborationen zu entsprechen. Doch tatsächlich ist etwas scheinbar Simples wie "Ressourcen aufzuteilen" bei dieser Größe alles andere als trivial. Während die kleinsten Rechner des GridKa-Clusters nur 16 CPU-Kerne bieten, wartet die neueste Generation mit 256 CPU-Kernen auf. Diese gleichmäßig, schnell und fair an die anstehenden Jobs aufzuteilen, ist auch weiterhin eine Herausforderung.

Doch nicht nur der GridKa-Cluster selbst entwickelt sich weiter, auch das Grid geht mit der Zeit. Die Verbindung zwischen Rechencluster und Grid sind die Compute Elements (CE): im Prinzip Submittierungsknoten für den Rechencluster, zugänglich über die Authentifizierungsverfahren des Grid. Nachdem das GridKa jahrelang die Grid Computing Middleware Nordugrid ARC CE betrieben hat, ist dieses Jahr auch hier ein Generationenwechsel vollzogen worden: seit August wird bei GridKa nun das HTCondor CE verwendet und bietet somit eine einheitliche Technologie sowohl lokal als auch für die Anbindung an die Welt.

The GridKa Compute Cluster – non-stop scientific computing!

For almost two decades, the Grid Computing Centre Karlsruhe (GridKa) is providing resources and services for data- and computation-intensive experiments in particle and astroparticle physics. In addition to tape and online storage, GridKa operates one of the largest computing clusters of the worldwide LHC Computing Grid. To ensure high availability, GridKa is in operation literally continuously and has therefore been undergoing a continuous evolution for years.

¹ SCC News 2/2018

² research.cs.wisc.edu/htcondor

³ www.erum-data-idt.de/post/cobald_tardis_covid-19/

⁴ www.scc.kit.edu/ueberuns/13531.php

Jupyter@SCC

Einsteiger und Studierende sehen sich bei der Nutzung der HPC-Systeme des SCC häufig mit einer steilen Lernkurve konfrontiert. Seit Oktober 2020 steht daher nun zusätzlich auch Jupyter als alternativer, graphischer und web-basierter Zugangsweg zu den HPC-Systemen zur Verfügung.

Jennifer Buchmüller, Samuel Braun und Simon Raffener

Der Zugriff auf HPC-Systeme erfolgt traditionell über textbasierte Kommandozeilenschnittstellen, die mit Hilfe von verschlüsselten Netzwerkprotokollen, z.B. Secure Shell (SSH), über Netzwerke wie das Internet genutzt werden. Wird diese Verbindung durch Protokolle wie X11 oder VNC erweitert, ist auch eine graphische Ausgabe möglich. Das klingt nicht nur kompliziert, insbesondere für Erstnutzerinnen und -nutzer ist es das auch.

Was ist Jupyter?

Das vor einigen Jahren gegründete Jupyter-Projekt¹ stellt eine einfache, web-basierte Entwicklungsumgebung zur Verfügung. Die darin verwendeten Programmiersprachen Julia, Python und R (daher der Name „Jupyter“) sind in einigen Disziplinen wie dem der Künstlichen Intelligenz und dem Machine Learning (KI/ML) stark verbreitet und besonders geeignet für den Einstieg ins [technisch-wissenschaftliche] Programmieren. Die Nutzenden können in einem Web-Frontend Dateien öffnen, diese in einem sogenannten „JupyterLab“ interaktiv bearbeiten und dort auch direkt ausführen. Die Ausgaben des Programmcodes sind direkt im Dokument sichtbar.

Seit Ende Oktober bietet das SCC Jupyter als Service an. Neben dem klassischen Zugang via SSH, ist nun auch der interaktive Zugriff auf alle HPC-Systeme des SCC via Webbrowser möglich.

Programmieren leicht gemacht

Jupyter ermöglicht damit eine neue Art des interaktiven Supercomputings. Der klassische Prozess der Programmierung einer Anwendung, die auf dem Supercomputer läuft, umfasst mindestens vier

Schritte, hier am Beispiel einer Python-Anwendung dargestellt:

- Aufbau einer SSH-Verbindung
- Editieren der Python-Dateien in einem Texteditor
- Einstellen eines Jobs in die Warteschlange
- Analyse der Ergebnisse

Ein JupyterLab kombiniert diese Schritte in einer einzigen, interaktiven Web-Ansicht. Die Verbindung zum HPC-System wird automatisch hergestellt.

Die Bearbeitung des Codes wird durch sogenannte „Code-Konsolen“ direkt im Browser realisiert. Diese bieten Syntax-Highlighting, integrierte Code-Prüfung und andere Komfortfunktionen ähnlich einer traditionellen Integrierten Entwicklungsumgebung (IDE) (Abbildung 1). Die meisten Jupyter-Code-Konsolen unterstützen Dateiformate wie Bilder, CSV- und JSON-Dateien, Markdown, PDFs, Vega- und Vega-Lite-Visualisierungen.

Die eigentlichen Laufzeitumgebungen, die für die Ausführung des Codes benutzt werden, liegen als sogenannte „Kernel“ vor. Über das Kernel-Backend können die Nutzerinnen und Nutzer die in der Code-Konsole bearbeiteten Codes sofort interaktiv ausführen. Einer der bekanntesten Kernel im Jupyter-Projekt ist der ipython-Kernel. Dieser ermöglicht die interaktive Ausführung von Python-Code. Das Jupyter-Projekt bietet eine ständig wachsende Liste von Kernen für Laufzeitumgebungen wie z.B. Julia, R, Octave, Haskell, C/C++ an. Sollte Jupyter eine Laufzeitumgebung noch nicht unterstützen, können der Code im Web-Frontend editiert und dann in der interaktiven Kommando-Konsole die zur Ausführung

des Codes nötigen Schritte (z.B. Aufruf eines Compilers) manuell durchgeführt werden.

Jupyter auf den HPC-Systemen

Das Einrichten eines JupyterLab auf einem Laptop oder PC nimmt weniger als 20 Minuten in Anspruch. Die Einführung von JupyterLab in der komplexen Infrastruktur eines HPC-Systems mit Hunderten von Rechenknoten, einem Queuing-System, einer Benutzerverwaltung mit einer Vielzahl von Projekten, komplexen Zugriffsregeln und strikten Sicherheitsregeln erfordert hingegen deutlich mehr Aufwand. Die verschiedenen Bestandteile eines JupyterLabs müssen voneinander entkoppelt, auf verschiedene Hardwarekomponenten aufgeteilt und mit einigen zusätzlichen Komponenten versehen werden.

Dazu wird auf einem speziellen Server, der sich in der Nähe des HPC-Systems befindet, die Software „JupyterHub“ installiert. Auf den HPC-Systemen werden die zu unterstützenden Jupyter-Kernel installiert, da die Berechnungen dort und nicht auf dem Web-Frontend laufen werden.

Das JupyterHub-Web-Frontend übernimmt zunächst im Zusammenspiel mit dem föderierten Identitätsmanagementsystem bwIDM die Zwei-Faktor-Authentifizierung mittels OpenID Connect. Nach erfolgter Authentifizierung wählen die Anwenderinnen und Anwender die gewünschten Ressourcen aus einer Liste vorgefertigter Optionen z.B. eine GPU für eine Dauer von einer Stunde. Aus dieser Auswahl wird automatisch eine Anforderung für das Queuing-System generiert. Dieses prüft die Anforderung und allokiert die Ressourcen auf den Compute-

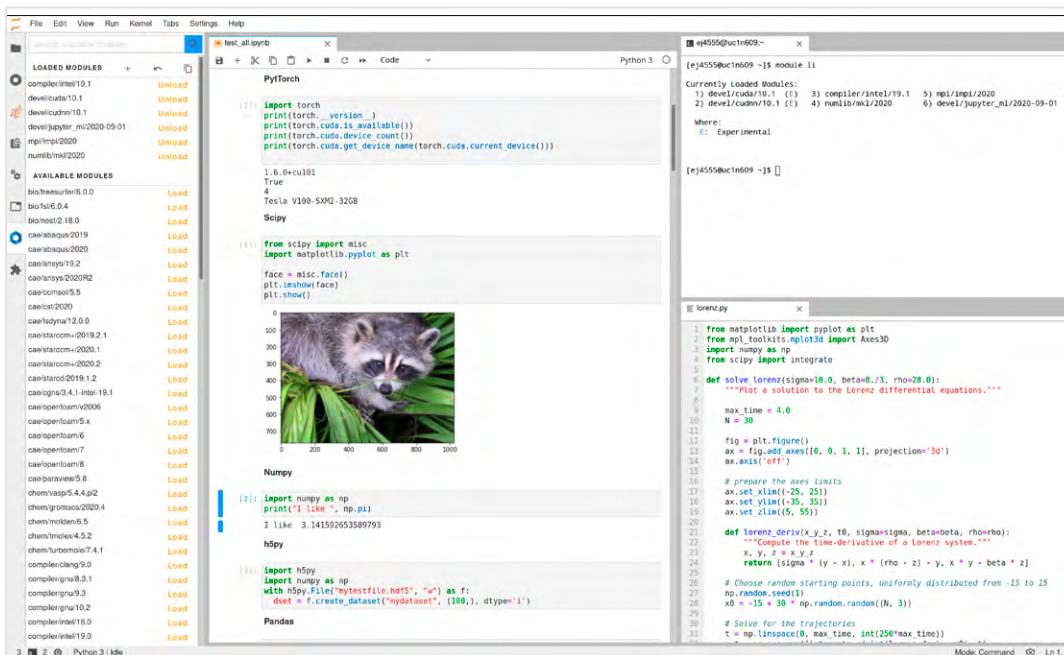


Abbildung 1: Beispiel einer Programmieraufgabe in der JupyterLab-Entwicklungsumgebung.

Knoten des HPC-Systems. Innerhalb dieser Ressourcen wird ein sogenannter „JupyterSpawner“ gestartet, der die Verbindung zwischen Web-Frontend und HPC-System realisiert. Gleichzeitig wird in der Web-Anwendung eine JupyterLab-Instanz gestartet und im Hintergrund eine Verbindung von dieser JupyterLab-Instanz zum zugehörigen JupyterRunner auf dem HPC-System geschaltet. Damit ist die Einrichtung komplett und alle im Web-Browser ausgelösten Aktionen und Programme laufen nun auf den HPC-Ressourcen. Die Ausgaben der Berechnungen werden wieder zurück an den Web-Browser übertragen und können dort sofort visualisiert werden. Die Ausführung des Codes findet komplett auf dem HPC-System statt, die Anwenderinnen und Anwender können also auch auf Software oder Hardware zugreifen, die lokal nicht installiert oder nicht vorhanden ist.

Herausforderung Jupyter-Software-Ökosystem

Auf den ersten Blick sieht dieses Konzept bereits wie eine sehr gute Lösung aus, allerdings wird es nun erst richtig kompliziert. Höchste Leistung und effizientes Arbeiten werden erst dann ermöglicht,

wenn verschiedenen wissenschaftlichen Gemeinschaften genau die Softwareumgebung zur Verfügung steht, die am besten zu ihnen passt. Die Komplexität dieser Aufgabe lässt sich bereits an einem einfachen Beispiel, nämlich der Berechnung einer Sinusreihe, zeigen. Sogar in Python, der derzeit meistverbreiteten Skriptsprache, gibt es dafür mindestens drei verbreitete Lösungen. In Python muss zuerst ein Modul wählen, das die Sinusfunktion implementiert und die Auswertung der Programmzeile `print(sin(x))` für `x in list(xi)` ermöglicht. Sowohl das integrierte `math`-Modul als auch die bei Wissenschaftlern weit verbreiteten Module `NumPy` und `SciPy` bieten entsprechende Funktionen. `NumPy` und `SciPy` wiederum können auf verschiedene numerische Hilfsbibliotheken (z.B. Intel MKL oder OpenBLAS) zurückgreifen. Am Ende hängt die Performance des Python-Codes auf einem bestimmten System von der richtigen Kombination aller genannten Komponenten ab. Dies ist ein sehr einfaches Beispiel, bei Anwendungen der Künstlichen Intelligenz (KI) und des Maschinellen Lernens (ML) wird es noch viel komplizierter.

Effektive Nutzerunterstützung

Das SCC vertritt schon lange die Philosophie, dass auf den HPC-Systemen sogenannte „Modulsysteme“ angeboten werden, die einen einfachen Zugriff auf gängige Softwarekomponenten wie wissenschaftliche Bibliotheken oder Compiler ermöglichen. Unter diesen Modulsystemen befinden sich auch solche, die verschiedene Python-Module, Bibliotheken und andere Softwarekomponenten für bestimmte Einsatzzwecke bündeln. Die KI/ML-Community findet beispielsweise im Modul

„jupyter_ml“ mehr als 30 gängige Python-Module und kann diese mit Jupyter sofort für Berechnungen auf den neuen GPUs nutzen. Die eingangs erwähnte steile Lernkurve, die bei einem textbasierten SSH-Zugang nötig wäre, entfällt. Jupyter ist sowohl auf dem deutschlandweit verfügbaren Tier-2-System ForHLR II als auch auf dem für das KIT und das Land betriebenen `bwUniCluster 2.0` verfügbar. Weitere Informationen zu Jupyter@SCC finden sich unter www.scc.kit.edu/dienste/JupyterHPC.

Jupyter@SCC

Since October 2020 the SCC now also offers Jupyter as an alternative access path to the HPC systems. Jupyter is a new way of “interactive supercomputing”, enabling interactive work with runtime environments like Python and data visualizations. Users only need a web browser on their local system to use Jupyter and have access to all resources on the HPC systems, including GPUs. Preconfigured software environments, e.g. for GPU-accelerated Machine Learning using Python, are available. Jupyter is available both on ForHLR II and `bwUniCluster 2.0`. More information can be found at www.scc.kit.edu/dienste/JupyterHPC.

Lohnt sich das Upgrade auf die neue NVIDIA A100 GPU?

Immer mehr Supercomputer der Spitzenklasse nutzen die Rechenpower von Hardware-Beschleunigern. Zum ersten Mal wird auch ein System mit den neuen NVIDIA A100 GPUs unter den Top500 gelistet. Die Hardware-Architektur der Beschleuniger ist dabei speziell für KI-Algorithmen entwickelt worden. Doch lohnt sich ein Umstieg auf diese neuen Systeme? Die Berechnungen für diese Studie wurden auf den im Juli am SCC in Betrieb genommenen Hochleistungsservern vom Typ DGX A100 durchgeführt.

Hartwig Anzt, Tobias Ribizel

In den letzten zehn Jahren haben Grafikkarten (GPUs) immer mehr Einzug in HPC-Plattformen (High Performance Computing) gehalten, und in der Top500-Liste vom November 2020 wurden sechs der zehn schnellsten Systeme mit GPU-Beschleunigern ausgestattet. Die gängigste Form von Beschleunigern sind GPUs. Die Juni 2020-Ausgabe der Top500 listet zum ersten Mal ein System mit NVIDIAs neuen A100-GPUs auf – eine GPU mit der HPC-zentrierten Ampere-Architektur, die für KI-Anwendungen entwickelt wurde. Drei DGX A100-Systeme mit jeweils acht NVIDIA A100 GPUs sind seit Juli am SCC im Einsatz¹. Jetzt, wo dieser neue Flaggschiff-Chip von NVIDIA auf dem Markt ist, fragen sich Anwendungswissenschaftler, die sich auf GPU-beschleunigte wissenschaftliche Simulationscodes verlassen, ob es nicht an der Zeit ist, ihre Hardware aufzurüsten.

