

Steinbuch Centre for Computing

NEWS



Startschuss für Landesprojekte bwIDM und bwLSDF

Project kick-offs bwIDM and bwLSDF

Peer Energy Cloud – Energie aus der Wolke

Energy from the cloud

HadoOptimizer – Lösung komplexer Optimierungsprobleme

Solution of complex optimization problems



INHALT

- 4
Landesprojekt bwIDM –
Föderatives Identitätsmanagement
in Baden-Württemberg
- 6
Landesprojekt bwLSDF –
Datenspeicherung für die
baden-württembergischen Hochschulen
- 8
bwGRiD – Das Computing Grid für
Baden-Württemberg
- 10
Der neue C++-Standard
Wissenschaftliches Kolloquium des SCC
mit Michael Wong
- 11
9. Internationale GridKa Summer
School am SCC
- 12
Peer Energy Cloud bringt
Energie aus der Wolke
Bürgermarktplatz für den Handel mit
erneuerbaren Energien
- 13
Das SCC stellt sich vor
In dieser Ausgabe: Die Abteilung
Verteilte Systeme und Grid (VSG)
- 20
Peer Energy Cloud –
Teilvorhaben am KIT
- 24
HadoOptimizer – Lösung komplexer
Optimierungsprobleme
- 28
Datenerhebung zur
EU-Trennungsrechnung
SCC unterstützt FIMA
- 31
Security Service Challenge 4:
GridKa bestes Tier-1-Zentrum
- 31
Gebündelte Ressourcen für das
Höchstleistungsrechnen in BW
- 32
Computational Methods in Science
and Engineering
Proceedings of the Workshop SimLabs@KIT
- 33
Informationskonsistenz im föderativen
Identitätsmanagement
Modellierung und Mechanismen
- 34
Software-Cluster „Ausgewählter Ort
2011 im Land der Ideen“



6



8



12



32



36

35
Software-Cluster entwickelt innovative
Dienstleistungen im zukünftigen Internet

36
Neuer Zugang zum Software Shop
mit Shibboleth

38
TIS – Testing IT Systems and
Support-Team
Unterstützung bei der Entwicklung neuer
IT-Systeme am SCC

EDITORIAL

Liebe Leserinnen und Leser,

das SCC befasst sich neben seinen klassischen Aufgaben als wissenschaftliches IT-Service-Zentrum nicht nur mit Forschung und Entwicklung auf dem Gebiet der angewandten Informationstechnologien, sondern ist auch Partner in zahlreichen nationalen und internationalen IT-Föderationen. Für drei dieser Föderationen, die sich auf Landesebene bewegen und deren Projektleitung das SCC wahrnimmt, fiel in den letzten Monaten der Startschuss: „bwIDM“, „bwLSDF“ und „bwGRiD – ergänzende Maßnahmen“. Ziel des Projekts „bwIDM“, das von allen neun Landesuniversitäten getragen und unterstützt wird, ist die Ermöglichung eines einfachen Zugriffs auf über das Land verteilt angebotene Ressourcen und Dienste aus einem lokalen Kontext heraus. Das Landesvorhaben „bwLSDF“ mit den Universitäten Heidelberg, Mannheim, Stuttgart und Ulm als Kernprojektspartner sieht eine Erweiterung der am SCC angesiedelten Large Scale Data Facility (LSDF) für die baden-württembergischen Hochschulen vor. Das Landesprojekt „bwGRiD“, ein Verbund aus neun Landeshochschulrechenzentren, stellt mit ca. 323.000 täglich angebotenen CPU-Stunden eine einzigartige Infrastruktur im Vergleich zu anderen Bundesländern bereit. An allen Standorten stehen den Nutzern nicht nur enorme Rechenressourcen zur Verfügung, sondern auch eine einheitliche Umgebung, die einen problemlosen Standortwechsel ermöglicht. Mit dem Projekt „bwGRiD – ergänzende Maßnahmen“ sollen nun die Weichen für die Zukunft gestellt werden: Die standortunabhängige Nutzung soll weiter verbessert, der Zugriff erleichtert und der Nutzerkreis sowie die zugehörige Betreuung ausgedehnt werden.

Ebenfalls neu ist das wissenschaftliche Kolloquium des SCC. Es startete Anfang September mit Michael Wong, der in einem voll besetzten Hörsaal zwei Stunden über den neuen C++-Standard referierte. Das Kolloquium fungierte auch als Abendveranstaltung der 9. Internationalen GridKa Summer School des SCC, eine der weltweit führenden Schulen für Grid und Cloud Computing. Auch in diesem Jahr wurde den knapp 100 Teilnehmern aus 11 Nationen in hochkarätigen Vorträgen, Workshops und praktischen Übungen der neueste Stand der Technik in den genannten Bereichen vermittelt.

Und noch ein Projekt mit KIT- und SCC-Beteiligung startete Anfang September: Das Partnerkonsortium „Peer Energy Cloud“ gehört zu den zwölf Gewinnern des Technologiewettbewerbs „Trusted Cloud“ des Bundeswirtschaftsministeriums und stellt die bahnbrechende Entwicklung von so genannten Cloud Enabled Smart Energy Micro Grids in Aussicht.

Viel Vergnügen bei der Lektüre wünschen Ihnen
Hannes Hartenstein, Bernhard Neumair und Achim Streit



Prof. Dr. Hannes Hartenstein
Foto: Privat



Prof. Dr. Bernhard Neumair
Foto: Privat



Prof. Dr. Achim Streit
Foto: Privat

IMPRESSUM

Dezember 2011

Herausgegeben im Auftrag des Direktoriums des Steinbuch Centre for Computing (SCC) von der Stabsstelle Öffentlichkeitsarbeit und Kommunikation

Anschrift: Steinbuch Centre for Computing (SCC)

Redaktion SCC-News

Zirkel 2

76131 Karlsruhe bzw.

Hermann-von-Helmholtz-Platz 1

76344 Eggenstein-Leopoldshafen

Fax: 0721/32550

<http://www.scc.kit.edu/publikationen/scc-news>

Redaktion:

Ursula Scheller (verantwortlich)

Telefon: 0721/608-44865

E-Mail: ursula.scheller@kit.edu

Layout und Bildredaktion: John Atkinson

Redaktionell bearbeitete Texte werden mit (red) gekennzeichnet. Nachdruck und elektronische Weiterverwendung von Texten und Bildern nur mit ausdrücklicher Genehmigung der Redaktion.

Landesprojekt bwIDM – Föderatives Identitätsmanagement in Baden-Württemberg



Kick-off-Veranstaltung zum Landesprojekt bwIDM Ende September am KIT mit den Teilnehmern aller neun Landesuniversitäten. Verantwortung und Leitung des Projekts liegen bei Prof. Dr. Hannes Hartenstein und Dr. Martin Nussbaumer vom SCC (siebter und neunter von links). Foto: Christina Petri

Die Hochschulen des Landes Baden-Württemberg stellen ihren Nutzern eine große und stetig steigende Anzahl von IT-Diensten und -Ressourcen zur Verfügung, die maßgeblich zur Qualität und Attraktivität der Hochschulen beitragen. Da diese Dienste oftmals auf die einfache Verwendung für die Mitglieder der eigenen Hochschule optimiert sind, ist ihre Verwendung für Standortfremde häufig nur mit größerem Aufwand möglich: Nutzer müssen dazu in der Regel für jeden Dienst und jede Ressource eigene Zugänge beantragen. Dabei handelt es sich meistens um ein zeitintensives und fehleranfälliges Verfahren, da die VergabeprozEDUREN von den lokalen Gegebenheiten abhängen. Ziel des Projekts bwIDM ist es, den Betreibern und Nutzern von IT-Diensten der Hochschulen des Landes Baden-Württemberg die Möglichkeit eines einheitlichen, einfachen und sicheren Dienstzugangs zur Verfügung zu stellen.

„Die mit dem Landeskonzept bwIDM verfolgte Vision ist die Ermöglichung eines nahtlosen Zugriffs auf über das Land verteilt angebotene Ressourcen und Dienste aus einem lokalen Kontext heraus“, so Prof. Dr. Hannes Hartenstein, Geschäftsführender Direktor des SCC und Projektverantwortlicher. „Nutzer sollen mit ihrer gewohnten Kennung aus ihrer gewohnten Umgebung auf die Dienste und Ressourcen anderer Hochschulen der Föderation zugreifen können, als wären diese Ressourcen physisch vor Ort.“

Dabei kommt den standortbezogenen Identitätsmanagement-Systemen eine besondere Rolle hinsichtlich der Bereitstellung landesweit einheitlicher Autorisierungsmerkmale zu. „In Zusammenarbeit mit den anzuschließenden Landesdiensten müssen daher Nutzungsrichtlinien erarbeitet werden, die – bereitgestellt durch die lokalen IDM Systeme – die Autorisierung am Landesdienst ermöglichen“, erklärt Projektleiter Dr. Martin Nussbaumer vom SCC. Dabei bilden die Landesuniversitäten Ulm, Freiburg, Konstanz und das KIT ein Kernteam, das sich mit den im Projekt anstehenden Aufgaben und deren Koordination befasst. Getragen und unterstützt wird das Vorhaben von allen neun Landesuniversitäten.

Struktur des Projekts

Die im Projekt bwIDM notwendigen Arbeiten werden im Wesentlichen den drei Arbeitsbereichen „Bestandsaufnahme und Prozessanalyse“, „Rahmenkonzept“ und „Fachkonzepte“ zugeordnet (s. Abbildung 1).