Um diese Frage zu beantworten, werfen wir einen Blick auf die Performanz, die iterative Löser und Batch-Routinen mit der NVIDIA A100-GPU erzielen, und quantifizieren die Beschleunigung gegenüber ihrem Vorgänger, der NVIDIA V100-GPU. Die Motivation, sich auf diese Routinen zu konzentrieren, liegt darin, dass viele wissenschaftliche Anwendungen entweder (1) auf Batch- und Sparse-Linear-Algebra-Bibliotheksroutinen basieren oder (2) aus Operationen mit sehr ähnlichen Eigenschaften bestehen. Folglich können Leistungsgewinne für diese Benchmarks ein guter Indizwert für die Beschleunigung sein, die man für wissenschaftliche Anwendungen erwarten kann, ohne zusätzliche Code-Modifikationen vornehmen zu müssen.

In Abbildung 1 sind die Beschleunigungen dargestellt, die sich ergeben, wenn man eine NVIDIA V100-GPU durch eine NVIDIA A100-GPU ohne Code-Modifikationen ersetzt. Während die Hauptspeicherbandbreite auf dem Papier von 900 GB/s (V100) auf 1.555 GB/s (A100) gestiegen ist, liegen die Beschleunigungsfaktoren für die STREAM-Benchmark-Routinen² bei großen Datensätzen zwischen 1,6 und 1,72. Gleichzeitig konnte beobachtet werden, dass beim Zugriff auf kleine Datensätze die Speicherbandbreite der A100-Architektur tatsächlich geringer ist als die Bandbreite der V100.

Für das Sparse Matrix-Vektor-Produkt (SpMV) – ein Schlüsselalgorithmus für

Sparse-Linear-Algebra- und wissenschaftliche Rechenanwendungen – hängen die Leistungsverbesserungen vom individuellen Datenformat, der Kernel-Implementierung und den spezifischen Problemeigenschaften ab. Die in Abbildung 1 dargestellten Beschleunigungswerte für die SpMV-Kernel aus NVIDIAs cuSPARSE-Bibliothek und der Open-Source-Bibliothek Ginkgo (s. SCC-News 01/2019) sind alle über die mehr als 2.800 in der Suite Sparse Matrix Collection verfügbaren Testmatrizen gemittelt. Da viele dieser Matrizen klein sind, sind die Kernel nicht in der Lage, die volle Speicherbandbreite zu erreichen. Folglich sind die Beschleunigungswerte für die SpMV-Kernel im Allgemeinen niedriger

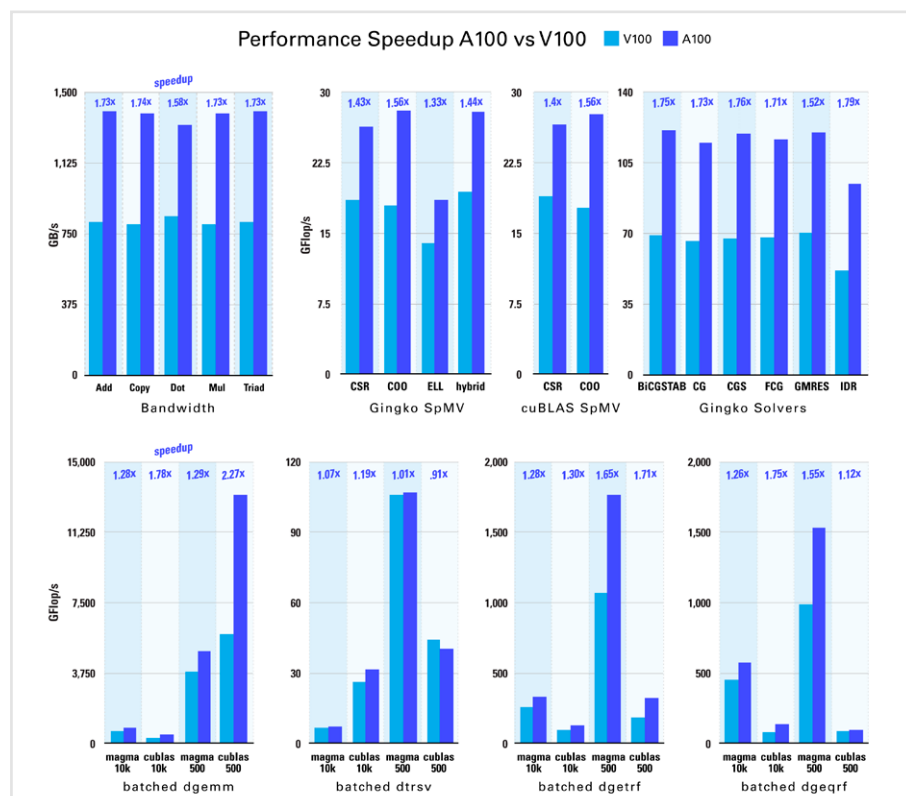


Abbildung 1: Die A100- versus V100-Architektur - ein detaillierter Beschleunigungsvergleich.

¹ www.scc.kit.edu/ueberuns/13772

² 2018. Evaluating attainable memory bandwidth of parallel programming models via BabelStream. Int. J. Comput. Sci. Eng. 17, 3 (January 2018), 247–262. dl.acm.org/doi/10.5555/3292750.3292751

als die für die STREAM-Benchmarks. Bei der Leistungsanalyse für die iterativen linearen Löser von Ginkgo konzentrieren wir uns auf große Testprobleme, um sicherzustellen, dass die Bandbreite in den Vektoroperationen voll ausgeschöpft wird. Je nach individuellem Algorithmus laufen die iterativen Löser von Ginkgo auf der A100-GPU im Vergleich zur V100-GPU zwischen 1,5 und 1,8-mal schneller.

Schließlich gilt es auch die Beschleunigung von Batch-Routinen zu untersuchen, die ebenfalls in wissenschaftlichen Rechenanwendungen üblich sind. Es ist wichtig zu erwähnen, dass MAGMAS Batch-Routinen zwar besonders für die V100-Architektur entwickelt wurden, durch Feintuning auf die A100-Architektur aber möglicherweise noch höhere Beschleunigungen möglich sind. Nichtsdestotrotz ergeben sich aus den Tests attraktive Leistungssteigerungen bis

zum 1,6-fachen, die "umsonst" kommen. Erwähnenswert ist zudem, dass der A100-GPU-Tensorecore-Beschleuniger Berechnungen in IEEE754 double precision durchführen kann. Dies ist eine neue Hardware-Fähigkeit, die bei den A100-Vorgängern nicht vorhanden war. Solche drastischen Architekturverbesserungen stellen eine Herausforderung für Open-Source-Bibliotheken wie MAGMA dar, die darauf abzielen, hochgradig abgestimmte numerische Software für eine breite Palette von Hardware-Architekturen bereitzustellen. Beispielsweise nutzen die vorhandenen, rechenintensiven Codes in MAGMA derzeit nicht die Vorteile der A100-Tensorecores für IEEE754 double precision. Dies bedeutet, dass diese Codes bestenfalls durch eine theoretische Spitzenleistung von 9,7 Teraflops (das ist etwa 1,3-mal besser als der V100) beschränkt sind. Wenn MAGMA jedoch die Vorteile der neuen Tensorcore-Beschleuni-

ger nutzen kann, beträgt die theoretische Spitzenleistung 19,5 Teraflops (das wäre um den Faktor 2,6 besser).

Angesichts dieser insgesamt konsistenten Ergebnisse können wir erwarten, dass auch komplexe wissenschaftliche Rechenanwendungen 1,3 bis 1,7-mal schneller ausgeführt werden können, ohne architekturenspezifische Änderungen vorzunehmen. Zwar kann die Frage, ob dies die Investition in neue Hardware rechtfertigt, hier nicht eindeutig beantwortet werden, aber es wird klar, dass die neue Architektur von NVIDIA mit Schwerpunkt auf KI-Algorithmen eine erhebliche Leistungssteigerung gegenüber dem Vorgängermodell bietet. Das SCC freut sich, einige der weltweit ersten A100 GPUs zu betreiben, die bereits von Anwendungswissenschaftlern eingesetzt werden.

Is the NVIDIA A100 GPU Performance Worth a Hardware Upgrade?

Over the last decade, GPU accelerators have seen an increasing rate of adoption in high-performance computing (HPC) platforms, and in the June 2020 Top500 list, eight of the ten fastest systems featured GPUs. The June 2020 edition of the Top500 is the first edition listing a system equipped with NVIDIA's new A100 GPU – the HPC-centric Ampere GPU designed with AI applications in mind. With this new flagship NVIDIA chip now on the market, domain scientists relying on GPU-accelerated scientific simulation codes wonder whether it is time to upgrade their hardware.

To help answer this question, we take a look at the performance we achieve on the NVIDIA A100 for sparse and batched computations and quantify the acceleration over its predecessor, the NVIDIA V100 GPU. The motivation for focusing on these routines is that many scientific applications are either (1) based on batched and sparse linear algebra library routines or (2) composed of operations with very similar characteristics. Consequently, the performance gains for these benchmarks may be indicative of the acceleration we may see when porting a scientific computing application from a V100 platform to the A100 architecture, without applying additional code modifications.

In Figure 1, we are visualizing the speedups we get when replacing an NVIDIA V100 GPU with an NVIDIA A100 GPU without code modification. While the main memory bandwidth has increased on paper from 900 GB/s (V100) to 1,555 GB/s (A100), the speedup factors for the STREAM benchmark routines range between 1.6x and 1.72x for large data sets.

For the sparse matrix-vector product (SpMV) the performance improvements depend on the individual sparse data format, the kernel implementation and the specific problem characteristics, and are generally much lower than those for the STREAM benchmarks.

In the performance analysis for Ginkgo's iterative linear solvers, we focus on large test problems to ensure the bandwidth is saturated in the vector operations. Depending on the individual algorithm, Ginkgo's iterative solvers run between 1.5x and 1.8x faster on the A100 GPU over the V100 GPU.

For batched routines, we observe performance gains up to 1.6x that come "for free" by just switching to newer hardware architecture. It is worth mentioning that the A100 GPU provides tensor core acceleration for FP64 arithmetic. This is a new hardware capability that did not exist on the A100 predecessors, and is hence not yet reflected in the algorithm implementations available in open source libraries such as MAGMA.

While we cannot answer the question of whether these performance improvements justify the investment, it is clear that the NVIDIA team succeeded in delivering an architecture with a new focus that delivers considerable performance improvement over its predecessor – not just incremental acceleration. At the SCC, we are happy to host some of the first A100 worldwide, ready to be used by application scientists.

EU-Projekt DEEP Hybrid DataCloud bringt Werkzeuge für Benutzer und Entwickler von KI Anwendungen

Das von der EU geförderte Projekt DEEP Hybrid DataCloud¹ (DEEP) bietet eine digitale Arbeitsumgebung für Nutzende aus Praxis und Entwicklung von Künstlicher Intelligenz (KI) mit unterschiedlichem Fachwissen. Im Kontext der European Open Science Cloud (EOSC) werden Forschenden Werkzeuge zum Erlernen als auch für das Teilen und Bereitstellen von Anwendungen für Maschinelles Lernen und Deep Learning (ML/DL) zur Verfügung gestellt. Diese Werkzeuge sind lokal sowie auf Hybrid-Cloud-Systemen nutzbar. Das Projekt lief von November 2017 bis April 2020 und wurde anschließend von der Europäischen Kommission für seine außergewöhnlichen Ergebnisse und unmittelbare und potenziell vereinfachte Verwendung von KI-Technologie gelobt. Hier geben wir einen kurzen Überblick über die Hauptkomponenten und die Plattform aus Anwendersicht.

Valentin Kozlov



Der zentrale Einstiegspunkt der DEEP-Plattform für alle Nutzenden ist der DEEP Open Catalog² (Abbildung 1). Er bietet einen Marktplatz mit detailliert beschriebenen ML/DL-Anwendungen, entsprechenden Links zu den dazugehörigen Docker-Containern als Datei (den Container Images) und zum Quellcode. Alle Anwendungen nutzen die DEEPaaS-API und die Docker Container-Technologie. Die DEEPaaS-API basiert auf der OpenAPI-Spezifikation und bietet einen einfachen Zugriff auf ML/DL-Modelle. Der Einsatz von Containern gewährleistet die Reproduzierbarkeit von ML/DL-Modellen und die Flexibilität bei der Bereitstellung auf unterschiedlichen Rechnerplattformen. Die Vorhersagefunktion von im DEEP-Marktplatz aufgelisteten ML/DL-Anwendungen kann man sofort online via Webbrowser erreichen (OpenWhisk-Endpunkte). Nutzende können die volle Funktionalität von Anwendungen erkunden, indem sie entsprechende Docker-Images herunterladen und entweder lokal, in einer Cloud oder auf einem Hochleistungsrechner ausführen. Im letzteren Fall wird das udocker-Tool ein-

gesetzt, welches es ermöglicht, Docker-Images in einem Mehrbenutzersystem ohne Administratorrechte auszuführen. Registrierte Nutzerinnen und Nutzer der DEEP-Plattform können über das Training Dashboard auch eigene Anwendungen in der DEEP-Hybridinfrastruktur bereitstellen. Diejenigen, die eine neue Anwendung entwickeln möchten, können den Container der DEEP-Entwicklungsumgebung nutzen und aus der dafür bereitgestellten DEEP-Cookiecutter³-Vorlage für ML/DL-Anwendungen ein neues Projekt erstellen. Sobald der DEEP GitHub-Organisation eine neue Anwendung hinzugefügt wurde, können Entwickler davon auf Grund des im Projekt akzeptierten und implementierten DevOps-Ansatzes direkt profitieren, z.B. Jenkins-basierte Continuous Integration und Delivery (CI/CD).

Während der Projektlaufzeit ist die Anzahl der bearbeiteten Anwendungsfälle, einschließlich externer Anwendungsfälle und Mehrzweckmodule von 5 auf 25 gestiegen, was das breite Interesse der ML/DL-Communities an der Plattform zeigt. Unter den externen Anwendungsfällen wurden erst kürzlich drei Module als Beitrag zur Bekämpfung der COVID-19-Pandemie hinzugefügt.

Vom SCC beteiligten sich Marcus Hardt und Valentin Kozlov am Projekt. Der Beitrag des SCC umfasst insbesondere

den ‚Mit gutem Beispiel voran‘-Ansatz, wobei das Einbinden und Erstellen von Modellen sowie das Verwalten der DEEP-Entwicklungsumgebung mit Vorlagen für Deep-Learning-Projekte im Mittelpunkt steht, auch das FLAAT-Paket für die Authentifizierung mit OIDC-Zugriffstoken, den Beitrag zur DEEPaaS-API. Zusammen mit zwei KIT-Studierenden sind dazu mehrere Deep-Learning-Anwendungen entwickelt und in den Open Catalog integriert worden.

Am KIT nutzen Astroteilchenphysiker und Meteorologen entweder die DEEP-Plattform oder starten Docker Container mittels udocker-Hilfsprogramm auf dem hauseigenen Hochleistungsrechner ForHLR II. Das Helmholtz AI-Team am SCC (Seite 20) verwendet udocker zum Testen und Ausführen der Anwendungen auf dem neuen NVIDIA DGX A100 Subsystem des KIT-Hochleistungsrechners. Die DEEP-Lösungen werden auch in Folgeprojekten wie EOSC-Synergy verwendet und weiterentwickelt. Hier profitiert beispielsweise der „KIT Service on Ozone Assessment (O3as)“ von der Containerisierung des Dienstes, dem udocker-Tool, FLAAT, dem Infrastructure Manager, der Jenkins Pipeline Library und den weiterentwickelten Lösungen zur Qualitätssicherung von Software.

¹ deep-hybrid-datacloud.eu/

² marketplace.deep-hybrid-datacloud.eu/

³ cookiecutter.readthedocs.io

Über die Ergebnisse und Erfolge des Projekts wurde regelmäßig auf verschiedenen Konferenzen, Seminaren und Workshops berichtet. Bei der EOSC hub-Woche 2020 gewann die von Lara Lloret (Spanischer Forschungsrat CSIC) und Valentin Kozlov (KIT/SCC) präsentierte DEEP-Demo den Demonstrationswettbewerb.

EOSC Digital Innovation Hub (EOSC-DIH) und DEEP haben eine Kooperationsvereinbarung geschlossen, um die Adoption der Technologie für kleine und mittlere Unternehmen zu fördern und Kooperationen aufzubauen. Über das EOSC-Portal wird die DEEPaaS-Training-Facility für Maschinelles Lernen, Künstliche Intelligenz und Deep Learning auch für alle

Forschenden in der EOSC bereitgestellt. Das DEEP-Konsortium ist dabei, eine Kooperationsvereinbarung (Memorandum of Understanding) abzuschließen, um die DEEP-Plattform nachhaltig zu unterstützen. Die technischen Details und die Gesamtarchitektur der DEEP-Arbeitsumgebung finden Sie in [1].

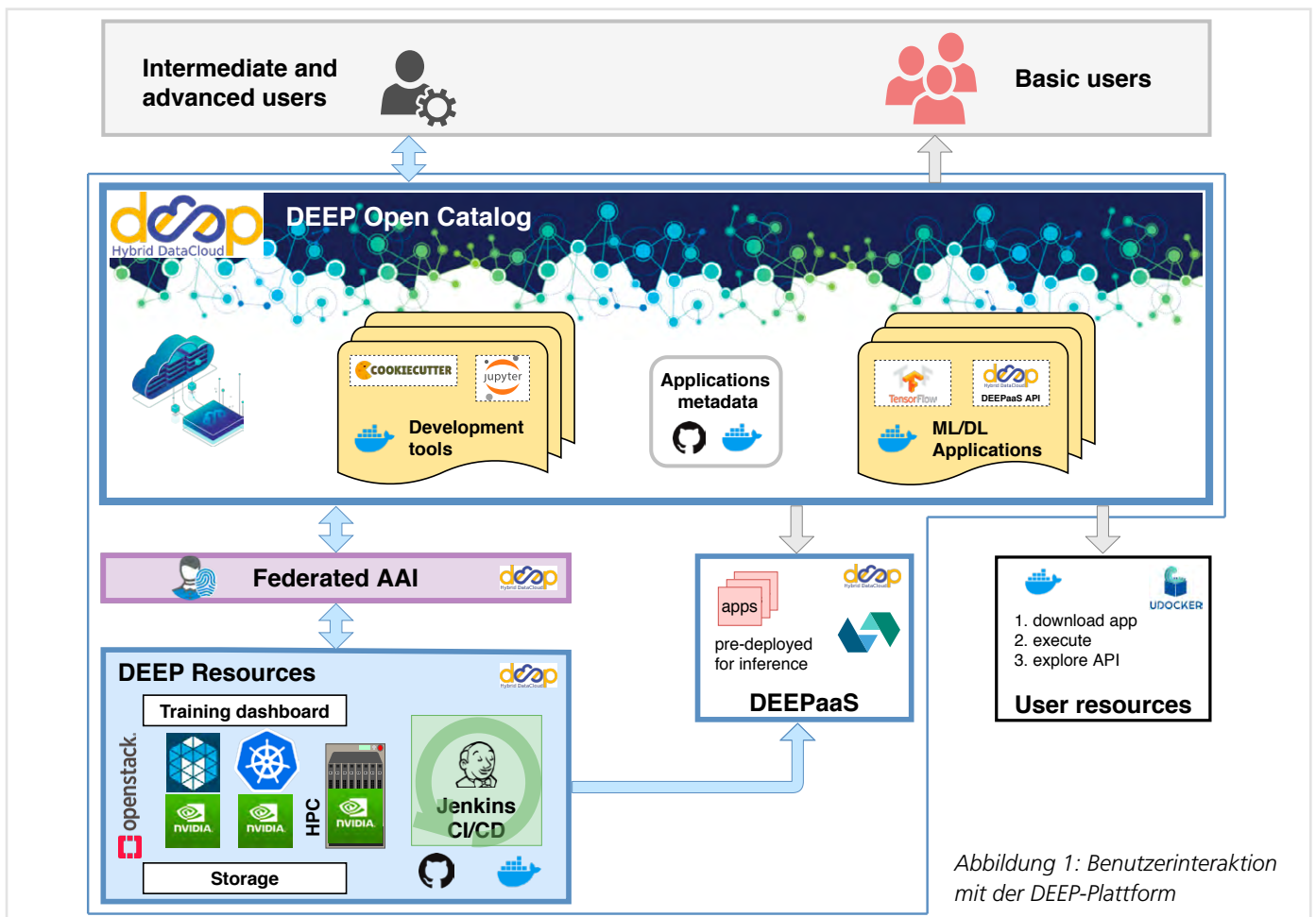


Abbildung 1: Benutzerinteraktion mit der DEEP-Plattform

The DEEP Hybrid DataCloud (DEEP) project¹, funded by the EU, offers a framework for users, practitioners and developers of Artificial Intelligence (AI) with various levels of expertise. It was successfully evaluated by the European Commission as the project that has delivered exceptional results with significant immediate and potential impact. The central entry point for all users is the DEEP Open Catalog², where AI applications are offered. Users get all necessary information and links, can immediately probe the applications for inference via OpenWhisk endpoints and download them as Docker images to explore the full functionality either locally, in a cloud, or in HPC. AI developers can start new projects from the existing DEEP template, leverage the DEEP development environment, and the Jenkins-based CI/CD. SCC scientists contributed to many aspects of the project. KIT researchers in Astroparticle physics, Meteorology, and the Helmholtz AI team either use the platform or benefit from the udocker tool on ForHLR-II or on the new NVIDIA DGX A100 system. EOSC-DIH and DEEP agreed to collaborate on business pilots. DEEPaaS Training Facility service is offered via EOSC-portal. The DEEP solutions are further used and developed in follow-up projects like EOSC-Synergy.

[1] A. Lopez Garcia et al., A cloud-based framework for machine learning workloads and applications, IEEE Access 8, 18681-18692 (2020), doi.org/10.1109/ACCESS.2020.2964386

Kompetenz in der KI-Forschung – Das Helmholtz AI Local Energy Consultant Team stellt sich vor

Seit dem 1. Oktober 2020 ist die „Helmholtz AI Local Energy Consultant“-Gruppe am SCC komplett. Das Team unterstützt in verschiedenen Projekten im Bereich der Energieforschung mit ihrer Expertise zur künstlichen Intelligenz. Über ein speziell dafür entwickeltes Helmholtz AI Voucher-System adressieren Forschungsgruppen am KIT und anderen Helmholtz-Zentren Consulting-Anfragen im Bereich Künstliche Intelligenz an die Gruppe.

Charlotte Debus

Künstliche Intelligenz (KI, engl. AI) ist eines der gefragtesten Forschungsfelder unserer Zeit. Die breite Verfügbarkeit von Rechenressourcen und Daten haben es ermöglicht, Maschinen das beizubringen, was das menschliche Gehirn intuitiv schon leistet: Mustererkennung und eine darauf basierte Entscheidungsfindung und Vorhersage. Das ist etwas, das sich sehr trivial anhört, und doch einen essentiellen Teil unseres analytischen Denkens und Verständnisses ausmacht. Bild- und Spracherkennung sind die bekanntesten Beispiele, mit denen jeder auf die ein oder andere Weise schon einmal Kontakt hatte, z.B. durch die Sprachsteuerung von Handys oder in Social Media-Plattformen beim Taggen von Personen auf Fotos. Aber auch Musikvorschläge bei Spotify, die Übersetzungen von Google Translate oder die vorgeschlagenen Navigationsrouten in verschiedenen Kartendiensten sind das Resultat cleverer KI-Algorithmen. In Zukunft werden solche Algorithmen in noch viel mehr Bereichen unseres alltäglichen Lebens auftauchen, man denke nur an Smart Homes oder Autonomes Fahren. KI-Anwendungen entspringen längst nicht mehr nur den Köpfen von Science Fiction-Autoren, sondern sind die Realität des 21. Jahrhunderts.

Helmholtz AI – kooperative KI-Forschung

Auch in der Forschung spielt KI und insbesondere das Maschinelle Lernen (ML) eine zunehmend zentralere Rolle. Dies spiegelt sich in der Auswertung wachsender Datenmengen, in der Überwachung von

Prozessen oder als Entscheidungshilfen im Design-Prozess neuer Experimente wider. Deshalb hat die Helmholtz-Gemeinschaft vor zwei Jahren die Helmholtz Artificial Intelligence Cooperation Unit (HAICU), seit kurzem unter dem Namen Helmholtz AI, ins Leben gerufen (www.helmholtz.ai). Unter der Koordination der Central Unit am Helmholtz Zentrum München (HMGU) wurden in den letzten Jahren fünf sogenannte Local Units an verschiedenen Helmholtz-Zentren implementiert (Abbildung 1), die sich jeweils mit einem der sechs Forschungsbereiche der Helmholtz-Gemeinschaft beschäftigen: Das Forschungszentrum Jülich mit Information & Schlüsseltechnologien, das Helmholtz-Zentrum Dresden-Rossendorf mit dem Forschungsbereich "Materie", das KIT mit Energie, das Deutsche Zentrum für Luft und Raumfahrt mit dem Kernthema Luftfahrt, Raumfahrt und Verkehr und das Helmholtz-Zentrum Geesthacht mit dem Themenfeld Erde und Umwelt. Das HMGU hat neben seiner koordinierenden Rolle auch die Funktion einer Local Unit zum Thema Gesundheit.

Die Local Units bestehen jeweils aus einer Helmholtz-Nachwuchsgruppe

(Young Investigator Group) und einem AI Consultant-Team. Während die Nachwuchsgruppen sich mit aktuellen KI-Fragestellungen in ihrem Forschungsfeld auseinandersetzen, stehen die Consultants anderen Forschungsgruppen mit ihrer KI-Expertise zur Seite, um gemeinsam angewandte KI-Forschungsfragen zu adressieren. Das Ganze funktioniert über ein sogenanntes Voucher-System: Forschungsgruppen aus allen Helmholtz-Zentren können ihre Arbeiten im Rahmen der jeweiligen Schwerpunkte in Form von kleinen bis mittleren Projektanträgen (Vouchern) einreichen und dadurch Hilfe von den Consultant-Teams anfordern. Voucher laufen zwischen zwei Wochen und sechs Monaten und können die verschiedensten Arten von Hilfestellung und Beratung beinhalten. Sei es, dass eine Gruppe noch völlig am Anfang steht mit ihrer KI-Forschung und Hilfe bei der Ideenfindung,

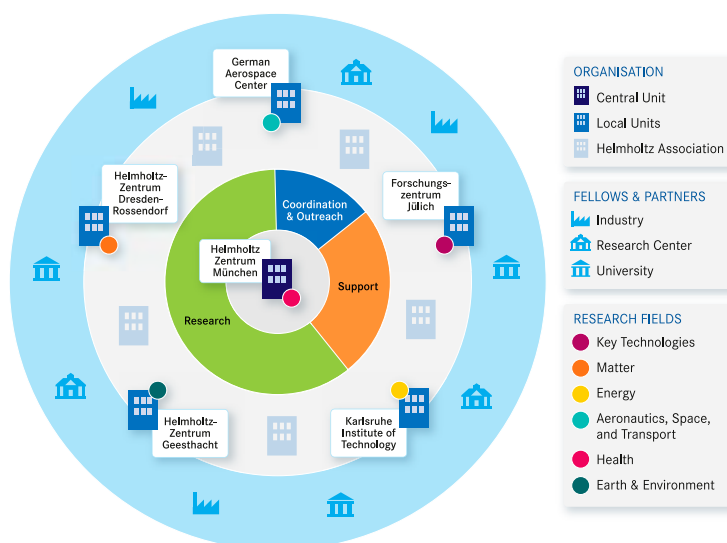


Abbildung 1: Die Struktur der Helmholtz AI Cooperation Unit

der Auswahl der Algorithmen und deren Implementierung benötigt, oder in einem Projekt arbeitet, in dem bereits erfolgreich KI-Methoden verwendet werden, diese jedoch verbessern oder auf Hochleistungsrechnersystemen zur Anwendung bringen will.

Das Consultant-Team am SCC



Das vierköpfige Team unter der kommissarischen Leitung von Markus Götz soll Arbeitsgruppen am KIT und anderen Helmholtz-Zentren bei ihren Forschungsfragen mit Methoden der Künstlichen Intelligenz unterstützen.

Die Voucher und Projekte kommen dabei aus verschiedensten Gebieten: z.B. Materialwissenschaften für die Batterieforschung, Entwicklung neuer Solarzellen oder Analysen von Stromnetzen oder chemischen Prozessen bei der Fusion. Diese Themengebiete haben einen gemeinsamen Nenner: Sie beschäftigen sich mit aktuellen Problemen der Energieforschung, dem Schwerpunkt der Helmholtz AI Local Energy-Einheit.

Bei der Arbeit an solch vielfältigen Projekten und Forschungsbereichen kommt dem Consultant-Team seine breite Expertise aus verschiedenen Anwendungsbereichen zugute. Als „Universaltalente“ bezeichnet Markus Götz deshalb seine drei Physiker scherzhaft. Er selbst ist Informatiker und beschäftigt sich mit skalierbaren und verteilten Algorithmen des Maschinellen Lernens. Er studierte an der Universität Potsdam Computer Science Engineering und promovierte zu Datenanalyse auf Hochleistungsrechnern an der Universität Reykjavik. Seit 2018 ist er am SCC tätig, zunächst als Projektleiter für das Pilotprojekt Helmholtz Analytics Framework und seit kurzem als kommissarischer Leiter des Consultant-Teams.

James Kahn ist Teilchenphysiker. Nach seinem Studium an der Universität Melbourne und der Promotion an der Ludwig-Maximilians-Universität München wechselte Kahn nach Karlsruhe ans SCC, wo er beim Grid Computing Centre Karlsruhe (GridKa) als Schnittstelle zum Belle II Experiment fungierte. Die Idee, Teilchenrekonstruktionen aus den Kollisionsexperimenten mittels tiefen neuronalen Netzen zu modellieren, brachte ihn schließlich in das Consultant-Team.



Daniel Coquelin hat sich während seines Masterstudiums mit der Entwicklung von Detektoren für die Teilchenphysik beschäftigt.

Im Anschluss war er am Forschungszentrum Jülich als einer der Hauptentwickler für das Helmholtz Analytics Toolkit (HeAT) zuständig. Seit März dieses Jahres ist er nun am SCC tätig und setzt seine Arbeiten zu datenparallelen neuronalen Netzwerken fort.