Eine Besonderheit bilden die bwIDM-Commitments, die Befürwortung und Sicherung der Nachhaltigkeit der Projektergebnisse ausdrücken. Im ersten Commitment werden die Befürwortung des Projekts und das Bereitstellen von Unterstützung durch die Partner-Universitäten artikuliert. Dadurch soll zum Ausdruck gebracht werden, dass das landesweite Vorhaben von allen teilnehmenden Landesuniversitäten gemeinsam getragen wird. Dieses Commitment ist in der ersten Phase des Projekts zu erbringen. Durch das zweite Commitment, das in der Phase Q2/2012 abgegeben wird, bekräftigen die teilnehmenden Universitäten ihre Bereitschaft, auf Basis der Ergebnisse aus den Arbeitspaketen eins und zwei, die notwendigen und zukünftigen Anpassungen an den lokalen Identitätsmanagement-Systemen vorzunehmen und nachzuhalten.

Der Arbeitsbereich „Bestandsaufnahme und Prozessanalyse“ dient der Analyse und Auswahl geeigneter föderativer Technologien, die für den Aufbau eines föderativen Identitätsmanagement für das Land Baden-Württemberg eingesetzt werden sollen (AP Föderative Technologien). Zusätzlich werden bestehende Dienste erfasst, die beim Aufbau eines föderativen Identitätsmanagements a priori betrachtet werden sollen (AP Status Quo). Bei diesen Landesdiensten handelt es sich um das Hochleistungsrechnen, bwGrid, bwLSDF (Large Scale Data Facility) und das landesweite integrierte Bibliothekssystem (IBS). Außerdem soll die Interoperabilität des bwIDM mit bestehenden Strukturen, sei es national oder auch internatio-

nal, berücksichtigt sowie Zusammenarbeitspotenziale ermittelt werden (AP Interoperabilität mit bestehenden Föderationen). Schließlich wird ein Integrationsmodell entwickelt, mit dessen Hilfe die Aufnahme zukünftiger Landesdienste in die bwIDM-Föderation erreicht wird (AP Erweiterungsfähigkeit des bwIDM).

Der Arbeitsbereich „Rahmenkonzept“ widmet sich den für Aufbau und Betrieb notwendigen technischen und organisatorischen Strukturen (AP Governance-Strukturen). Des Weiteren werden Anforderungen zur Gewährleistung des Datenschutzes und der IT-Sicherheit betrachtet und Sicherheitsrichtlinien für die Föderation erarbeitet (AP Datenschutz und Sicherheit). Darüber hinaus müssen auf Basis der für das bwIDM gewählten föderativen Technologie entsprechende Schnittstellen und Datenmodelle für die später anzubindenden Fachkonzepte erarbeitet werden (AP Schnittstellen und Datenmodell). Die universitätslokalen Identitätsmanagement-Systeme werden auf Basis der gewählten föderativen Technologie in die bwIDM-Föderation integriert und notwendige Anpassungen hinsichtlich der verabredeten Richtlinien erbracht (AP Einbezug und Integration lokaler IDMs in die Föderation).

Der Arbeitsbereich „Fachkonzepte“ umfasst das Erstellen der Anbindung von a priori identifizierten Diensten in die bwIDM-Föderation. Dabei werden zunächst dienstübergreifende Anforderungen an Fachkonzepte für bwIDM aufgestellt (AP Fachkonzepte). Darauf aufbauend wird dann für die identifizierten Landesdienste ein dediziertes Fachkonzept entworfen, implementiert und in Betrieb gesetzt.

(red)

“bwIDM” – federated identity management in Baden-Wuerttemberg

The universities of the State of Baden-Wuerttemberg provide their users with a large and steadily increasing number of IT services and resources. These IT services contribute significantly to the quality and attractiveness of the universities. While they are often optimized for ease of use for members of the local site, the distributed use for non-local users is many times only possible with great efforts: users have to apply for a dedicated account at each service site, which is frequently time consuming and error-prone because of the assignment procedures that depend on local prevailing conditions. The goal of the bwIDM project is to provide operators and users of the universities' IT services in the State of Baden-Wuerttemberg with a uniform, simple and secure service access.

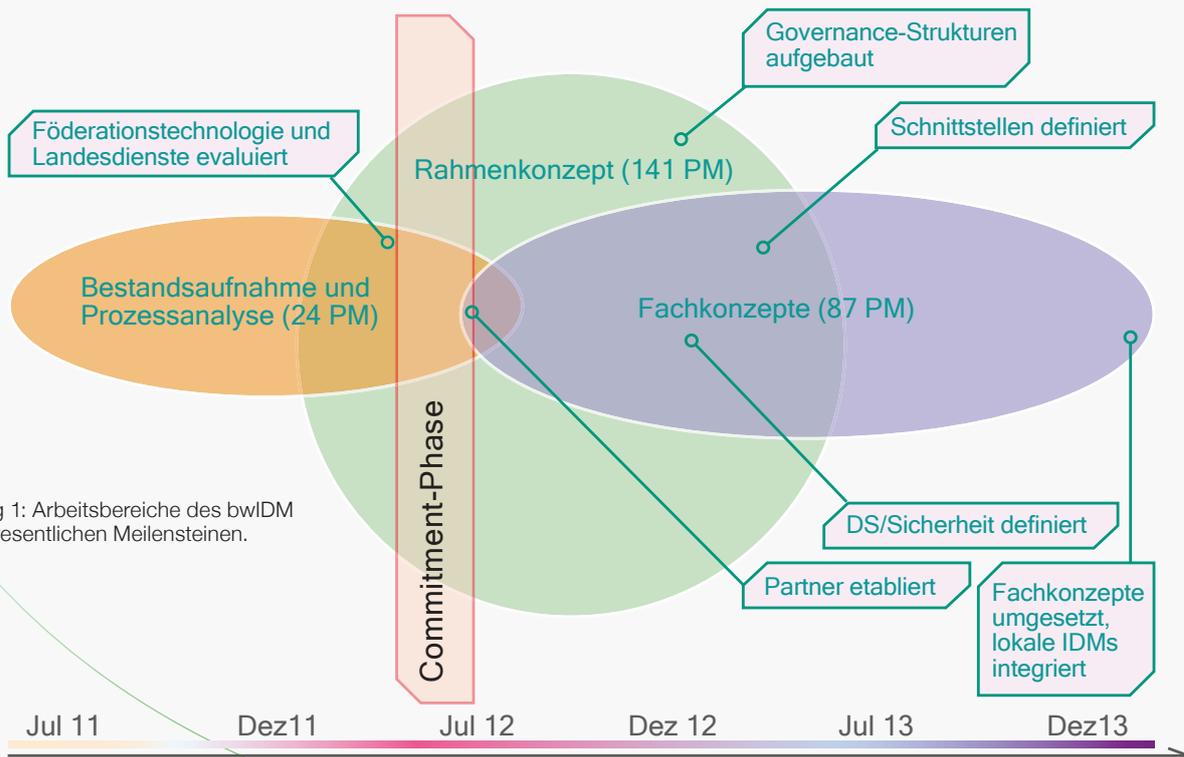


Abbildung 1: Arbeitsbereiche des bwIDM mit den wesentlichen Meilensteinen.

Landesprojekt bwLSDF – Datenspeicherung für die baden-württembergischen Hochschulen

Am SCC ist zum 1. Oktober 2011 das Landesprojekt bwLSDF gestartet. Unter der Leitung von Prof. Dr. Bernhard Neumair und Dr. Sebastian Rieger wird im Projekt mit Unterstützung des Ministeriums für Wissenschaft, Forschung und Kunst Baden-Württemberg eine Erweiterung der Large Scale Data Facility (LSDF) zur Datenspeicherung für die baden-württembergischen Hochschulen entwickelt und exemplarisch umgesetzt. Kernprojektpartner sind neben dem SCC die Universitäten Heidelberg, Mannheim, Stuttgart und Ulm.

Das Projekt baut auf existierenden Konzepten für Landesdatenspeicher im Rahmen der Large Scale Data Facility (LSDF) auf. Gegenüber den bestehenden Konzepten enthält das bwLSDF-Projekt eine Erweiterungsplanung in zwei Bereichen, die basierend auf den Anforderungen der baden-württembergischen Universitäten in den gemeinsamen Projektantrag eingebracht wurden. Die Erweiterungen (s. Abbildung 1) gliedern sich in die landesweite Bereitstellung eines Überlaufspeichers sowie die standortübergreifende Realisierung von virtuellen Fileservern.

Hochschulen), ergänzend zur bestehenden Funktionalität als Überlaufspeicher für wissenschaftliche Massendaten. Hochschulen könnten die virtuellen Fileserver verwenden und wahlweise auf den Betrieb eigener Fileserver verzichten, sofern die spezifischen Anforderungen für die optimale Integration in die IT-Infrastruktur des Standorts umgesetzt werden können. Damit würde das bisherige Datenspeicherkonzept auf den „täglichen Datenspeicherbedarf“ von Mitarbeitern und Studierenden der Hochschulen ausgedehnt. Im Rahmen des Projekts werden mögliche Verfahren hierfür getestet und in Pilotinstallationen prototypisch hinsichtlich ihrer Netzwerkanforderungen erprobt.

Beide Ansätze bilden Erweiterungen des Zugriffs auf die LSDF bzw. eine neue Art von verteilten Datenservices der baden-württembergischen Universitäten für die Hochschulen, Forschungsprojekte, Archive und Bibliotheken im Land. Sie ermöglichen eine für die Benutzer transparente und ubiquitäre Bereitstellung von Speicher-Ressourcen über das bestehende Landeshochschulnetz BelWü. Im Rahmen des Vorhabens soll daher die Nutzung der bestehenden Infrastruktur der LSDF auf weitere Hochschulstandorte ausgedehnt und so als Landesdienst weiter ausgebaut werden.

Höhere Betriebseffizienz und Entlastung der Hochschulen

Die im bwLSDF-Projekt geplanten Erweiterungen bezüglich des landesweiten Zugriffs auf die LSDF-Kapazitäten ermöglichen darüber hinaus durch die Nutzung von Synergien (zum Beispiel durch einheitliches Management und Bündelung von Know-how) eine nachhaltige Steigerung der Betriebseffizienz wissenschaftlicher Rechen- oder Informationszentren im Bereich des Datenmanagements.

Zielsetzung ist ein aus Sicht der Benutzer der Hochschulen ortsunabhängiger und transparenter Zugriff auf die von ihnen benötigten Daten. Ausgehend von den am Projekt beteiligten Kernpartnern erhält das Vorhaben unter Koordination des KIT einen Modellcharakter für die Nutzung der LSDF durch weitere Hochschulen des Landes. Die Realisierung der Erweiterungen im Rahmen eines Verbundprojekts der Kernpartner ermöglicht eine Vereinheitlichung des Managements und die Nutzung von Skaleneffekten bei Beschaffungen. Das übergreifende Betriebskonzept vereinfacht zudem die Prozesse und entlastet insbesondere Hochschulen, die keine eigenen Speicherinfrastrukturen haben, von administrativen, operativen und technischen Tätigkeiten für die Bereitstellung von lokalen Datenspeichern. Neben der Evaluierung geeigneter Lösungen und Zugriffsverfahren für die genannten Teilprojekte, bilden beispielsweise geeignete Abrechnungs- und

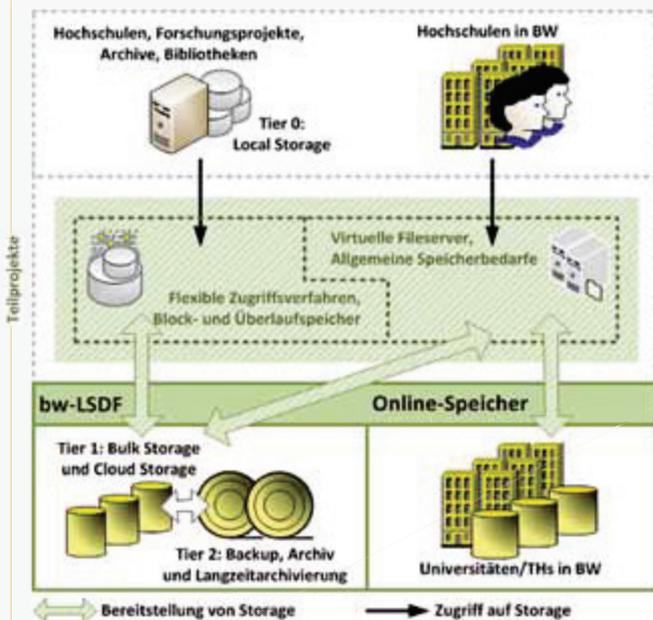


Abbildung 1: Teilprojekte des bwLSDF-Projekts.

Ubiquitäre Bereitstellung von Speicher-Ressourcen

Für die Bereitstellung des Überlaufspeichers ist dessen Erweiterung um flexible Zugriffsverfahren erforderlich. Dies umfasst die Evaluierung von technischen Speicherzugriffsverfahren, die die Einsatzmöglichkeiten der LSDF als Überlaufspeicher für die Universitäten verbessern, die Anbindung an nutzende Systeme flexibilisieren, die Übertragungsleistung optimieren sowie die Verfügbarkeit und Stabilität erhöhen. Hierbei wird auch evaluiert, ob und wie neben File-basierten zusätzlich Block-basierte Verfahren wie beispielsweise FC, FCoE und iSCSI für den Zugriff auf den Tier-1-Speicher verwendet werden können und welche netzseitigen Voraussetzungen hierfür gegebenenfalls zu schaffen sind.

Darüber hinaus wird im zweiten Teilprojekt die standortübergreifende Bereitstellung von virtuellen Fileservern evaluiert und pilotiert. Dies geschieht auf der Basis der LSDF des KIT und weiterer Anbieter (Universitäten, Technische

“bwLSDF” – Data storage for the universities in Baden-Wuerttemberg

Authentifizierungsverfahren, Supportstrukturen, die notwendige hohe Verfügbarkeit und Bandbreite der angebotenen Dienste sowie die Gewährleistung der erforderlichen IT- bzw. Datensicherheit und des Datenschutzes übergeordnete Anforderungen an das Gesamtprojekt. Die im vorliegenden Projekt erarbeiteten Lösungen für die Adressierung dieser Herausforderungen können auch für andere Landesprojekte Modellcharakter haben. Durch die enge Abstimmung mit bestehenden und angrenzenden Landesprojekten während des Projektzeitraums (wie beispielsweise bwGRiD und bwIDM) wird ein entsprechender Austausch gefördert und die kooperative Bereitstellung von Landesdiensten hochschulübergreifend gestärkt.

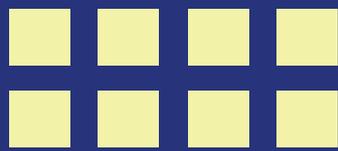
Gemeinsam mit den Kernprojektpartnern sowie Teilnehmern weiterer Universitäten des Landes Baden-Württemberg und des Landeshochschulnetzes BelWü fand Mitte Oktober am SCC ein Kick-off Meeting zum Projekt statt. Auf der Veranstaltung wurde die Ausgangsbasis des Vorhabens anhand existierender Lösungen, Herausforderungen und Netzwerkanforderungen präsentiert und gemeinsam diskutiert. Im nächsten Schritt werden die individuellen Anforderungsprofile der Projektpartner im Rahmen einer Bedarfsanalyse ermittelt, und die Arbeit in den Arbeitspaketen zur Evaluierung geeigneter Zugriffsverfahren begonnen.

Dr. Sebastian Rieger

In order to establish a distributed storage infrastructure for research and teaching in the State of Baden-Wuerttemberg, SCC has started the project bwLSDF (Large Scale Data Facility for Baden-Wuerttemberg) with four core project partners of the universities of Heidelberg, Mannheim, Stuttgart and Ulm. The goal of the project being funded by the Ministry of Science, Research and the Arts Baden-Wuerttemberg is to provide academic users in the state with a ubiquitous and transparent access to storage resources. In this context the project focuses on two solutions to be evaluated and developed: a swap storage allowing the transfer of especially large or less frequently used data to a remote data center in Baden-Wuerttemberg and the provisioning of virtual file servers in order to offer decentralized file access across different sites. Both solutions require the development of new storage access methods.



Auf dem Kick-off Meeting zum bwLSDF-Projekt Mitte Oktober am SCC diskutierten die Projektpartner die Ausgangsbasis des Vorhabens.
Foto: Rainer Kupsch



bwGRiD

bwGRiD – das Computing Grid für Baden-Württemberg

bwGRiD – the Computing Grid for Baden-Wuerttemberg

Within the scope of the bwGRiD project in recent years an IT infrastructure has been built in the State of Baden-Wuerttemberg which is unique concerning the way of cooperation of the involved universities' computing centers and in comparison with other federal states in Germany. bwGRiD offers its users a daily computing power of about 323 thousand CPU hours in a standardized environment. The computing resources are open for all members of the state's universities, once they have registered for the so called bwGRiD-VO. The strong infrastructure formed and extended by bwGRiD and its co-project "Ergänzende Maßnahmen" (Complimentary Measures), which has started in July 2011, is planned to be continuously operated in a sustainable way.



Abbildung 1: bwGRiD-Standorte (rot).

In den letzten Jahren wurde in Baden-Württemberg im Rahmen des bwGRiD-Projektes eine IT-Infrastruktur aufgebaut, die sowohl im Hinblick auf die Zusammenarbeit der beteiligten Hochschulrechenzentren als auch im Vergleich zu anderen deutschen Bundesländern einzigartig ist. An allen acht bwGRiD-Standorten finden die Nutzer seit Mitte 2010 eine Einheitsumgebung vor, so dass sie mit ihren Anwendungen problemlos von einem Standort zum anderen wechseln können und reichlich Rechenressourcen flexibel zur Verfügung haben.

Etwa 323.000 CPU-Stunden bietet das bwGRiD insgesamt seinen Nutzern täglich an. Zusätzlich zu den sieben ursprünglich beschafften Clustern an den Standorten Freiburg, Heidelberg, Karlsruhe, Mannheim, Stuttgart, Tübingen und Ulm ging Mitte Mai 2010 ein Cluster mit 180 Knoten an der Hochschule Esslingen in Betrieb. Damit besteht der bwGRiD-Verbund aus neun (die Ressourcen für Konstanz werden in Ulm betrieben) Hochschulrechenzentren mit insgesamt ca. 1680 Knoten á 8 Cores, das heißt 13.450 Cores. Ergänzt wird diese Rechenkapazität durch insgesamt über 700 TB nutzbare Speicherkapazität in den lokalen Cluster File-Systemen und im zentralen Speichersystem in Karlsruhe.