Die vierte im Bunde ist Charlotte Debus, die erst seit Oktober am KIT ist. Zuvor war sie als wissenschaftliche Mitarbeiterin am Deutschen Zentrum für Luft- und Raumfahrt in Köln, wo sie ebenfalls am Helmholtz Analytics Toolkit mitarbeitete. Debus hat in Heidelberg Physik studiert und am Deutschen Krebsforschungszentrum zum Thema Analyse von medizinischen Bildgebungsdaten mittels Maschinellen Lernen promoviert.



Zusammen bringen die Vier umfassendes Fachwissen im Bereich der Zeitreihen- und Bilddatenanalyse, Anomalie-Detektion, Netzwerk-Architekturen, Hyperparameteroptimierung und High Performance Computing an einen Tisch.

Forschungsunterstützung und weitere Aktivitäten

Die ersten Voucher sind schon gestartet, wie z.B. das Projekt BaumBauen zur Rekonstruktion von Teilchenschauern aus Belle II BB+-Zerfällen gemeinsam mit dem Institut für Experimentelle Teilchenphysik (ETP) des KIT und den Universitäten Bonn und Straßburg, oder im Projekt PerSeuS mit dem Lichttechnischen Institut (LTI) des KIT, in dem es um die Überwachung des Trocknungsprozesses bei der automatischen Herstellung von Perowskit-Solarzellen geht.

In den nächsten Wochen finden zudem die HelmholtzAI-Roadshows im Bereich Energie statt, bei denen das Konzept der Consultants den Arbeitsgruppen im Bereich Energieforschung an verschiedenen Helmholtz-Zentren vorgestellt und Ideen zu weiteren gemeinsamen Projekten gesammelt werden. Das Consultant-Team rechnet mit einer Vielzahl von Voucheranfragen, die unter anderem Biotreibstoffe, erneuerbare Energie oder auch Stromnetzplanung betreffen.

Das Helmholtz AI Local Energy Consulting-Team ist jederzeit offen für Projekt-Ideen, die KI-Methoden und -Expertise benötigen und nimmt diese unter consultant-helmholtz.ai@kit.edu gerne entgegen.

Competence in AI research – The Helmholtz AI Local Energy Consultant Team introduces itself

Since October 1st, the Helmholtz AI Consultant Team at SCC is running at full power. Under the comissionary lead of Markus Götz, scientists Charlotte Debus, James Kahn, and Daniel Coquelin will apply their AI expertise to assist in a range of projects in the area of energy research. Research groups at KIT and other Helmholtz centers can apply for assistance for short- to mid-term projects through the so-called voucher system.

Auf zu neuen Herausforderungen: Let's go Quantum

Das Quantencomputing ist nicht nur ein vielversprechendes Forschungsfeld, sondern verheißt mit potentiell exponentiellen Rechenbeschleunigungen für bestimmte Anwendungen ein hohes Innovationspotenzial. Insbesondere die Eigenschaft, effizient Strukturen in Daten zu erkennen, macht dessen Anwendung für Maschinelles Lernen fast schon offensichtlich. Ein abteilungsübergreifendes Team des SCC beschäftigt sich daher mit diesem Forschungsfeld und hat das Ziel, Algorithmen zu definieren und in Anwendungen zu übersetzen.

Eileen Kühn

Die Einflüsse des Quantencomputing auf Gesellschaft und Wirtschaft sind kaum absehbar. Insbesondere die Möglichkeit, mit Hilfe von Quantencomputern bestimmte Berechnungen durchführen zu können, bei denen klassische Computer an ihre Grenzen stoßen, trägt zu einem hohen Innovationspotenzial bei. Bereits im Jahr 1985 wurde der erste Quanten-Algorithmus von David Deutsch vorgestellt¹. Dieser Algorithmus benötigte weniger Rechenschritte als ein herkömmlicher Computer. Ein weiteres prominentes Beispiel ist der Algorithmus zur Primfaktorzerlegung, der von Peter Shor im Jahr 1994 vorgestellt wurde² und sogar exponentielle Rechenbeschleunigungen erzielen kann. Seitdem sind vielfältige Fortschritte in den unterschiedlichsten Forschungsfeldern, von Quantensimulationen in der Chemie, Materialwissenschaft und Hochenergiephysik, bis hin zu Optimierungproblemen beim Maschinellen Lernen, zu verzeichnen.

Auch für die Helmholtz-Gemeinschaft gehört das Quantencomputing zu einem der fünf Hauptbereiche, die im Rahmen der Quantentechnologien als strategisch relevant gelten³. Zu diesen zählen außerdem die Quantenkommunikation, Quantensensorik, Quantenmaterialien und Grundlagenforschung sowie Simulation und numerische Methoden. Als

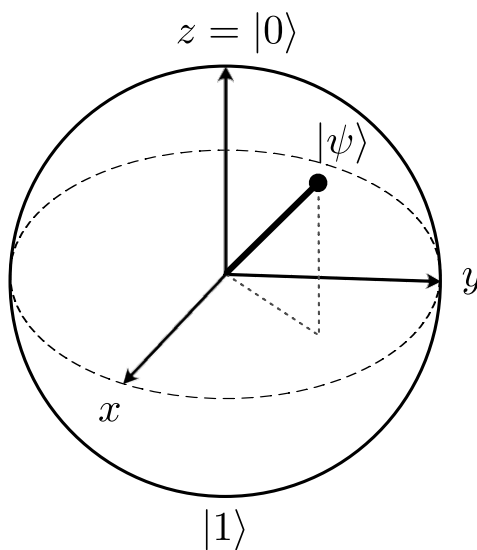


Abbildung 1: Die Bloch-Kugel wird im Quantencomputing genutzt, um die Überlagerung der Zustände eines Qubits grafisch darzustellen.

nationaler und europäischer Treiber und wissenschaftlich-technischer Wegbereiter trägt die Helmholtz-Gemeinschaft gemeinsam mit Partnern an Hochschulen, Forschungseinrichtungen und Industrie so auch direkt zu dem European Quantum Technology Flagship⁴ und dem Programm "Quantentechnologien – von den Grundlagen zum Markt" des BMBF⁵ bei.

Aktuelle Herausforderungen des Quantencomputing

Das Quantencomputing ist allerdings nicht nur vielversprechend, sondern geht gleichermaßen mit umfangreichen Herausforderungen einher. Eine dieser Herausforderungen ist die Realisierbarkeit universeller Quantencomputer mit einer skalierbaren Anzahl an Qubits und somit die technische Verfügbarkeit von Quantencomputern. Heutzutage arbeiten diverse Firmen an Cloud-Angeboten für ihre Quantencomputer, so zum Beispiel auch IBM, bei denen die breite Öffentlichkeit bis zu 15 Qubits (Stand November 2020) nutzen kann. Viele Quantencomputer mit eingeschränktem Zugriff existieren aber auch in Laboren und Forschungseinrichtungen und operieren mit 50 - 100 Qubits. Bei diesen Quantencomputern handelt es sich allerdings nicht um universelle Quantencomputer. Stattdessen wurde für diese Generation der Quantencomputer der Begriff Noisy Intermediate-Scale Quantum (NISQ) geprägt (siehe Box).

Trotz der großen Fortschritte in den letzten Jahren steht die anwendungsorientierte Forschung im Quantencomputing erst am Anfang. Auch die Quantentechnologie muss am ehesten noch mit den klassischen Computertechnologien

¹ David Deutsch: Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. In: Proc. R. Soc. Lond. A 8 July 1985 vol. 400 no. 1818, S. 97–117. doi:10.1098/rspa.1985.0070

² Peter W. Shor: Algorithms for quantum computation: Discrete logarithms and factoring. In: Proceedings, 35th Annual Symposium on Foundations of Computer Science, Santa Fe, NM, November 20–22, 1994. IEEE Computer Society Press, 1994, S. 124–134

³ www.helmholtz.de/en/research/quantum-technologies/

⁴ qt.eu

⁵ www.bmbf.de/de/quantentechnologien-7012.html

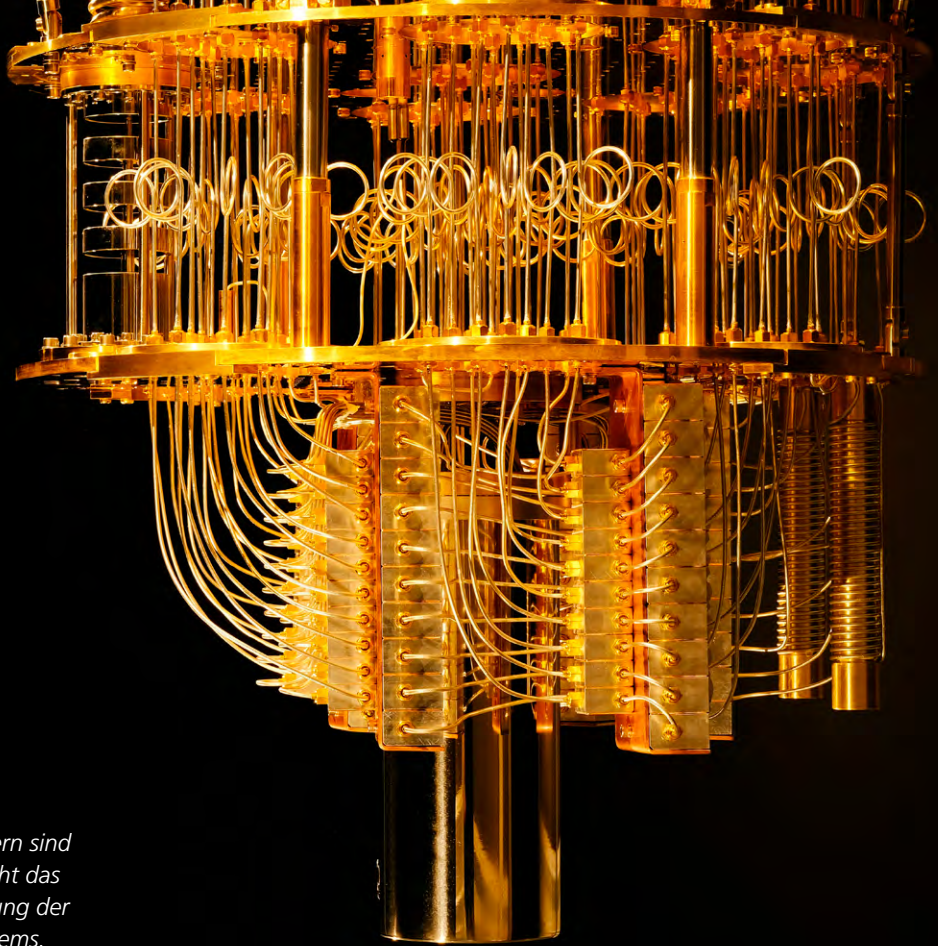


Abbildung 2: Die Fotos von Quantencomputern sind ziemlich beeindruckend, zeigen allerdings nicht das eigentlich Herzstück, die technische Realisierung der Qubits, sondern die Komplexität des Kühlsystems. (Credit: Graham Carlow, IBM)

der 1950er Jahre verglichen werden. Nichtsdestotrotz schreitet die technische Entwicklung unaufhaltsam voran: Beispielsweise sieht die Roadmap von IBM vor, noch in dieser Dekade Rechnergebilde aus 1 Million Qubits zu präsentieren.

Eine weitere Herausforderung besteht in der Identifikation der Algorithmen und Anwendungsfälle, die effizient durch einen Quantencomputer verarbeitet werden können, also beispielsweise eine exponentielle Beschleunigung ermögli-

chen, auf NISQ-Computern ausgeführt werden können oder sogar neuartige Algorithmen beschreiben.

Rechnen mit Qubits

Während ein herkömmlicher Rechner mit Bits rechnet, nutzt ein Quantencomputer Quanten-Bits – die sogenannten Qubits. Ein Qubit kann nicht nur den Wert 0 oder 1 annehmen, sondern auch jede beliebige Kombination aus beiden, also prinzipiell unendlich viele Zustände (siehe Abbildung 1). Dieser Zustand kann im Allgemeinen allerdings nicht ausgelesen werden: Die Messung des Qubits ergibt stattdessen zufällig einen der beiden möglichen Messwerte, wobei die Wahrscheinlichkeit jedes Messwertes durch den Zustand des Qubits vor der Messung bestimmt wird. Auch ein erneutes Messen erlaubt es nicht, weitere Messwerte dieser Wahrscheinlichkeitsverteilung auszulesen, da die Messung den Zustand des Qubits ändert.

Eine weitere besondere Eigenschaft von Quantencomputern ist die Möglichkeit der Verschränkung von Qubits. Salopp gesagt, hängen

die Eigenschaften miteinander verschränkter Qubits voneinander ab: Verändert man nun die eine Eigenschaft eines Qubits, wirkt sich diese Änderung gleichzeitig auch auf die mit ihm verschränkten Qubits aus. Die verschränkten Qubits können daher nicht mehr getrennt voneinander betrachtet werden, sondern nur noch als Gesamtsystem.

Eine Konsequenz dieser Verschränkung ist, dass man nicht mehr von den Zuständen der einzelnen Qubits sprechen kann. Sie lassen sich nur noch mit einer gemeinsamen Funktion beschreiben. Insofern kann man auch nur das Gesamtsystem messen. Wenn eines der verschränkten Qubits gemessen wird, so betrifft dies auch den Zustand des mit ihm verschränkten Qubits, unabhängig davon, wie weit diese voneinander entfernt sind. Die Messung eines einzelnen Qubits legt also das Messergebnis für das andere Qubit fest.

Noisy Intermediate-Scale Quantum Computer

Wenn wir über Quantencomputer reden, meinen wir in der Regel universelle, fehlertolerante Geräte. Diese Quantencomputer wären in der Lage, unter anderem die von Peter Shor vorgestellte Primfaktorzerlegung durchzuführen. Dafür sind allerdings Millionen von Qubits nötig, um Fehlerkorrekturen durchzuführen. Diese sind notwendig, da durch Wechselwirkung der Qubits untereinander und mit ihrer Umgebung sehr große Störeffekte und somit enorme Rechenfehler bei Quantencomputern auftreten können. Da die meisten Algorithmen aber extrem anfällig für Rauschen sind, werden mehr physikalische Qubits benötigt, um fehlertolerante logische Qubits abzubilden. Anlässlich einer Keynote-Präsentation im Jahr 2017 prägte John Preskill, einer der führenden Köpfe des Quantencomputings, erstmals den Begriff Noisy Intermediate-Scale Quantum Computer. Gemeint ist ein Quantencomputer mit etwa 50 bis 100 Qubits, der insbesondere zwei Eigenschaften erfüllt: Er stellt nur eine relativ kleine Anzahl an Qubits zur Verfügung und ist sehr fehleranfällig, da die Anzahl der Qubits nicht für eine ausreichende Fehlerkorrektur ausreicht. Auch die Fehlerrate der Operationen, der sogenannten Quanten-Gatter, liegt bei 1:100 bis zu 1:1000 und somit um viele Größenordnungen über der Fehlerrate herkömmlicher Computer.

Quantum Machine Learning – ein neues Forschungsfeld des SCC

Ein besonderer Vorteil von Quantencomputern besteht in ihrer Fähigkeit, Strukturen in Daten zu erkennen. Diese Eigenschaft begründet unter anderem das Forschungsfeld des Quantum Machine Learnings. Ähnlich wie bei dem übergeordneten Thema des Quantencomputings, ist auch hier die Forschungsgemeinschaft noch auf der Suche nach neuen Algorithmen und Implementierungen, um die Überlegenheit von Quantencomputern im Bereich des Maschinellen Lernens zu beweisen und praktische Implementierungen zu liefern.

Vor diesem Hintergrund beschäftigt sich ein kleines, aber hochmotiviertes Team von Mitarbeitenden aus unterschiedlichen Abteilungen des SCC seit wenigen Monaten mit theoretischen und praktischen Fragestellungen des Quantum Machine Learnings. Insbesondere diese abteilungsübergreifende Zusammenarbeit ermöglicht es, von den Erfahrungswerten der theoretischen und experimentellen Physik, der Informatik sowie der langjährigen Expertise in unterschiedlichen Projekten des Maschinellen Lernens zu profitieren. Eine der zentralen Forschungsfragen ist, welche konkreten Anwendungsszenarien oder Teilprobleme des Maschinellen Lernens sich für die Berechnung mit

einem Quantencomputer eignen und wie sich die notwendigen Algorithmen dafür entwickeln und in Applikationen übersetzen lassen. Ziel ist es, Bausteine für Quantenalgorithmen zu definieren, um sie für die Anwendungsprogrammierung einfach nutzbar zu machen und so zukünftig auch die Integration von Quanten-Beschleunigern zu ermöglichen und frühzeitig die notwendige Fachkompetenz aufzubauen.

Towards new challenges: Let's go Quantum

Quantum computing is not only one of the most promising fields of research today, but also promises a high innovation potential with potentially exponential speedups for certain tasks. Since the first quantum algorithms were designed, such as the first quantum algorithm in 1985 by David Deutsch or Peter Shor's prime factorisation in 1994, a variety of advances have been made in a wide range of research fields. But even today, the research field still comes with extensive challenges. One of these challenges is the feasibility of universal quantum computers with a scalable number of qubits and thus the technical availability of quantum computers. Today, one speaks of so-called Noisy Intermediate Scale Quantum (NISQ) computers, which operate with 50 - 100 qubits.

Despite the great progress in recent years, application-oriented research is still in its infancy and can best be compared to the classical computer technologies of the 1950s. Also identifying algorithms and use cases that can be efficiently processed on a quantum computer, for example, with exponential speedup, also remains a current challenge. Especially the ability to efficiently recognise structures in data makes the application in the field of machine learning almost obvious. An interdepartmental team of SCC is therefore engaged in this research area and has the goal of defining algorithms and translating them into applications.

NFDI – Nationale Forschungsdateninfrastruktur am SCC

Forschungsdaten, die bisher oft nur lokal, unstrukturiert und vorübergehend gespeichert werden, sollen in der Nationalen Forschungsdateninfrastruktur (NFDI) in standardisierter und systematischer Weise erfasst, erschlossen, nachhaltig gesichert und langfristig zugänglich und nachnutzbar gemacht werden. Im Oktober haben die ersten 9 von mittelfristig geplanten 30 Fachkonsortien damit begonnen, Dienste und Standards für das Forschungsdatenmanagement zu entwickeln. Das SCC ist in an zwei der ersten Konsortien in der NFDI maßgeblich beteiligt.

Felix Bach, Rainer Stotzka



Abbildung 1: Sabine Brünger-Weilandt (Direktorin FIZ Karlsruhe), Holger Hanselka (Präsident des KIT), Eva Lübke (kaufmännische Leiterin der NFDI), York Sure-Vetter (KIT, Direktor der NFDI) und Frank Mentrup (OB Karlsruhe) durchschnitten das rote Band zur Eröffnung der NFDI-Geschäftsstelle in Karlsruhe. Foto: KIT

Forschungsdaten werden bisher oft nur lokal, unstrukturiert und vorübergehend gespeichert. Dadurch gehen wertvolle Daten für eine Nachnutzung verloren. Mit der Nationalen Forschungsdateninfrastruktur (NFDI) soll sich dies ändern.

Sie wird Lösungen dafür erarbeiten, wie Daten in Wissenschaft und Forschung in standardisierter und systematischer Weise

erfasst, erschlossen, nachhaltig gesichert und nachnutzbar zugänglich gemacht werden können. Die NFDI ist als vernetzte Struktur eigeninitiativ agierender Konsortien mit unterschiedlicher fachlicher Ausrichtung aufgebaut, die in drei Förderrunden Anträge erarbeiten und einreichen. Das SCC ist an zwei erfolgreichen Konsortien der ersten Runde¹ beteiligt, die seit 1.10.2020 gefördert werden: NFDI4Chem

und NFDI4Ing. Die Geschäftsstelle der Nationalen Forschungsdateninfrastruktur (NFDI) ist in Karlsruhe angesiedelt (Abbildung 1).

¹ www.kit.edu/kat/pi_2020_052_nationale-forschungsdateninfrastruktur-drei-konsortien-mit-beteiligung-des-kit-gefördert.php

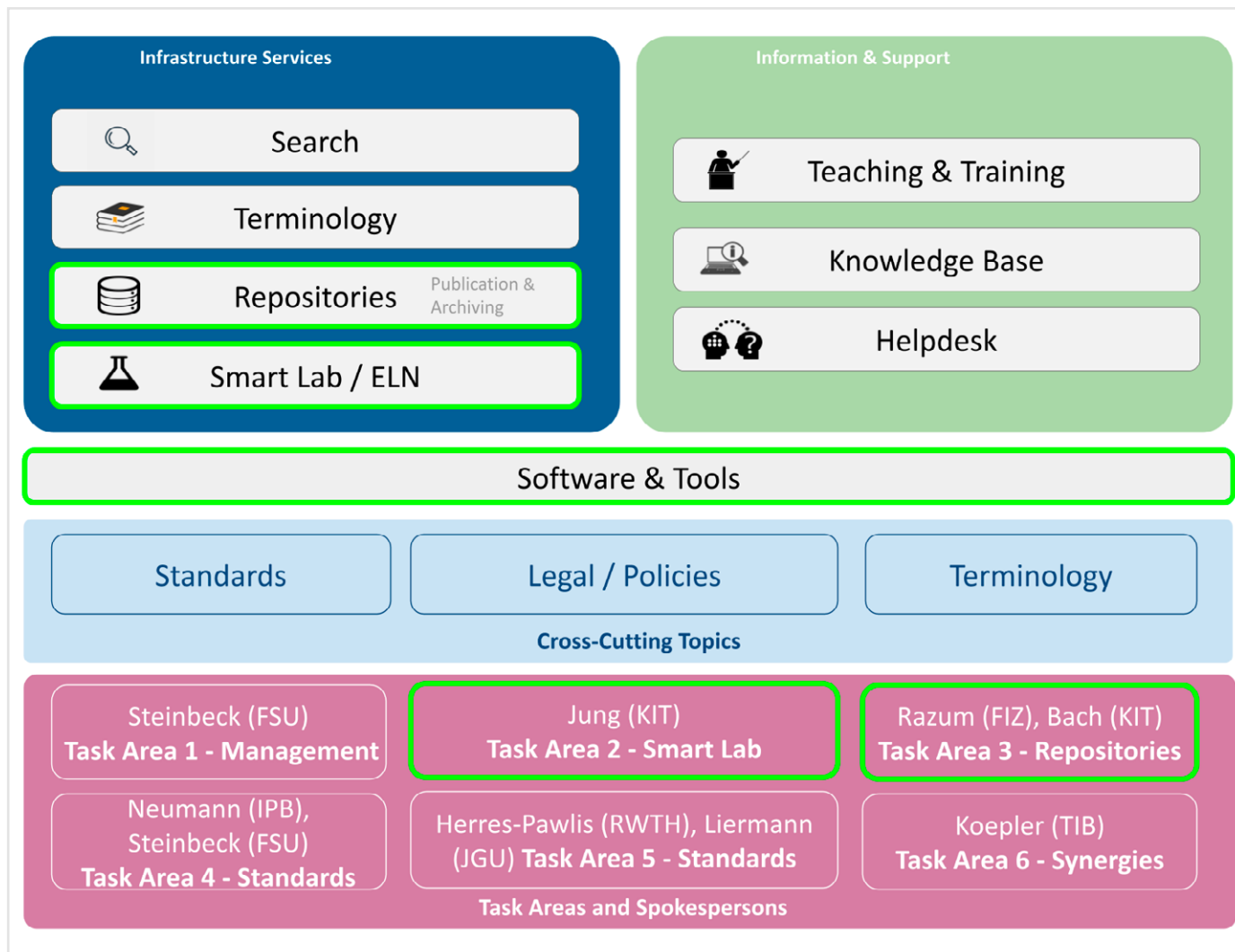


Abbildung 2: Datenflüsse und Arbeitsbereiche der NFDI4Chem, mit grün markierten SCC-Schwerpunkten



Die Vision von **NFDI4Chem²** ist die Digitalisierung aller Arbeitsprozesse in der chemischen Forschung. Hierzu sollen Dienste auf- und ausgebaut werden, die Forschende dabei unterstützen, Forschungsdaten zu sammeln, zu speichern und zu archivieren, zu verarbeiten und zu analysieren. Darüber hinaus sollen die Daten in Repositorien zusammen mit beschreibenden Metadaten und dauerhaften Identifikatoren (DOIs) publiziert und somit referenzierbar und wiederverwendbar

gemacht werden können. NFDI4Chem vertritt als Fachkonsortium alle Disziplinen der Chemie und arbeitet dazu eng mit den großen Fachgesellschaften zusammen. Die Chemie unterstützt und beeinflusst zahlreiche andere Wissenschaftsdisziplinen wie Medizin, Biologie, Materialforschung, Ingenieurwesen oder die Energieforschung, weshalb auch eine interdisziplinäre Nutzung der Daten mitbedacht wird. In der Anfangsphase liegt der Fokus darauf, Daten zu Molekülen und deren Reaktionen sowie Daten für deren experimentelle und theoretische Charakterisierung für Chemikerinnen und Chemiker strukturiert erfassbar und gemäß den FAIR Prinzipien³ nachnutzbar zu machen. Hierfür wurde im Vorfeld eine Studie⁴ zu den Arbeitsweisen und Anforderungen in den diversen chemischen Disziplinen aus Nutzersicht

durchgeführt. Das KIT ist mit mehreren Instituten sowie einer Co-Sprecherin und einem Co-Sprecher an der NFDI4Chem beteiligt: Nicole Jung vom Institut für Organische Chemie (IOC) und vom Institut für Biologische und Chemische Systeme (IBCS) leitet die Task Area Smart Lab, Felix Bach vom SCC die Task Area Repositorien. Beide bringen mit dem Science Data Center MoMaF (Molekulare Materialforschung)⁵, dem elektronischen Laborbuch Chemotion⁶ ELN und dem Chemotion Repositoryum starke gemeinsame Vorarbeiten ein. Am SCC werden in den Themenbereichen Smart Lab und Repositorien in enger Zusammenarbeit mit den Projektpartnern Dienste aus- und aufgebaut und miteinander vernetzt (Abbildung 2). Details finden sich im NFDI4Chem-Antrag⁷, der im RIO Journal publiziert wurde.

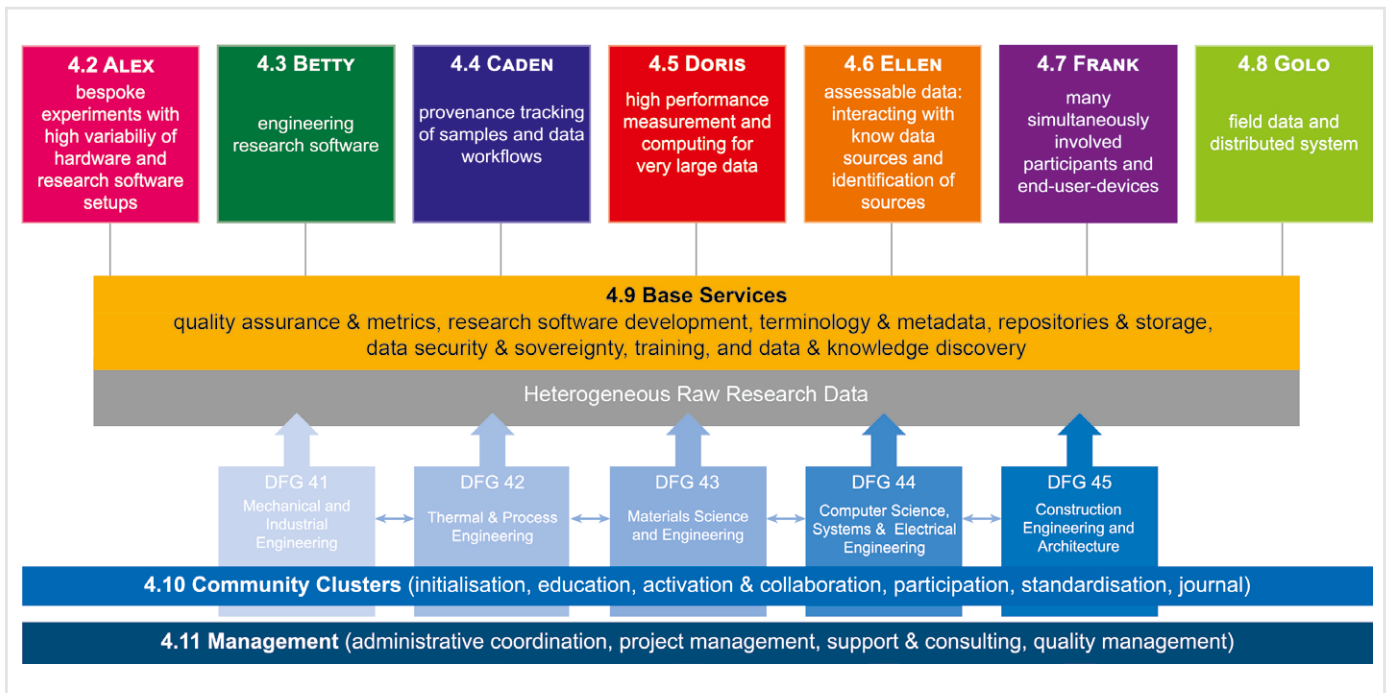


Abbildung 3: Die strukturgebenden Elemente in NFDI4ing sind die Archetypen, die für Klassen von Problemstellungen im Forschungsdatenmanagement stehen. Archetypen decken typische, datengenerierende Ingenieurstätigkeiten ab, für die im Konsortium Lösungen entwickelt werden. (Bildquelle NFDI4ing.de)



NFDI4ing⁸ ist ein Konsortium der Ingenieurwissenschaften und fördert das Management technischer Forschungsdaten. Die NFDI4ing-Initiative wurde schon 2017

gegründet und steht in engem Austausch mit Forschenden aller ingenieurwissenschaftlichen Fachgebiete (Abbildung 3). Das Konsortium bietet einen einzigartigen methodenorientierten und nutzerzentrierten Ansatz, um technische Forschungsdaten FAIR - auffindbar, zugänglich, interoperabel und wiederverwendbar - zu machen. Eine wichtige Herausforderung ist dabei die große Anzahl von Teildisziplinen der Ingenieurwissenschaften und ihre fachspezifischen Eigenheiten. Das KIT ist mit einer Co-Sprecherin, Britta Nestler aus dem Institut für Angewandte Materialien

(IAM), und einem Co-Sprecher, Achim Streit aus dem Steinbuch Centre for Computing (SCC), beteiligt.