Vielfältige Synergieeffekte

Die erwähnte Einheitsumgebung kam zustande, nachdem Ende März 2010 eine Initiative zur Vereinheitlichung der Benutzerumgebung auf den Clustern im bwGRiD gestartet wurde, die im September erfolgreich beendet werden konnte. Damit wurden zum einen die Hürden zur Ausführung von benutzergeneriertem Programmcode an unterschiedlichen Standorten noch einmal deutlich reduziert. Zum anderen entstehen bei der Administration der Systeme vielfältige Synergieeffekte, indem Erfahrungen, Konfigurationen und ganze Softwarepakete zwischen den Standorten ausgetauscht werden, oder erfahrene Kollegen bei komplexen Änderungen direkt vor Ort Hilfe leisten. Durch Profilbildung bei der Auswahl der unterstützten Software entstand zudem eine landesweite Kooperation bei der Nutzerunterstützung.

bwGRiD größter Nutzer im D-Grid

Das landesweite bwGRiD-Projekt reiht sich ein in das bundesweite, übergeordnete Projekt D-Grid. Für das D-Grid-Projekt werden halbjährlich Nutzungsstatistiken erstellt, aus denen sich die Aufteilung der genutzten Ressourcen auf die einzelnen D-Grid-VOs und lokale Benutzer entnehmen lässt. Die Zahlen zeigen, dass die VO bwGRiD mit einigen Hundert Mitgliedern bereits heute der größte Ressourcennutzer (im Sinne verbrauchter CPU-Stunden) im D-Grid-Verbund ist. Die bwGRiD-Systeme sind in die Monitoring-Dienste des D-Grid eingebunden und auf diese Weise auch deutschlandweit sichtbar.

Webbasierter Zugang zu Applikationen

Ein wichtiges Teilprojekt des bwGRiD ist das so genannte Portal-Projekt. Im Rahmen dieses Vorhabens wird eine umfassende Plattform zur Bereitstellung eines einfachen webbasierten Zugangs zu Applikationen am bwGRiD geschaffen. An insgesamt sechs Standorten in Baden-Württemberg werden auf Basis des Portals Web-Schnittstellen zu Applikationen entwickelt. Jenseits der Basisfunktionalität entstehen zurzeit mehrere vielversprechende Applikations-Portlets in verschiedenen Wissenschaftsbereichen. Hier ist für das KIT aus den Ingenieurwissenschaften die CAE-Software Fluent zu nennen, für die eine Portallösung entwickelt wird.

Die Koordination der Arbeiten im bwGRiD erfolgt hauptsächlich durch eine alle zwei Wochen stattfindende Videokonferenz. Darüber hinaus werden Ankündigungen von allgemeinem Interesse, wie zum Beispiel neu angebotene Softwareversionen, über einen entsprechend großen E-Mail-Verteiler kommuniziert.

Projektleitung am SCC

Seit Januar 2010 wurde die Projektleitung des bwGRiD von Stuttgart ans SCC nach Karlsruhe übertragen. Die Projektleitung des bwGRiD ist auch auf den dreimal jährlich stattfindenden Sitzungen des D-Grid-Beirats vertreten.

Durch einen zum 1. Juli 2011 zusätzlich bis Mitte 2013 bewilligten Projektantrag mit dem Arbeitstitel „bwGRiD - ergänzende Maßnahmen“ sollen die Weichen für die Zukunft des bwGRiD gestellt werden: Die standortunabhängige Nutzung soll weiter verbessert, der Zugriff weiter erleichtert und der Kreis der Nutzer sowie die zugehörige Betreuung ausgeweitet werden.

Für die Zukunft ist davon auszugehen, dass der Betrieb von bwGRiD-Ressourcen an einigen wenigen Standorten fortentwickelt wird. Die Entwicklung von einfach zu nutzenden Zugangsmechanismen, Portalen, Softwarelösungen und anderen Diensten (wie zum Beispiel dem bwGRiD-weiten 1st-Line-Support, aber auch Monitoring u.a.) sowie die Gewinnung, Betreuung und Förderung von Nutzern und Communities wird an Bedeutung noch weiter zunehmen und geographisch verteilt erfolgen.

Nachhaltiger Betrieb der Infrastruktur

Die durch das bwGRiD und die ergänzenden Maßnahmen gebildete, erweiterte und gestärkte Infrastruktur soll nachhaltig betrieben werden. Entwickelte Prototypen, etwa bei der Kopplung von lokalem Helpdesk mit dem NGI-DE-Helpdesk, sollen produktionsreif gemacht werden und in ganz Baden-Württemberg zum Einsatz kommen. Es sollen Nutzungsricht-

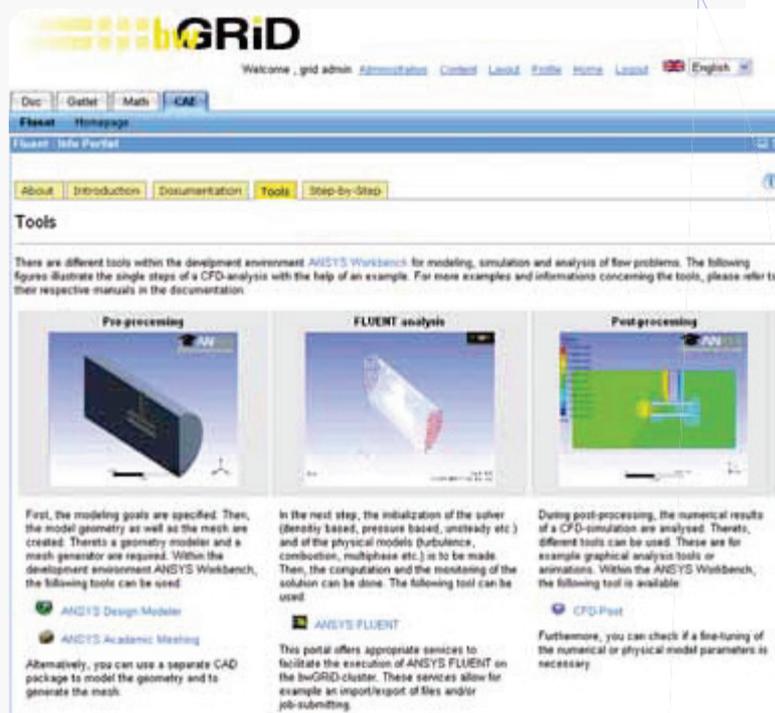


Abbildung 2: Das am SCC entwickelte Applikations-Portlet Fluent.

linien erarbeitet werden, etwa um die Anmeldeprozeduren beim Zugang zum bwGRiD zu vereinfachen, und diese Richtlinien sollen dauerhaft über den Förderungszeitraum hinaus Gültigkeit haben. Der angestrebte übergreifende Betrieb der Cluster soll dauerhaft möglich sein. Materialien zur Verbreitung, Förderung, Schulung und der Zugang für Nutzer werden auch nach Projektende verfügbar bleiben. Die Erfahrungen bei der Integration zusätzlicher wissenschaftlicher Nutzer und Fachgruppen sollen in Form von Berichten festgehalten werden und zur Verfügung stehen. Das Anwendersoftware-Kompetenz-Netzwerk soll ebenfalls über das Projekt hinaus Bestand haben. Auch das bwGRiD-Portal soll, nachdem es einmal im Produktionsmodus ist, zu einer dauerhaften Einrichtung für die Wissenschaftler in Baden-Württemberg werden. Um diese nachhaltigen Vorhaben zu unterstützen, wird eine Verlängerung des derzeitigen bwGRiD-Projekts oder ein entsprechendes neues Folgeprojekt angestrebt.

Dr. Sven Hermann

The new C++ standard — scientific colloquium with Michael Wong

As inaugural talk of its new scientific colloquium and on the occasion of the release of the new C++ standard C++11, SCC invited on 6th September Michael Wong - representative for IBM and Canada in the OpenMP- and C++-Standards committees.

The new C++ standard which has just been released is often referred to as C++0x. As a language, C++ plays a particularly important role in high performance computing and systems programming. Being close to the physical hardware, C++ can provide unrivalled performance, but has also to abide by stricter rules than are common for semi-interpreted languages. In the past, this has led to a very conservative development style – the last official update to the C++ standard dates back to 2003.

With the new standard, many of the more daunting limitations will be removed. Users will benefit from increased performance as well as a wide array of new capabilities, both on the core language side of C++0x and on the side of the standard library.

The presentation introduced the new standard using specific code examples, including a discussion of limitations that still remain. A particular focus lay on C++0x's new features related to concurrency. In addition possible migration paths for existing projects were discussed and an insight into the standards process given.

The interest in the topic proved to be so big that the lecture hall was filled to the last seat, and additional chairs had to be organized. Michael Wong thus agreed on short notice to give a second presentation in the context of SCC's GridKa summer school, with additional details on the topic of concurrent programming and memory models.

SCC will build on this successful start and continue to invite internationally renowned experts who will inform the interested audience on current topics in information technology.

Der neue C++-Standard

Wissenschaftliches Kolloquium des SCC mit Michael Wong

Mit einem furiosen Start läutete das SCC am 6. September 2011 sein neu eingerichtetes wissenschaftliches Kolloquium ein. Den Anfang machte Michael Wong als Vertreter von IBM und Kanada in den C++-Standard- und OpenMP-Komitees.

Vor dem mit mehr als 120 Teilnehmern voll besetzten Hörsaal informierte Michael Wong die Zuhörer über die Neuerungen des gerade verabschiedeten C++-Standards „C++11“.