Im Rahmen von NFDI4ing entwickelt und implementiert das SCC in enger Kooperation mit den Partnern die Konzepte für förderierte Forschungsdateninfrastrukturen, Datenmanagement-Prozesse, Repositorien und Metadatenmanagement. Der NFDI4ing-Antrag⁹ beschreibt die geplante Forschungsdateninfrastruktur im Detail.

NFDI – National Research Data Infrastructure at SCC

Research data contain important scientific results that may be unique to produce, but are still often stored locally, unstructured and temporarily. In consequence, valuable data can't be reused by other scientists. The National Research Data Infrastructure (NFDI) is designed to change this. It will develop solutions for the standardized and systematic collection, cataloguing, sustainable storage and reuse of data in science and research. The NFDI is a collaborative structure of consortia representing a variety of disciplines that act on their own initiative. They develop and submit proposals in three funding rounds for a five year project with the option to be extended by additional five years. The SCC is involved in two successful consortia of the first round that are starting their work on first of October 2020: NFDI4Chem and NFDI4ing.

² nfdi4chem.de

³ www.force11.org/group/fairgroup/fairprinciples

⁴ onlinelibrary.wiley.com/doi/full/10.1002/zaac.202000339

⁵ momaf.scc.kit.edu

⁶ chemotion.net

⁷ riojournal.com/article/55852/

⁸ nfdi4ing.de

⁹ doi.org/10.5281/zenodo.4015200

Landesprojekt bwCard: Digitale Identitätskarte

Ziel des Projektes bwCard ist die Entwicklung von Lösungen und Konzepten, wie vorhandene Chipkarten, seien es Studenausweise oder Ausweise von Mitarbeitenden, landesweit eingesetzt werden können. Das Projekt gliedert sich in drei Themenschwerpunkte, die föderative Zusammenarbeit, die Konzeption von Pilotanwendungen und die Evaluation einer E-ID. Mit Abschluss des Projektes Ende 2021 werden die Grundlagen für eine landesweit einsetzbare Chipkarte geschaffen sein, die für viele Dienste die Basisinfrastruktur zur Verfügung stellt.

Axel Maurer

Daniel studiert an der Pädagogischen Hochschule (PH) in Karlsruhe Lehramt Grundschule. Zu Beginn des Studiums hat er eine Chipkarte als Studenausweis bekommen. Die Karte dient gleichzeitig als Bibliotheksausweis und als Bezahlkarte für die Mensa. Während einer Seminararbeit stellte Daniel fest, dass die Literatur an der PH-Bibliothek thematisch nicht ausreicht. Eine Recherche ergibt, dass er noch Bücher aus der Badischen Landesbibliothek (BLB) braucht, und einige Zeitschriftenartikel bei der KIT-Bibliothek wären auch sehr hilfreich. Daniel registriert sich und seine Karte bei den beiden Bibliotheken und kann ab jetzt mit dem Studenausweis auch die BLB und die KIT-Bibliothek nutzen. Hier kann er mit seiner Karte nicht nur Bücher ausleihen, sondern auch den 24-Stunden-Zugang nutzen. Und das alles mit nur einer Karte, genauso wie es die Studierenden und Mitarbeitenden nahezu aller Hochschulen in Karlsruhe können.

Bis das alles so reibungslos funktioniert, mussten einige Hürden überwunden und Verabredungen getroffen werden. Schließlich müssen die Einrichtungen elektronisch erkennen, dass es sich um eine berechnete Karte handelt, und die Bibliotheken müssen in Realzeit über eventuelle Sperren informiert werden. Die Basis bildet dabei das Kartenmanagementsystem der KIT-Card¹. Alle beteiligten Hochschulen nehmen an der Karlsruher Kartenkooperation teil und nutzen als Produktionsgemeinschaft damit das Kartenmanagementsystem, das am SCC betrieben wird, für Produktion und Verwaltung ihrer Chipkarten. Über dieses werden die technischen Vereinbarungen umgesetzt und Sperrinformationen bereitgestellt, die im Bedarfsfall auch von

der BLB und der Stadtbibliothek abgerufen werden².

Trotz der vielen Arbeit hat Daniel noch etwas Freizeit und war in Heidelberg. Prompt hat er sich dort verliebt und verbringt jetzt viel Zeit in Heidelberg. Natürlich soll die Seminararbeit nicht zu kurz kommen. Dazu möchte er die Universitätsbibliothek Heidelberg nutzen. Hier stößt er auf unerwartete Hürden. Er braucht einen neuen Bibliotheksausweis und muss dazu eine Studienbescheinigung seiner Hochschule vorlegen. In der Mensa kann er weder seinen Studenausweis zum Bezahlen nutzen, noch zum Studierendenpreis essen gehen.

Die bwCard soll in Zukunft all das und noch mehr an allen Hochschulen in Baden-Württemberg ermöglichen. Die Universitäten in Baden-Württemberg haben dazu das Projekt „bwCard: Digitale Identitätskarte“ aufgesetzt. Es wird vom Ministerium für Wissenschaft, Forschung und Kunst (MWK) gefördert, wurde im Dezember 2018 begonnen und läuft noch bis Ende November 2021. Die Projektleitung liegt beim SCC. Die Ergebnisse des Projektes stehen allen Hochschulen in Baden-Württemberg offen³.

Das Projekt gliedert sich in drei Themenschwerpunkte:

- Föderative Zusammenarbeit: Verlässliche bwCard (Federführung KIT)
- Konzeption von Pilotanwendungen (Federführung Universität Konstanz)
- Evaluation einer E-ID (Federführung Universität Mannheim)

Föderative Zusammenarbeit

Im Themenschwerpunkt 1 geht es darum, die oben geschilderten Szenarien auf Landesebene zu ermöglichen. Aus Sicht der bwCard wäre eine einfache Lösung, alle Hochschulen an die Produktionsgemeinschaft anzuschließen; diese macht dann entsprechende Vorgaben, vergleichbar mit der bereits in Karlsruhe eingesetzten Lösung. Die Erweiterung der Produktionsgemeinschaft ist in der Tat ein wichtiger Teil des Projektes. Man kann jedoch nicht einfach alle Kartenmanagementsysteme an allen Hochschulen ersetzen. Das wäre für einige Universitäten zu teuer und zu aufwändig, da damit viele angebundene Systeme, wie Bibliothekssysteme oder Zutrittssysteme, erweitert werden müssten.

In Fällen, in denen man etwas gemeinsam erreichen möchte, aber teilweise an lokale Gegebenheiten gebunden ist, hat sich der föderative Ansatz bewährt. Dieser hat im Land auch in anderen Bereichen zu verlässlichen Lösungen geführt, wie z.B. das föderierte Identitätsmanagement bwIDM zeigt. Der Sinn dahinter ist, man einigt sich auf gemeinsame Parameter und überlässt es den Teilnehmern der Föderation, wie sie diese Parameter im Detail realisieren. Im Fall von bwCard hat man sich auf drei gemeinsame Merkmale verständigt, die die Karten einer Einrichtung erfüllen müssen. Zum einen werden alle bwCards mit einem sichtbaren Logo ausgestattet, sie haben den gleichen Chip und bekommen eine einheitliche Kennung auf dem Chip, die „bwCardApp“ (Abbildung 2). Die bwCard-App dient zur maschinenlesbaren Nutzung der Karte.

Anhand der dort gespeicherten Daten kann jede Karte innerhalb des bwCard Verbundes eindeutig zugeordnet werden. Die Daten der bwCard-App haben keinen direkten Personenbezug, sondern dienen ausschließlich der Identifikation der Karte. Zusätzlich besteht für jedes Mitglied der bwCard-Föderation, nach Freigabe durch die Karteninhaberin oder den -inhaber, die Möglichkeit, jederzeit den Status der Karte auf Basis der Daten in der bwCard-App bei der ausgebenden Einrichtung abzufragen.

Die Statusabfrage orientiert sich sehr stark an den bewährten Lösungen aus bwIDM und wird bei den Ausführungen zur E-ID zum Themenschwerpunkt 3 näher betrachtet.

Neben der Föderation wird im Themenschwerpunkt 1 auch die Erweiterung der Produktionsgemeinschaft behandelt. Der grundsätzliche Aufbau des Kartenmanagementsystems ist in den SCC-News 1/2018, S. 21 beschrieben. Jede teilnehmende Einrichtung hat auf dem System einen eigenen Mandanten. Sowohl die

Daten als auch die Konfiguration, die Definition der Prozesse bis hin zu Layout und Kodierung der Karte sind komplett getrennt und können von der Einrichtung selbst bestimmt werden. Beteiligt an der Produktionsgemeinschaft mit jeweils eigenem Mandanten sind die Hochschule Karlsruhe, die Duale Hochschule Baden-Württemberg Karlsruhe, die Pädagogische Hochschule und die Hochschule für Musik. Mit der Hochschule für Gestaltung und der Staatliche Akademie der Bildenden Künste Karlsruhe laufen Beitrittsprojekte. Darüber hinaus ist die Universität Freiburg bereits Mitglied der Produktionsgemeinschaft. Mit den Universitäten Konstanz, Stuttgart und Hohenheim sowie weiteren Hochschulen laufen ebenfalls Beitrittsprojekte.

bwCard connector

Ausweisproduktion bedeutet immer auch die Verarbeitung personenbezogener Daten. Die meisten Einrichtungen drucken zumindest den Namen und oft auch ein Bild der Karteninhaberin auf die Karte. Diese Daten müssen zur Produktion der

Karte im Kartenmanagementsystem zur Verfügung stehen. Daher spielt der Datenschutz in der Produktionsgemeinschaft eine erhebliche Rolle. Hier wurde eine Lösung vorangetrieben, die man mit „Datenschutz by Design“ bezeichnen kann. Personenbezogene Daten werden immer nur bei der teilnehmenden Hochschule gespeichert und nur anlassbezogen an das zentrale Kartenmanagement am KIT übertragen, wo sie nach der Verarbeitung sofort wieder gelöscht werden. Sie sind immer nur im jeweiligen Mandanten sichtbar, der ausschließlich durch Mitarbeitende der jeweiligen Einrichtung bedient werden kann. Sowohl der Zugang zum System als auch die Übertragung personenbezogener Daten erfolgt ausschließlich verschlüsselt und authentifiziert mit Zertifikat und Passwort. Damit die Mitglieder der Produktionsgemeinschaft möglichst einfach die lokalen Daten zur Verfügung stellen können und dabei auch beide Modelle zur Verfügung haben, wurde eine eigene Softwarekomponente entwickelt, der „bwCard connector“.

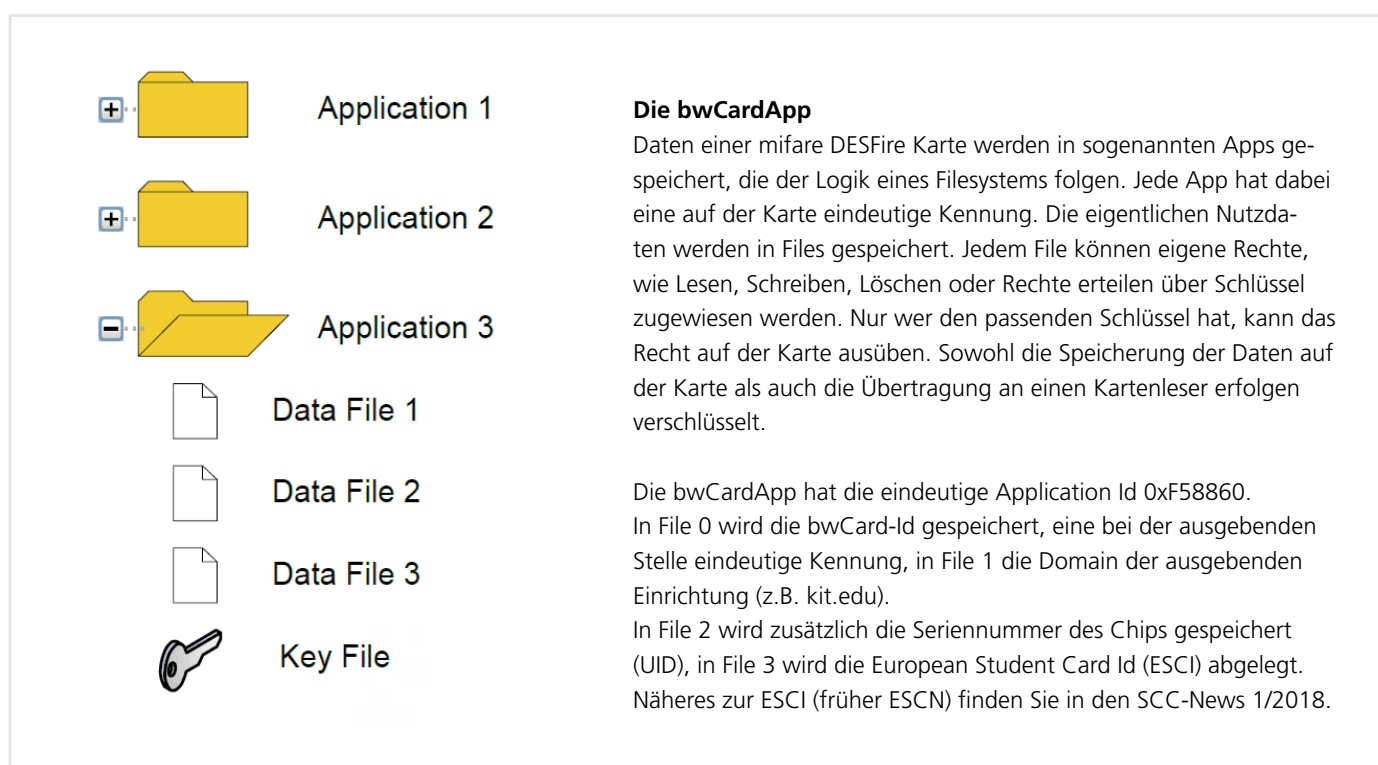


Abbildung 2: Die bwCardApp

Cashless Campus

Daniel ist es gewohnt, an vielen Stellen mit seinem Studenausweis Leistungen zu bezahlen. Sei es in der Mensa, eventuell anfallende Mahngebühren in der KIT-Bibliothek oder mal ein Mineralwasser am Getränkeautomat. Die Bezahlung mit der Chipkarte hat allerdings auch ihre Grenzen. Werden Gebühren fällig oder möchte Daniel auf Exkursion gehen, kann es passieren, dass er wieder zu Alternativen oder zum bewährten Bargeld greifen muss. Das mag für ihn zwar nur ein kleines Ärgernis sein, für die Hochschulen stellt der Umgang mit Bargeld jedoch eine zunehmend große und teure Herausforderung dar.

Auf dem Campus jeder Hochschule steht heute ein breites Angebot an kostenpflichtigen Zusatzdienstleistungen zur Verfügung, wie das Mensaangebot, ergänzt durch Getränke- und Essensautomaten, den Hochschulsport, die Dokumentlieferdienste der Bibliothek oder das Bezahlen von fälligen Gebühren.