Dieser beinhaltet nicht nur vielfältige Verbesserungen hinsichtlich der Sprache an sich, sondern auch bezüglich der Standardbibliothek. Als größte Neuerung dürfte die Einführung von Threads als Teil des Sprachstandards sowie die Etablierung des zugehörigen Memory-Modells gelten. Michael Wong erklärte sich angesichts des hohen Interesses an seinem Vortrag spontan bereit, den Aspekt „nebenläufige Programmierung mit Threads“ im Rahmen der internationalen GridKa School des SCC weiter zu vertiefen.



Michael Wong berichtete vor vollbesetztem Hörsaal zwei Stunden lang über den neuen C++-Standard C++11.
Foto: Christoph-Erdmann Pfeiler

Vielfältige weitere Neuerungen resultieren in einer insgesamt „runderen“ und deutlich ausdrucksstärkeren Pro-

grammiersprache. So erhöhen etwa das „auto“-Keyword zur automatischen Typfestlegung von Variablen, ein „foreach“-Äquivalent, die Initialisierung von Variablen im Klassenrumpf oder das Überladen von Konstruktoren den Komfort erheblich. Die „move“-Semantik erhöht ihrerseits die Performance. Lambda-Funktionen als Teil des Standards ermöglichen es, deutlich einfachere und lesbarere Programme zu schreiben.

Das letzte Update zum C++-Sprachstandard stammte dabei von 2003. Angesichts dieses langen Intervalls und der Konkurrenz durch andere Sprachen mit im Vergleich dynamischerer Entwicklung dürften die nun verfügbaren neuen Möglichkeiten nicht zuletzt durch Nutzer aus der Forschung willkommen geheißen werden. Angesichts ihrer Hardwarenähe existieren kaum Alternativen zu C++, wenn - wie im High Performance Computing häufig - die Performance einer Applikation ein bestimmender Faktor ist. Gerade für Nutzer der einschlägigen Einrichtungen des SCC dürfte der neue Standard also teils erhebliche Verbesserungen in puncto Komfort bei der Programmentwicklung sowie der resultierenden Applikations-Performance bringen. Hierzu trägt auch bei, dass der Technical Report 1 aus dem Jahr 2005 – dort wurden Empfehlungen für die Erweiterung des Standards gegeben – nun zum überwiegenden Teil in C++11 umgesetzt und damit für die Anbieter von Compilern bindend geworden ist. Wichtige Neuerungen umfassen hier etwa die Einführung generalisierter Binder (std::bind) oder eine verbesserte Art von Smart Pointern (std::shared_ptr). Der gerade im wissenschaftlichen Umfeld häufig eingesetzte GNU C++ Compiler „g++“ verfügt übrigens bereits heute über eine sehr weitreichende Abdeckung von C++11.

9. Internationale GridKa Summer School am SCC

Vom 5. bis 9. September veranstaltete das SCC zum neunten Mal die internationale GridKa Summer School, eine der weltweit führenden Sommerschulen für Grid und Cloud Computing. In hochkarätigen Vorträgen, Workshops und praktischen Übungen wurde den knapp 100 Teilnehmern aus 11 Nationen auch in diesem Jahr wieder der neueste Stand der Technik vermittelt.

„Die GridKa School spiegelt in ihrem Programm das wachsende Interesse an Cloud Computing wider, ohne dabei das seit Jahren erfolgreich verwendete Grid Computing zu vernachlässigen“, erläutert GridKa School-Leiter Dr. Christopher Jung. „In beiden Techniken, und damit auch in der Schule, spielen Virtualisierung und Datenhaltung heute eine Hauptrolle.“

Die GridKa School arbeitet seit Jahren mit allen wichtigen Grid Computing-Projekten in Deutschland und Europa zusammen. Erstmals fungierte sie dieses Jahr als Schweizer Grid-Schule und konnte so ihre internationale Ausrichtung einmal mehr unterstreichen.

Sowohl Anwender als auch Administratoren gehören zur Zielgruppe der Schule. Mit zahlreichen praktischen Übungen, die an drei Tagen in parallelen Kursen stattfanden, konnten die Organisatoren gezielt auf ihre Interessen eingehen. Besondere Highlights der Veranstaltung waren der Eröffnungsvortrag von Jeff Templon (NIKHEF, nationales niederländisches Forschungszentrum für subatomare Physik in Amsterdam) zum Thema „Grids and Clouds“ und die Closing Keynote vom Direktor der „European Grid Initiative“, Steven Newhouse, der dem Grid Computing auch für die Zukunft eine wichtige Rolle vorhersagte.

Das Programm der Schule wurde durch den SCC-Kolloquiumsvortrag von Michael Wong (s. Seite 10), eine Führung durch das GridKa Computing Centre am SCC, einen Kennenlern-Abend mit Flammkuchen und das offizielle School Dinner abgerundet. „Die GridKa School-Teilnehmer nehmen neben dem gewonnenen Wissen immer auch gute Kontakte mit nach Hause“, so das Fazit von Christopher Jung.

(red)



Großes Interesse aus aller Welt: Die Teilnehmerinnen und Teilnehmer der GridKa School 2011 mit Schulleiter Dr. Christopher Jung (zweite Reihe, dritter von links).
Foto: Marek Domiter



Neben exzellentem Fachwissen sorgte auch der Enthusiasmus des Sprechers für „sein“ Thema für einen rundum gelungenen Start des wissenschaftlichen Kolloquiums.

Foto: Christoph-Erdmann Pfeiler

Das wissenschaftliche Kolloquium fungierte auch als Abendveranstaltung der GridKa School. Viele Zuhörer entstammten damit der wissenschaftlichen Nutzergemeinde des High Performance Computing. Die Organisatoren freuten sich jedoch besonders, dass die Zuhörerschaft sich mindestens ebenso sehr aus anderen, auch KIT-externen Bereichen rekrutierte und dass insbesondere viele Teilnehmer aus der Wirtschaft anwesend waren.

Das SCC möchte auf diesen gelungenen Beginn der neuen Vortragsreihe aufbauen und wird auch weiterhin international renommierte Experten einladen, die interessierte Zuhörer des KIT und der Region über aktuelle Themen der Informatik informieren.

Dr. Rüdiger Berlich

9th International GridKa Summer School at SCC

From 5th to 9th September SCC organized for the ninth time the International GridKa Summer School, one of the leading schools for grid and cloud computing worldwide. In top-class presentations, workshops and practical exercises international scientists imparted knowledge about state-of-the-art technology to the almost 100 participants from eleven nations.

Peer Energy Cloud bringt Energie aus der Wolke

Bürgermarktplatz für den Handel mit erneuerbaren Energien

Das Partnerkonsortium Peer Energy Cloud, bestehend aus dem Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI), dem Karlsruher Institut für Technologie (KIT), AGT Germany, der SEEBURGER AG und den Stadtwerken Saarlouis ist zum 1. September 2011 gestartet. Das Projekt gehört zu den zwölf Gewinnern des Technologiewettbewerbs „Trusted Cloud“ des Bundeswirtschaftsministeriums und stellt die bahnbrechende Entwicklung von Cloud Enabled Smart Energy Micro Grids in Aussicht. Auf der Kick-off-Veranstaltung am 19. Oktober in Saarlouis wurde der Startschuss bekannt gegeben.



Gelungener Auftakt: Die Protagonisten der Kick-off-Veranstaltung zum Projekt Peer Energy Cloud am 19.10. in Saarlouis. (Von links nach rechts) Prof. Dr. Lutz Heuser (CEO, AGT Germany), Prof. Dr. Wolfgang Wahlster (CEO, DFKI), Dr. Simone Peter (Ministerin für Umwelt, Energie und Verkehr des Saarlandes), Prof. Dr. Wilfried Juling (CIO/CSO, KIT), Holger Kirchner (SEEBURGER AG), Bernd Seeburger (CEO, SEEBURGER AG), Dr. Ralf Levacher (Geschäftsführer Stadtwerke Saarlouis), Roland Henz (Oberbürgermeister Saarlouis). Quelle: Peer Energy Cloud

tungsweisenden Marktplatz im Energiebereich deutliche Akzente im lokalen Energiehandel gesetzt werden können. Mit unserer umfangreichen Erfahrung im Bereich Cloud/Marktplätze sind wir für dieses Vorhaben bestens gerüstet“, erklärte Bernd Seeburger, Geschäftsführer der SEEBURGER AG.

„Aufgrund der Energiewende hat unser Projekt höchste energiepolitische Relevanz. Erneuerbare Energie muss möglichst lokal dort verbraucht werden, wo sie dezentral zum Beispiel über private Photovoltaik-Anlagen erzeugt wird. Nur so kann der notwendige Netzausbau so gering wie möglich gehalten und zu starke Netzschwankungen vermieden werden. Dabei kommt digitalen Bürgermarktplätzen, die sicher und preiswert den privaten und lokalen Energiehandel ermöglichen, eine Schlüsselstellung zu. Wir haben im Saarland durch dieses Projekt die einmalige Chance, bundesweit auf diesem Gebiet zum Pionier und Technologieführer zu werden“, so Prof. Dr. Wolfgang Wahlster, Vorsitzender der Geschäftsführung des Deutschen Forschungszentrums für Künstliche Intelligenz, DFKI.

„Peer Energy Cloud“ erhielt auf der Veranstaltung den Titel „Projekt im Software-Cluster“. Diese Auszeichnung für innovative Projekte im Bereich der Unternehmenssoftware wird vom Strategieboard des Software-Clusters verliehen, dem namhafte Persönlichkeiten der deutschen IT-Wirtschaft sowie international renommierte Wissenschaftler angehören. Der Software-Cluster rund um die Städte Darmstadt, Kaiserslautern, Karlsruhe, Saarbrücken und Walldorf ist Europas leistungsstärkstes Netzwerk von Unternehmen, Ausbildungs- und Forschungseinrichtungen im Bereich der Software-Entwicklung.

Ziel des Projekts ist die Entwicklung von vertrauenswürdigen Lösungen zur intelligenten Verteilung und Nutzung erneuerbarer Energien und die Erforschung sicherer, energiebezogener Peer-to-Peer Cloud-Dienste für den Energie-Marktplatz der Zukunft.

In einer spezialisierten Cloud-Infrastruktur entsteht beispielhaft ein neuartiger Marktplatz zum An- und Verkauf von Energiekontingenten für Prosumer (Erzeuger und Verbraucher). Dabei geht es praktisch darum, den Stromverbrauch in privaten Haushalten mit der Erzeugung in den Kraftwerken in Balance zu bringen. Ein Micro Grid bestehend aus ca. 100 Haushalten mit mehreren Photovoltaikanlagen im Stadtgebiet von Saarlouis wird von den dortigen Stadtwerken aufgebaut und betreut. Die Wohneinheiten sind neben dem Stromnetz mit einem Glasfasernetz zur Datenübertragung miteinander verbunden. Zur Übertragung vertraulicher, sensibler Daten innerhalb dieses Micro Grids steht eine gesonderte, gesicherte Glasfaserleitung zur Verfügung.

„Informations- und Kommunikationstechnologien spielen bei der Peer Energy Cloud eine zentrale Rolle. In unserem Projekt soll daher eine vertrauenswürdige Umgebung geschaffen werden, in

der diese Herausforderung sicherer und zuverlässiger gemeistert wird“, so Prof. Dr. Wilfried Juling, Chief Science & Information Officer (CSO/CIO) des KIT.

Die privaten Haushalte verbrauchen rund ein Drittel des Stroms, der von Kraftwerken produziert und über die Netze der rund 700 Stadtwerke geliefert wird. Aktuell ist der tatsächliche Strombedarf zu einem bestimmten Zeitpunkt für die Stromerzeuger jedoch nicht verbraucher-genau voraussagbar. Standardlastprofile sind ungenau, der Lastgang lässt sich kaum optimieren, der Netzausbau muss sich an Spitzenlasten orientieren, Backup-Kraftwerke sind betriebsnotwendig. Innerhalb des Projekts sollen daher innovative Erfassungs- und Prognoseverfahren für die Lastgangentwicklung mit dem Ziel der Lastflussoptimierung entwickelt sowie ein virtueller Marktplatz für den Stromhandel etabliert werden.

„Wir freuen uns als führender Anbieter von Produkten und Lösungen zur Integration von Geschäftspartnern mit dabei zu sein. Basierend auf unseren Erfahrungen in der erfolgreichen Umsetzung der Marktkommunikation im Energiebereich, gehen wir jetzt gemeinsam mit den beteiligten Partnern die Stufe zwei der Energiewende an. Wir gehen davon aus, dass durch diesen neuartigen rich-

(red)

Das SCC stellt sich vor

In dieser Ausgabe: Die Abteilung Verteilte Systeme und Grid (VSG)



Dr. Andreas Heiss ist seit dem 1. November 2011 Leiter der Abteilung VSG. Er studierte Physik an der Universität Karlsruhe und kam nach einem knapp zweijährigen Aufenthalt am Fermilab bei Chicago und seiner Promotion im Bereich der Elementarteilchenphysik an das Institut für Wissenschaftliches Rechnen des damaligen Forschungszentrums Karlsruhe. Dort arbeitete er an einem F&E-Projekt im Zusammenhang mit Hochgeschwindigkeitsnetzwerken sowie im Projekt CampusGrid mit. 2004 wechselte er in die damalige Abteilung Grid-Infrastruktur und -Service und übernahm 2008 die Projektleitung des Grid Computing Centre Karlsruhe (GridKa).

Dr. Torsten Antoni ist seit dem 1. November 2011 stellvertretender Leiter der Abteilung VSG. Er studierte Physik an der Universität Karlsruhe und promovierte an der Universität Heidelberg im Bereich Astroteilchenphysik. Diplom- und Doktorarbeit fanden im Rahmen des KASCADE-Experiments am damaligen Forschungszentrum Karlsruhe statt. 2004 kam er an das Institut für Wissenschaftliches Rechnen und übernahm die Leitung der Arbeitsgruppe „Global Grid User Support (GGUS)“. Seit 2010 ist er Projektmanager für die nationalen Grid-Initiativen „NGI-DE“ und „DGI-2“.



Dr. Holger Marten war bis zum 1. November 2011 Leiter der Abteilung Verteilte Systeme und Grid (VSG) und ist gewählter Sprecher des KIT-Kompetenzfeldes Hochleistungsrechnen und Verteilte Systeme. Er studierte Physik an der Christian-Albrechts-Universität zu Kiel und promovierte anschließend in Astrophysik. 1995 kam er an das damalige Forschungszentrum Karlsruhe und wechselte 1999 in die Hauptabteilung Informations- und Kommunikationstechnik. Hier übernahm er 2001 die Koordination für die Konzeption des späteren Grid Computing Centre Karlsruhe (GridKa) sowie in der Folge ab 2002 die Leitung der neu gegründeten Abteilung Grid-Infrastruktur und -Service (heutige Abteilung VSG) und die Projektleitung für GridKa bis 2008. Seit dem 1. November 2011 ist er Leiter der Abteilung Scientific Computing und Simulation (SCS).



Die Abteilung VSG (von links nach rechts): Manfred Alef, Dr. Torsten Antoni, Marian Zvada, Dr. Andreas Heiss, Dimitri Nilsen, Andreas Petzold, Dr. Christopher Jung, Ursula Epting, Stefan Bozic, Günter Grein, Dr. Holger Marten, Andreas Motzke, Dr. Sven Hermann, Dr. Jie Tao, Tobias König, Sabine Reißer, Helmut Dres, Melanie Ernst, Foued Jrad, Wilhelm Bühler.
Nicht im Bild: Oleg Dulov, Hans-Ulrich Hohn, Ingrid Schöffner, Dr. Pavel Weber.

Fotos: Markus Breig

Grid Computing ist eine Schlüsseltechnologie, die Wissenschaftlern und Ingenieuren aus Forschung und Wirtschaft weltweit verteilte Daten und IT-Ressourcen zur Verfügung stellt. Über schnelle Weitverkehrsnetze integrieren Grids heute Rechenressourcen von Desktop-Rechnern über Cluster- und Visualisierungssysteme bis hin zu Höchstleistungsrechnern. Dazu zählen auch verteilte Datenspeicher und -archive vom Terabyte-bis in den Multi-Petabyte-Bereich sowie Daten unterschiedlichster Typenklassen und Disziplinen, wie Gendatenbanken, Klimadaten und -simulationen oder Messungen mit Synchrotron-Strahlungsquellen, Satelliten, Teleskopen, Teilchendetektoren und anderen wissenschaftlichen Großexperimenten. Der transparente und standortunabhängige Zugriff auf eine solche verteilte Infrastruktur erlaubt es, komplexe wissenschaftliche und gesellschaftliche Fragestellungen zu bearbeiten und in neuartiger Weise interdisziplinär zu kooperieren. Im Rahmen nationaler und internationaler Projekte beteiligt sich die Abteilung „Verteilte Systeme und Grid“ (VSG) seit vielen Jahren am Aufbau und Betrieb dieser Infrastruktur und arbeitet als Service- und Ressourcen-Provider an der Schnittstelle zwischen Nutzern und Anbietern.

Gegenseitige Vertrauensstellung bei der Nutzung von Grids

Grids bilden Daten- und IT-Infrastrukturen, in denen Anbieter ihre Dienste organisations- und länderübergreifend zur Verfügung stellen. Unabdingbare Basis für eine derartige Zusammenarbeit ist eine gegenseitige Vertrauensstellung. Bereits frühzeitig hat sich die Abteilung VSG daher an dem Aufbau einer PKI (Public Key Infrastructure) im Rahmen des internationalen Grid-Trust-Netzwerkes IGTF beteiligt. x.509-Zertifikate sollen gewährleisten, dass es sich bei einem Nutzer, aber auch bei einem verwendeten Dienst, tatsächlich um denjenigen handelt, für den er sich ausgibt. Die am SCC aufgebaute GridKa-CA stellt solche international anerkannten „Grid-Zertifikate“ für Personen und Dienste in Deutschland aus und greift zur Identitätsprüfung von Personen auf ein etabliertes Netzwerk aus Registrierungsstellen an zahlreichen deutschen Hochschulen und wissenschaftlichen Einrichtungen zurück.

Neben x.509-Zertifikaten bilden sogenannte Virtuelle Organisationen (VOs) den zweiten Vertrauensbaustein für die Nutzung des Grid. Erst als Mitglied einer solchen oft auch als „Community“ bezeichneten Personengruppe stehen dem Nutzer die für diese VO freigeschalteten Ressourcen oder Dienste zur Verfügung, und der Anbieter kann sich mit seinen (gegebenenfalls spezialisierten) Diensten auf bestimmte Kundengruppen fokussieren bzw. eine darüber hinausgehende Fremdnutzung seiner Ressourcen einschränken.