Allen diesen Angeboten ist gemeinsam, dass die meisten zu bezahlenden Beträge kleiner als 50 € sind und sehr häufig anfallen. Bei diesen Kleinbeträgen verursacht die direkte Abbuchung über ein Bankkonto hohe Kosten auf Seiten der Hochschule, welche in keinem Verhältnis zum eigentlichen Abbuchungsbetrag stehen. Bezüglich des Einsatzes von Bargeld gibt es weitgehende Regelungen, wie im öffentlichen Dienst mit Bargeld umgegangen werden muss⁴, die zu hohem administrativen Aufwand führen. Als Alternative haben sich in den letzten Jahren Guthabekarten (elektronische Geldbörse), oft betrieben von den Studierendenwerken, etabliert. Hier kann ein mittlerer Geldbetrag (z.B. 100 €) entweder bar oder über Lastschriftverfahren aufgebucht und damit die Kleinbeträge direkt über

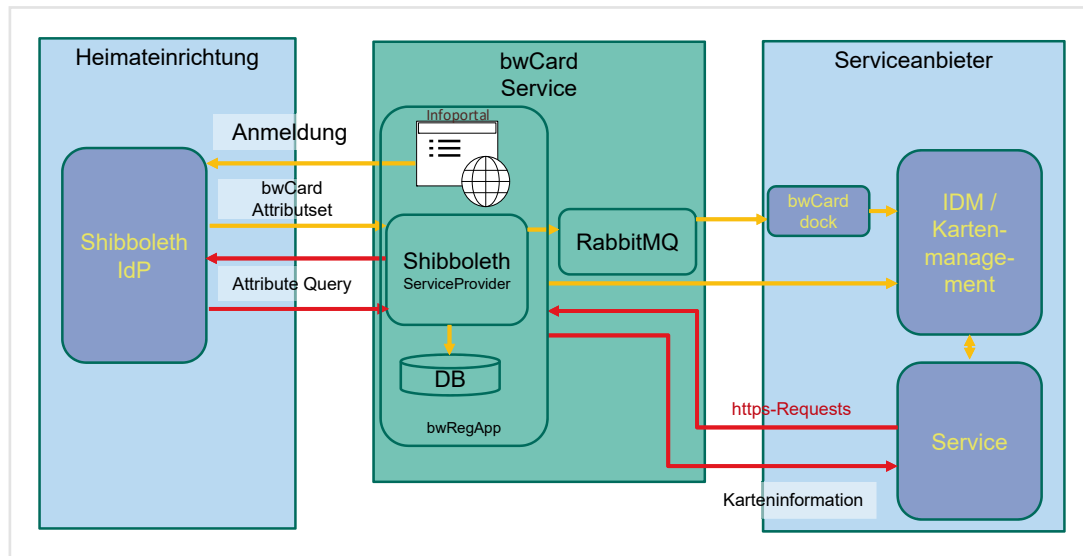


Abbildung 3: Schematischer Ablauf der bwCard-Statusabfrage

elektronische Kassen abgebucht werden. Diese Lösungen sind in den Hochschulen weit verbreitet.

Nachteile dieser Lösungen sind:

- Die hohen Lizenzkosten der Anbieter für eine elektronische Geldbörse.
- Die relativ hohen Kosten für die notwendige IT-Infrastruktur zum Auf- bzw. Abwerten. Damit scheidet ein solches Verfahren für Kassen mit einem sehr geringen Umsatz aus.
- Die Bezahlmöglichkeit ist häufig auf einen lokalen Anbieter innerhalb der Hochschulregion beschränkt (üblicherweise die jeweiligen Studierendenwerke). Für weitere Bezahlpartner muss das Umbuchen der Geldflüsse (Clearing) einzeln vereinbart werden, was teilweise hohen Aufwand verursacht.
- Die Beschränkung auf einen lokalen Anbieter verhindert, dass Bezahlmöglichkeiten auch in anderen Hochschulregionen genutzt werden können.

Ziel des Arbeitspaketes 2, das federführend von der Uni Konstanz vorangetrieben wird, ist unter anderem, ein Konzept zu entwerfen, wie in Zukunft mit dem

Thema der Bezahlung mit und ohne Bargeld an einer Universität umgegangen werden kann.

In einem ersten Schritt wurden Befragungen der Studierenden und der Universitäten durchgeführt. Als Fazit lässt sich festhalten, dass über 80% der Studierenden die bargeldlose Bezahlung akzeptieren, auch wenn sie die Barzahlung in der Mensa als wichtig ansehen. Hier ist aber der Durchdringungsgrad der bargeldlosen Bezahlung ein wichtiger Punkt. Bei Hochschulen, die bereits ausschließlich auf bargeldlose Bezahlung umgestellt haben, lag der Wunsch für Barzahlung bei unter 20%, bei Unis, die noch Barzahlung in der Mensa anbieten, lag der Wert teilweise bei über 70%. Bei bargeldloser Bezahlung mit Guthabekonto, wie es derzeit an den meisten Unis praktiziert wird, war der Wunsch nach flexibler Aufladung sehr ausgeprägt. Rund 90% der Befragten wünschen sich ihre Bezahlkarte online über das Handy aufladen zu können, während die direkte Zahlung per Handy-App nur zu einem geringen Anteil erwartet wurde.

Bei der Befragung der Universitäten wurde festgestellt, dass die rechtssichere Barzahlung, allein durch die erforderlichen Sicherheitsmaßnahmen, ein sehr kostenintensiver Prozess ist. Die Kosten

liegen in der Regel merklich über denen der bargeldlosen Zahlung, selbst wenn dort beachtliche Gebühren anfallen. Viel Wert legen die Universitäten auf eine digitalisierte Zuordnung der Gelder in ihr Finanzschema. Im Rahmen von bwCard wurde auf Grundlage dieser Umfrageergebnisse ein Leitfaden erarbeitet, mithilfe dessen Unis eine Voranalyse für ein Cashless Campus-Projekt durchführen können. Parallel dazu werden von der Uni Konstanz derzeit Pilotprojekte mit Herstellern von bargeldlosen Bezahlssystemen durchgeführt.

Evaluation einer E-ID

Daniel hat seinen Studenausweis verloren. Er meldet das bei seiner Hochschule, die den Ausweis sperrt. Gleichzeitig müssen auch die Landesbibliothek und die KIT-Bibliothek von der Sperre in Kenntnis gesetzt werden, damit zu dem Ärger, sich einen neuen Ausweis holen zu müssen, nicht noch jemand anderes Bücher mit der vermissten Karte ausleiht.

Erste Analysen im Projekt haben ergeben, dass die aktuelle Lösung im Karlsruher Verbund, welche einen direkten und gleichzeitig eingeschränkten Zugriff auf den Kartenstatus untereinander erlaubt, nicht für eine Landeslösung skaliert. Allerdings besteht mit bwIDM⁵ bereits eine landesweite Lösung zur Verwaltung von Identitätsmerkmalen. Es lag also nahe, sich an dieser Lösung zu orientieren. Genauere Untersuchungen der bwCard-Arbeitsgruppen haben einen Weg aufgezeigt, wie man die speziellen Anforderungen der bwCard komplett in bwIDM integrieren kann.

Dazu wird im Portal „bwRegApp“⁶ zur Registrierung von Landesdiensten die Möglichkeit geschaffen, sich dort auch für bwCard-Dienste zu registrieren. Will also in Zukunft eine Mitarbeiterin oder ein Mitarbeiter der Uni Konstanz einen Dienst an der Uni Stuttgart nutzen, registriert er bzw. sie sich im bwIDM-Portal für den entsprechenden Dienst. In diesem Fall wird der Dienst an der Uni Stuttgart davon informiert und erhält auch die Be-

rechtigung, den Kartenstatus der Person abzurufen.

Diese bleibt so lange erhalten, bis sie sich für den Dienst deregistriert oder aus anderen Gründen die Nutzungsberechtigung des Dienstes verliert.

Die Statusanfrage erfolgt ebenfalls über etablierte bwIDM-Mechanismen. Dazu wird der Personendatensatz, der beim Login an die Serviceeinrichtung übermittelt wird und bei jedem Login geprüft werden kann, um die Information einer gültigen bwCard erweitert. Selbstverständlich werden diese Informationen nur an die Dienste übermittelt, die diese Daten auch erfordern. Wird nun der Kartenstatus benötigt, erfolgt durch den Service-Anbieter eine sogenannte „Attribute-Query“⁷, die dann auch die bwCard-Nummer als Attribut enthält. Wird keine bwCard-Nummer mitgeliefert, gibt es keine gültige bwCard für diesen bwIDM-Account. Das bedeutet im Umkehrschluss, dass eine eventuell vorgelegte bwCard gesperrt ist.

Abbildung 3 zeigt den schematischen Ablauf einer Statusabfrage einer bwCard mit den bwIDM-Mechanismen. Gelbe Pfeile zeigen den Datenfluss zum Zeitpunkt der Registrierung für einen Dienst bei dem die bwCard genutzt werden soll. Rote Pfeile zeigen den Ablauf, wenn ein Dienst den Status der Karte ermitteln will. Für Endnutzer ist lediglich wichtig zu wissen, dass sie sich erst für einen Dienst registrieren müssen, bevor sie dort die bwCard einsetzen können.

Neue Kartentechnik oder Handytechnik

Dieses Arbeitspaket steht noch am Anfang. Es sollen Alternativen zur eingesetzten Technologie untersucht werden, wie etwa Übertragungsmechanismen für Karten und Handys. Weiterhin sollen neue und innovative Technologien geprüft werden, z.B. inwieweit für Verwaltungsaufgaben und Authentifizierung mit der bwCard oder anderen Datenträgern eine Blockchain genutzt werden kann. Ebenfalls soll im Rahmen von AP

3.2 untersucht werden, ob kontaktlose Karten aktuell oder in Zukunft der eIDAS-Verordnung⁸ entsprechen können. Diese Verordnung stellt die Basis des elektronischen Personalausweises dar, insbesondere im Zertifikataustausch und Unterschriftersatz.

Mit Abschluss des Projektes bwCard Ende 2021 werden die Grundlagen für eine landesweit einsetzbare Chipkarte geschaffen sein, die für viele Dienste die Basisinfrastruktur zur Verfügung stellt. Trotzdem wird es noch einige Zeit dauern, bis alle Möglichkeiten ausgeschöpft werden, für die das Projekt bwCard Konzepte und Lösungen erarbeitet hat.

Daniel wird seine Seminararbeit wahrscheinlich früher abgeschlossen haben.

State Project bwCard: Digital Identity Card

The bwCard project develops solutions and concepts for the state-wide use of existing chip cards, whether they are student ID cards or employee ID cards. It is run by the universities of the state of Baden-Württemberg and funded by the Ministry of Science, Research and the Arts. The project is divided into three main topics: The federal cooperation, the conception of pilot applications and the evaluation of an E-ID. The first priority area focuses on the organisation and implementation of different variants of cooperation. "Cashless Campus", an aspect of the second focal point, develops concepts and templates that can serve as a basis for a project to reduce cash on campus. In the third key area, the topics Evaluation of an E-ID and future technologies are considered.

¹ www.scc.kit.edu/downloads/oko/SCCnews_01_2018_web.pdf

² www.bibliothek.kit.edu/chipkarten-foederation.php

³ www.bwcard.de

⁴ www.landesrecht-bw.de/jportal/?quelle=jlink&query=VVBW-FM-20181220-SF&psml=bsbawueprod.psml&max=true&aiz=true

⁵ www.bwidm.de

⁶ bwidm.scc.kit.edu

⁷ wiki.shibboleth.net/confluence/display/SP3/QueryAttributeResolver

⁸ eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32014R0910

MathSEE Modeling Week – Studierende präsentieren Ergebnisse

Das KIT-Zentrum „MathSEE“, das die interdisziplinäre mathematische Forschung am KIT bündelt, richtete zum zweiten Mal eine Modeling Week für Studierende aus. Studierende sowie zwei Stipendiaten des Projekts Simulierte Welten präsentierten ihre Forschungsergebnisse aus den Themenbereichen Glasfasermaterialien, Mechatronische Systeme und Solarkraftwerke.

Kirsten Wohak

Studierende eines mathematisch-methodischen Faches und der „SEE-Fächer“ – Sciences, Engineering, and Economics – nahmen vom 16. bis 21. August 2020 an der MathSEE Modeling Week teil, die pandemiebedingt bis auf die Abschlussveranstaltung via Videokonferenz stattfand (Abbildung 1). Das KIT-Zentrum „MathSEE“ bündelt seit Oktober 2018 die interdisziplinäre mathematische Forschung am KIT und richtete zum zweiten Mal eine Modeling Week für Studierende aus. In der Abschlussveranstaltung am 21. August 2020 am Campus Süd stellten die Studierenden vor wenigen Zuschauern (Teilnehmer/innen und Organisatoren der Woche sowie Problemsteller/innen) ihre Ergebnisse vor (Abbildung 2). An dieser Veranstaltung nahmen auch zwei Förderstipendiaten des Projekts Simulierte Welten teil. Sie haben ein ganzes Schuljahr unter Betreuung von Sarah Schönbrodt (SCC) an der Optimierung eines solarthermischen Kraftwerks geforscht. Auch sie präsentierten ihre Erkenntnisse.

Am 16. August 2020 versammelten sich Studierende verschiedener Studiengänge in einem „virtuellen Klassenzimmer“, um an der MathSEE Modeling Week 2020 teilzunehmen. Dieses Format war für alle Teilnehmenden komplett neu. Sie würdigen sich die nächsten Tage virtuell mit spannenden Problemstellungen aus der realen Forschung des KIT-Zentrums MathSEE beschäftigen.

Nach einem ausgiebigen gemeinsamen Kennenlernabend begann die Arbeit in den Gruppen. Die betreuenden wissenschaftlichen Mitarbeitenden Pia Stammer (SCC) und Simon Rapp (Institut für



Abbildung 2: Abschlussveranstaltung mit der Verleihung von Zertifikaten an die Studierenden.

Produktentwicklung am KIT) führten in die Problemstellungen ein. Die Studierenden beschäftigten sich dieses Jahr mit zwei Themen:

- Die Faserorientierung in Glasfasermaterialien, welche einen entscheidenden Einfluss auf die mechanischen Eigenschaften von faserverstärkten Polymeren hat: Hierbei geht es darum, einen Algorithmus zu schreiben, der möglichst genau die Orientierungen von gegossenen Glasfasermaterialien bestimmt (Abbildung 3).
- Maschinenlesbare Beschreibung der Entstehung neuer mechatronischer Systeme, welche auf Basis bestehender Systeme entwickelt werden: Untersucht wurde, wie sich die Veränderung eines bekannten Systems auf das Potential und die Risiken des neuen Systems auswirkt.

Inhaltlich liegen diese beiden Problemstellungen weit auseinander, dennoch haben sie eine Gemeinsamkeit: Beide Probleme

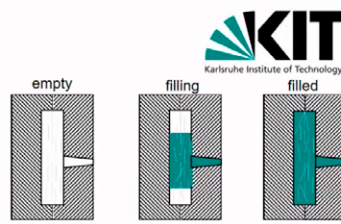
waren bisher ungelöst und stammen aus der aktuellen Forschung des KIT-Zentrums MathSEE. Die Aufgabe der Gruppen war es, Ansätze zu ihren Problemstellungen zu entwickeln. Dafür lernten sie die mathematische Modellierung kennen und wendeten sie ausgiebig an. So erhielten sie einen Einblick in den Forschungsalltag am KIT und mussten die dort auftretenden Probleme selbst lösen. In der Woche haben die Studierenden über die Kommunikations- und Kollaborationsplattformen Mattermost, MS Teams und CoCalc¹ zusammengearbeitet. Dadurch konnten sie nicht nur miteinander kommunizieren und über die Problemstellungen diskutieren, sondern auch gemeinsam am gleichen Code programmieren.