GridKa – deutscher Beitrag zum LHC Computing

Getrieben durch die Notwendigkeit, weltweit mehreren tausend Wissenschaftlern Ressourcen zur Datenauswertung der künftigen LHC-Experimente am CERN von jährlich etwa 15 Petabyte zur Verfügung zu stellen, war die internationale Gemeinschaft der Hochenergiephysiker in den vergangenen Jahren besonders stark an der Grid-Entwicklung beteiligt. Genauer wurde zu diesem Zweck neben dem Bau des LHC-Beschleunigers sowie der vier Detektoren ALICE, ATLAS, CMS und LHCb das World Wide LHC Computing Grid-Projekt WLCG als sechstes Teilprojekt gegründet, an dem auch das SCC mit der Abteilung VSG maßgeblich beteiligt ist. Gemeinsam mit den deutschen Kern- und Elementarteilchenphysikern wurde am SCC das Grid Computing Centre Karlsruhe (GridKa) als deutscher Beitrag zum LHC Computing entwickelt und speziell für den hohen Datendurchsatz der LHC- sowie weiterer Hochenergiephysikexperimente mit deutscher Beteiligung optimiert. Mit über 1.200 Compute-

Knoten (mehr als 13.000 CPU-Kernen) sowie je etwa 10.000 Terabyte Festplatten- und Magnetbandspeicher stellt GridKa heute ein Zentrum mit beachtlicher Leistungsfähigkeit dar, das den vier LHC-VOs (mit ihren gleichnamigen Detektoren) im Verbund mit 10 weiteren, vergleichbaren Zentren der Schicht Tier-1, weltweit über 120 kleineren Rechenzentren der Schicht Tier-2 sowie CERN als Datenquelle und Tier-0 die Auswertung ihrer Experimentdaten auf dem Grid ermöglicht.

Tatsächlich sind die Ressourcen-Provider dieses WLCG [1] selbst wiederum Mitglieder mehrerer, weitaus größerer Grid-Infrastrukturen, auf denen Dutzende Virtueller Organisationen unterschiedlicher Wissenschaftsdisziplinen arbeiten. Speziell in Europa ist hier die Europäische Grid-Initiative EGI zu nennen, unter deren Dach sich nationale oder multinationale Grid-Initiativen zu einer gemeinsamen Infrastruktur zusammengeschlossen haben. Das amerikanische Pendant hierzu stellt Open Science Grid dar, asiatische Länder und Australien arbeiten gemeinsam am Asia Pacific Grid. Keine dieser verteilten Infrastrukturen ist in sich abgeschlossen. Vielmehr besteht die Herausforderung darin, trotz unterschiedlicher eingesetzter Basistechnologien den Nutzergemeinschaften sämtliche Ressourcen über geeignet standardisierte Schnittstellen weltweit zur Verfügung zu stellen.

Grid Core und Site Services

Um den transparenten Zugriff auf die Ressourcen zu ermöglichen, ist eine Reihe von Diensten notwendig, die gemeinhin in so genannte Grid Core und Site Services eingeteilt werden. Core Services ermöglichen beispielsweise das Auffinden freier Compute-Ressourcen, eine entsprechende Verteilung der Rechenaufträge von Nutzern, das Auffinden der Daten anhand von File-Katalogen, den transparenten File Transfer zwischen lokalen Speichersystemen an verschiedenen Ressourcen-Standorten usw. Die dafür ebenfalls notwendigen, zentralen Informationsdienste werden von entsprechenden Site Services an den Ressourcen-Standorten gespeist. Diese stellen eine Vermittlungsschicht zwischen den lokalen Systemen des Ressourcen-Betreibers und den zentralen Grid Core Services dar und geben beispielsweise zu jedem Zeitpunkt Auskunft über die Anzahl und Art lokal verfügbarer Job Slots oder Speicherressourcen unterschiedlicher Speicherklassen, übermitteln die Rechenaufträge an das lokale Batch-System, liefern den Standard-Output zurück oder übermitteln anonymisierte Accounting-Informationen an ein zentrales System, um einem VO-Manager den Überblick über verbrauchte Ressourcen

„seiner“ VO zu ermöglichen. Die Abteilung VSG betreibt im Rahmen der nationalen Grid-Initiativen NGI-DE und D-Grid mehrere solcher Core Services als Verbindungsglied zur europäischen Grid-Infrastruktur sowie Site Services für diverse lokale Grid-Ressourcen, die in regionalen, nationalen und internationalen Projekten zur Verfügung gestellt werden.

Monitoring-System für Ressourcen und Dienste

Der Betrieb einer solchen, verteilten IT-Infrastruktur ist nicht trivial. Zu jedem Zeitpunkt muss gewährleistet sein, dass die angebotenen Dienste gewissen Qualitätsanforderungen entsprechen, und sowohl VOs als auch Ressourcen-Provider benötigten Werkzeuge, mit denen die über Service Level Agreements vereinbarten Service-Qualitäten überprüft bzw. nachgewiesen werden können. Es wurden daher in den letzten Jahren umfangreiche Monitoring-Systeme entwickelt, die die Dienste bis hinunter zu Teilfunktionalitäten überwachen, Fehlersituationen erkennen und automatisiert Experten-Teams benachrichtigen, um die Fehler zu beheben. Die Abteilung VSG ist maßgeblich an der Entwicklung und dem Betrieb eines solchen Monitoring-Systems für die Ressourcen und Dienste deutscher und schweizerischer Anbieter beteiligt und arbeitet im Verbund mit mehreren deutschen und schweizerischen Institutionen in einem verteilten Operations-Team in den Bereichen der Service-Überwachung und Unterstützung von Ressourcen-Anbietern bei der Fehlerbeseitigung.

Global Grid User Support System (GGUS) – zentrale Plattform für Grid-Nutzer

Schließlich gilt aber auch aus der Sicht eines einzelnen Nutzers ein zentrales Problem zu lösen: An wen wendet sich dieser eigentlich, wenn seine Anwendung irgendwo auf der Welt nicht (mehr) so funktioniert wie erwartet? Zu diesem Zweck hat eine Arbeitsgruppe innerhalb der Abteilung VSG das Global Grid User Support System GGUS entwickelt. GGUS stellt eine zentrale Plattform dar, in der Nutzer wie Administratoren technische Anfragen stellen und deren weitere Bearbeitung verfolgen können. Im Hintergrund sind Dutzende weltweit verteilter Expertenteams – von Administratoren lokaler und globaler Services über Anwendungsunterstützer bis hin zu Grid-Middleware-Entwicklern – angeschlossen, denen solche Anfragen über entsprechende Workflows zugewiesen werden. Die GGUS-Plattform wird inzwischen in mehreren, sehr großen europäischen IT-Projekten verwendet und wurde in den vergangenen Jahren zu einem Alleinstellungsmerkmal des KIT herausgearbeitet.

Ausbildung

In der Abteilung VSG werden regelmäßig Diplom- und Doktorarbeiten im Bereich der Integration von Grid und Cloud Computing angefertigt. In Zusammenarbeit mit der Dualen Hochschule Baden-Württemberg übernimmt VSG auch die praktische Ausbildung von Bachelorstudenten im Bereich Informationstechnik. Sie leistet ferner Beiträge zur Lehre, zum Beispiel durch Beteiligung an der im Wintersemester 2011/12 am KIT erstmals angebotenen Vorlesung „Verteilte Systeme: Grid und Cloud“, und ist seit 2003 Ausrichter der jährlichen Sommerschule für Grid und Cloud Computing.

Projekte

Europäische Grid-Initiative (EGI) / EGI-InSPIRE

Die EU-Kommission hat in ihrem 7. Rahmenprogramm umfangreiche Fördermittel für eine europäische Grid-Initiative (EGI) zur Verfügung gestellt. Über einen europäischen Dachverband (EGI.eu) sollen nationale Grid-Initiativen (NGIs) zu einer leistungsfähigen e-Infrastruktur verbunden und für die europäische Wissenschaftslandschaft nachhaltig betrieben werden. Deutschland hat auf diese Initiative reagiert, indem der kürzlich gegründete Gauß-Allianz e.V., bestehend aus 13 deutschen Hochleistungsrechenzentren – darunter das KIT – den Aufbau und Betrieb der deutschen Grid-Initiative NGI-DE an ein Konsortium aus den vier Partnern DESY (Hamburg), FZJ (Jülich), KIT und LRZ (München) übertragen hat. Das KIT ist hier Konsortialführer und war maßgeblich an der Projektplanung von EGI-InSPIRE (Integrated Sustainable Pan-European Infrastructure for Researchers in Europe) beteiligt.

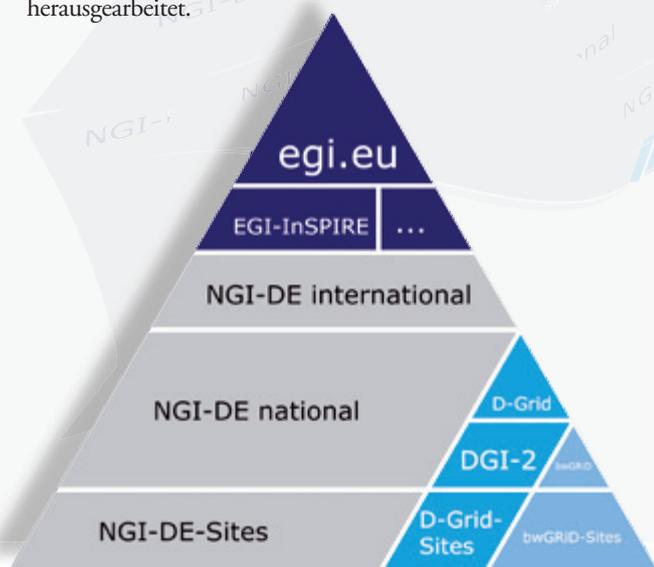
Nationale Grid-Initiative Deutschland (NGI-DE)

NGI-DE (Nationale Grid-Initiative Deutschland) ist eine von etwa 50 nationalen Grid-Initiativen, die an EGI beteiligt sind. Innerhalb der NGI-DE wirkt VSG insbesondere in den Bereichen „Projektbüro und Nachhaltigkeit“, „zentrales Monitoring“, „Helpdesk“ sowie „Sicherheit“ mit und stellt mit GridKa eines der größten Rechenzentren.