Am 21. August 2020 präsentierten die Studierenden und zwei Schüler des Förderstipendiums Simulierte Welten ihren Problemstellern und Problemstellerinnen ihre beeindruckenden Forschungsergebnisse. Trotz erschwelter Bedingungen

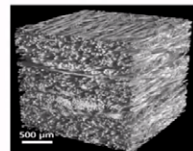
Motivation and Background

Why is fiber orientation important?

- Fiber orientation has a crucial impact on mechanical behavior of fiber re-inforced polymers (FRPs)
 - Especially fiber-fiber contacts inhibit orientation and may lead to fiber breakage
 - During injection molding, the FRP is injected in the cavity and the fibers are suspended in the polymer matrix
 - Depending on the flow field, the fibers orientate during mold filling
- Want to be able to estimate effect of different fiber orientations on number of fiber-fiber contact points



Schematic illustration of the mold filling process in injection molding



CT-Scan of glass fiber orientation in a FRP



Abbildung 3: Virtuelle Präsentation einer Problemstellung.

3

For internal use only

<https://www.kit.edu>

of Vehicle System Technology – Lightweight Technology

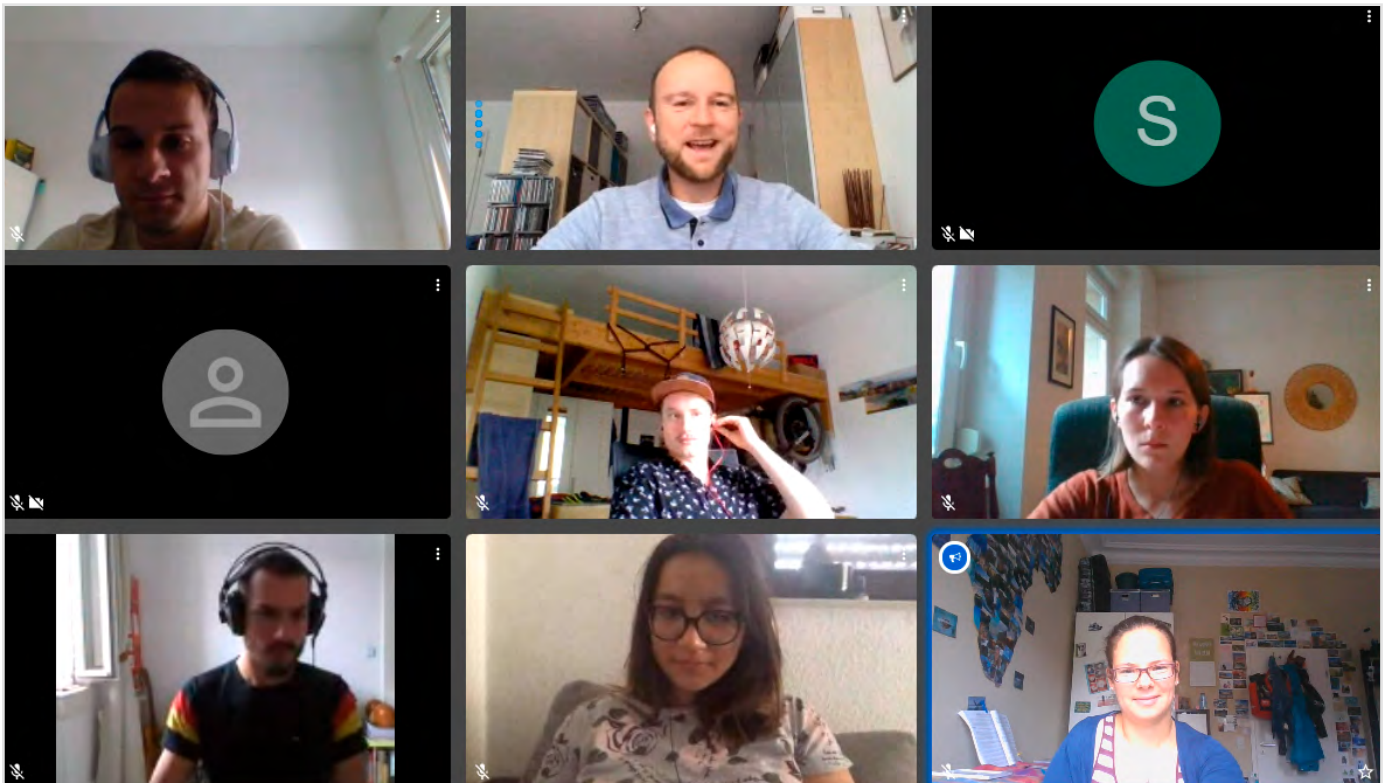


Abbildung 1: Diskussionsphase im virtuellen Kommunikationsportal.

durch die Corona-Auflagen war auch dieser Durchlauf erfolgreich.

Allen, die bei dieser interessanten und lehrreichen Woche mitgewirkt haben, indem sie entweder ein Problem gestellt, die Werbetrommel gerührt oder bei den Vorbereitungen geholfen haben, sei an dieser Stelle noch einmal ein riesengroßes Dankeschön ausgesprochen!

Students present their results from the MathSEE Modeling Week

Students of a mathematical-methodical subject and the „SEE subjects“ – Sciences, Engineering, and Economics – participated in the virtual MathSEE Modeling Week from August 16th – 21st 2020. Since October 2018, the KIT Center "MathSEE" has been bundling interdisciplinary mathematical research at KIT and organizing a Modeling Week for students for the second time. With the help of scientific assistants, the students dealt with fiber-fiber contact points in discontinuous glass fiber networks and machine-readable description of the development of new mechatronic systems. At the closing event on August 21st 2020 at Campus South, the problem posers followed the findings and results of the students with great interest.

KIT wird Zentrum für Nationales Hochleistungsrechnen

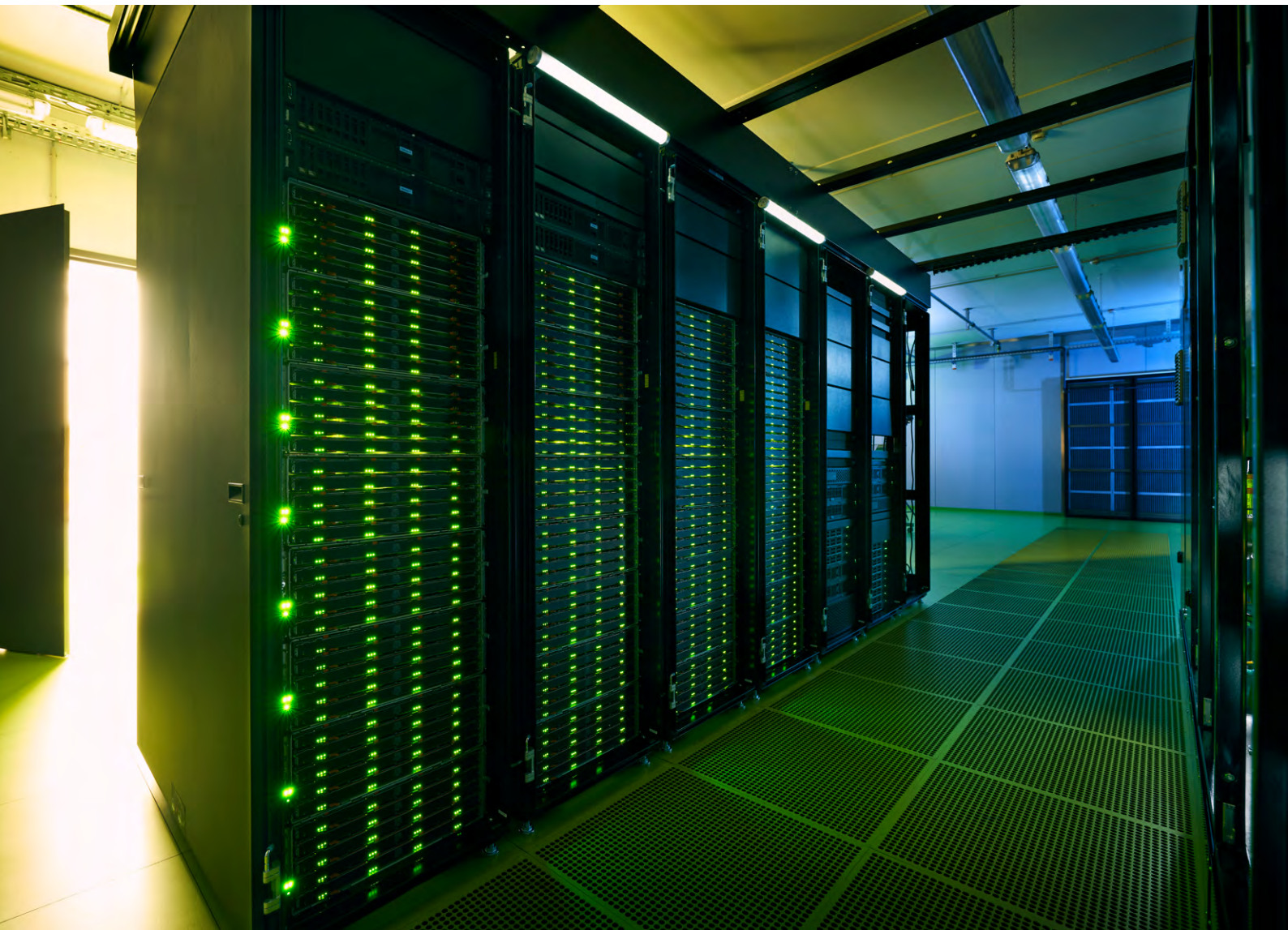
Das Karlsruher Institut für Technologie (KIT) wird Zentrum für Nationales Hochleistungsrechnen (NHR). Das hat die Gemeinsame Wissenschaftskonferenz (GWK) am 13.11.2020 beschlossen – das Gremium koordiniert die Wissenschaftsförderung von Bund und Ländern. Somit werden Wissenschaftlerinnen und Wissenschaftler zukünftig noch gewaltigere Hochleistungsrechner am KIT nutzen können. Die NHR-Allianz hat ein Jahresbudget von 62,5 Millionen Euro, ein hoher einstelliger Millionenbetrag geht jährlich ans KIT.

Felix Mescoli¹

„Baden-Württemberg ist in Sachen Supercomputing europaweit führend und international konkurrenzfähig. Ein Beleg dessen ist auch die heutige Aufnahme des KIT als Zentrum des Nationalen Hoch-

leistungsrechnens. Die Öffentlichkeit darf gespannt sein, zu welch bemerkenswerten Forschungsergebnissen die beeindruckende Rechenleistung und das geballte Methodenwissen am KIT der computer-

gestützten Spitzenforschung verhelfen werden“, so die baden-württembergische Ministerin für Wissenschaft, Forschung und Kunst, Theresia Bauer.



Die Bewältigung der Energiewende, die Entwicklung neuer Materialien, die Suche nach den kleinsten Bauteilen unseres Universums oder neuen Therapien für schwere Krankheiten wie Covid-19: Wollen Forscherinnen und Forscher natürliche oder technische Vorgänge in ihrer ganzen Komplexität verstehen, sind sie zunehmend auf Computer mit gewaltiger Rechenleistung angewiesen. „Sei es in der Energie-, Mobilitäts- und Umweltforschung oder Bereichen wie Cybersicherheit und Medizin: In der Wissenschaft fallen heute enorme Datenmengen an. Diese müssen wir nicht nur bewältigen, sondern aktiv nutzen, um die drängenden Herausforderungen unserer Gesellschaft zu meistern“, sagt der Präsident des KIT, Professor Holger Hanselka. „Ich freue mich sehr, dass das KIT mit seiner Erfahrung im Supercomputing nun Standort eines NHR-Zentrums wird.“

Neue Dimensionen bei der Rechenleistung

Der bestehende Hochleistungsrechner Karlsruhe (HoreKa) wird im Sommer 2021 voll einsatzbereit sein und eine Rechenleistung von rund 17 PetaFLOPS

erbringen – also etwa 17 Milliarden Rechenoperationen in der Sekunde. Das entspricht der Leistung von mehr als 150 000 Laptops. Mit der Aufnahme des KIT in den NHR-Verbund ist schon jetzt die Finanzierung der Nachfolgesysteme von HoreKa gesichert. „Deren Rechenleistung wird noch in ganz andere Dimensionen vorstoßen“, sagt Professor Martin Frank, Direktor des Steinbuch Centre for Computing (SCC) des KIT. Doch biete das KIT innerhalb der NHR-Allianz viel mehr als bloße Rechenleistung: „Wir wollen insbesondere junge Wissenschaftlerinnen und Wissenschaftler dabei unterstützen, schneller und effektiver zu Ergebnissen zu kommen“, so Frank. Um mit einem datenverarbeitenden Kraftpaket wie HoreKa umzugehen, brauche es besonderes Know-how. Hier seien intensive Beratung und engmaschige Betreuung der Nutzerinnen und Nutzer aus der Wissenschaft gefragt. „Unser Schwerpunkt wird also die umfassende Userunterstützung sein.“

Weitere Informationen zu HoreKa:

<https://www.scc.kit.edu/dienste/horeka.php>

KIT Will Be Center for National Supercomputing

Karlsruhe Institute of Technology (KIT) will become center for national supercomputing. This was decided by the Joint Science Conference on November 13, 2020. The Joint Science Conference coordinates funding of science by the Federation and the Federal States. In future, scientists will be able to use even more powerful supercomputers at KIT. The National Supercomputing Alliance has an annual budget of EUR 62.5 million. Of this, annual funding in the high single-digit million range will go to KIT.

IMPRESSUM

SCC news

Magazin des Steinbuch Centre for Computing

Herausgeber

Präsident Professor Dr.-Ing. Holger Hanselka
Karlsruher Institut für Technologie (KIT)
Kaiserstraße 12
76131 Karlsruhe

Anschrift

Steinbuch Centre for Computing (SCC)
Karlsruher Institut für Technologie (KIT)
Redaktion SCC-News
Zirkel 2
76131 Karlsruhe
oder:
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen
Fax: +49 721 608-24972

Redaktion

Achim Grindler (verantwortlich),
Karin Schäufele, Andreas Ley
E-Mail: redaktion@scc.kit.edu

Gestaltung, Satz und Layout

Heike Gerstner, Hella Grolmus
AServ – Crossmedia – Grafik
Karlsruher Institut für Technologie (KIT)
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen

Titelfoto

Markus Breig

Fotos

SCC, KIT, Markus Breig

Druck

Systemedia GmbH, 75449 Wurmberg

Erscheinungstermin dieser Ausgabe

Dezember 2020

www.scc.kit.edu/publikationen/scc-news

Der Nachdruck und die elektronische Weiterverwendung sowie die Weitergabe von Texten und Bildern, auch von Teilen, sind nur mit ausdrücklicher Genehmigung der Redaktion gestattet.



Karlsruher Institut für Technologie (KIT)
Steinbuch Centre for Computing (SCC)

ISSN: 1866-4954

www.scc.kit.edu
www.scc.kit.edu/twitter
contact@scc.kit.edu