Mit dem EGI-InSPIRE-Arbeitspaket „Grid Operations and Support Centre“ (GOSC) hat das KIT eine strategische Schlüsselrolle für die Anbindung der NGI-DE an die europäische Grid-Initiative EGI. Im internationalen Umfeld verfolgt das GOSC drei wesentliche Ziele:

- Die Entwicklung von Basis-Technologien sollen Wissenschaftlern unterschiedlichster Disziplinen einen optimalen Zugang zu verteilten Ressourcen in Deutschland und Europa ermöglichen,
- die Entwicklung und Umsetzung von Policies soll im europäischen Verbund den nachhaltigen und zuverlässigen Betrieb der nationalen Rechen- und Datenressourcen gewährleisten,
- die Weiterentwicklung eines internationalen Helpdesk-Systems (GGUS) soll ermöglichen, dass Nutzer und Betreiber verteilter Ressourcen im internationalen Umfeld die jeweils bestmögliche Unterstützung bei Support-Anfragen erhalten.

Abbildung 1: Grid-Projekte auf internationaler, nationaler und regionaler Ebene.



Auf nationaler Ebene soll das GOSC eine Konsolidierung von Grid-Ressourcen unterstützen, die von verschiedenen Projekten innerhalb Deutschlands zur Verfügung gestellt werden. Unter anderem sollen der Zugang zu den Ressourcen, das Monitoring, die Nutzungskonditionen und der User Support vereinheitlicht werden. Über das GOSC werden auch die Nutzer und Anbieter der LSDF- und SimLab-Ressourcen bzw. -Services in entsprechende Support-Strukturen eingebunden.

Deutsche Grid-Initiative (D-Grid)

Zur Förderung des Grid Computing wurde 2004 zusammen mit dem BMBF unter Beteiligung des KIT die Deutsche Grid-Initiative (D-Grid) gestartet. Für dieses Forschungsverbundvorhaben mit zahlreichen Großforschungseinrichtungen, Hochschulen und Unternehmen der Wirtschaft hat das BMBF in acht Jahren insgesamt über 100 Millionen Euro bereitgestellt.

Am KIT sind neben dem SCC mit mehreren Abteilungen (IDA, ISM, VSG) auch die Institute AIFB, IAI, IISM und IPE am D-Grid-Verbund beteiligt oder beteiligt gewesen. Das SCC ist seit Beginn Konsortialführer für das Integrationsprojekt des D-Grid (DGI-2 bzw. davor DGI), das für den Betrieb und die Weiterentwicklung des D-Grid sorgt. Bei VSG sind dies insbesondere die Bereiche „Projektbüro und Nachhaltigkeit“, „zentrales Monitoring“, „Helpdesk“ und „Referenzinstallation“.

Grid-Projekt des Landes Baden-Württemberg „bwGRiD“

Das baden-württembergische Grid-Projekt „bwGRiD“ kann als regionales Infrastrukturprojekt verstanden werden, das in Teilen vom BMBF über die D-Grid-Initiative und in Teilen vom baden-württembergischen Ministerium für Wissenschaft, Forschung und Kunst finanziert wird. Im Januar 2010

übernahm das SCC vom HLRS in Stuttgart hierfür die Projektleitung. Die seither in Baden-Württemberg im Rahmen des bwGRiD entstandene Infrastruktur ist sowohl in Bezug auf die Zusammenarbeit der acht Hochschulrechenzentren als auch im Vergleich zu anderen deutschen Bundesländern einzigartig (s. Seite 8). Im Rahmen des bwGRiD-Portalprojektes arbeitet VSG in enger Zusammenarbeit mit der Universität Ulm an der Entwicklung des Portal-Templates Gatlet. Ziel von Gatlet ist es, Wissenschaftlern eine Schnittstelle zu nationalen und regionalen Grid-Infrastrukturen, zum Beispiel NGI-DE und bwGRiD, an die Hand zu geben, auf deren Basis eigene Anwendungen ins Grid gebracht werden können. Mit Gatlet können dann mithilfe eines Webbrowsers Rechenjobs abgeschickt, überwacht und anschließend deren Ergebnisse abgeholt werden.

Im Juli 2011 wurde in enger Kooperation mit dem bwGRiD-Projekt außerdem das Projekt „bwGRiD - ergänzende Maßnahmen“ begonnen. Dieses Projekt soll die bereits vorhandene bwGRiD-Infrastruktur stärken und vor allem den Nutzerkreis erweitern. Das Projekt arbeitet eng mit den in anderen Abteilungen am SCC begründeten baden-württembergischen Projekten „bwIDM“ (Identitäts-Management) und „LSDF“ (Large Scale Data Facility) zusammen. Dabei soll die durch die genannten Projekte aufgebaute Infrastruktur nachhaltig betrieben werden.

Betrieb

GridKa – deutsches Tier-1-Zentrum für den LHC

Das Grid Computing Centre Karlsruhe (GridKa) wurde 2002 als Rechen- und Datenzentrum für die deutsche Teilchenphysik gegründet und unterstützt mittlerweile 10 Experiment-Communities aus der Teilchen- und Astroteilchenphysik. Im

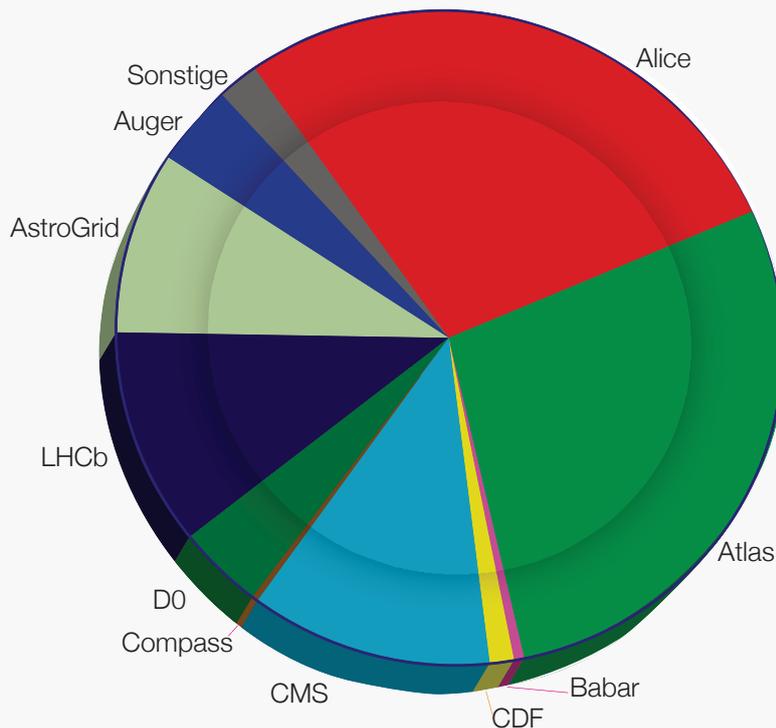


Abbildung 2: Nutzung des GridKa-Clusters durch die beteiligten Experimentgruppen im September 2011.

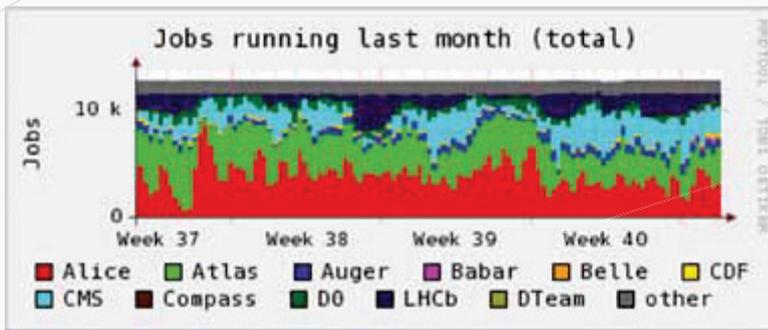


Abbildung 3: Am GridKa ausgeführte Benutzerjobs (Monatsübersicht).

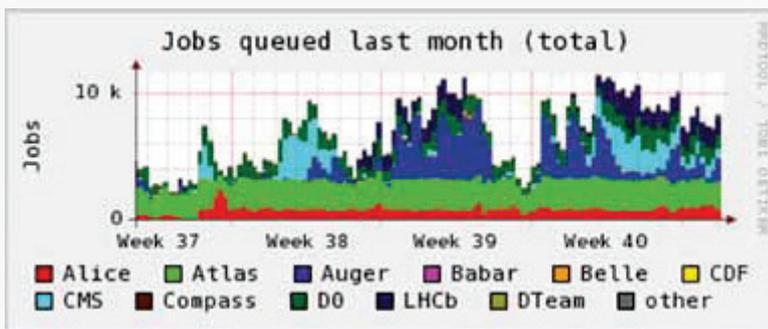


Abbildung 4: Am GridKa auf Ausführung wartende Benutzerjobs (Monatsübersicht).

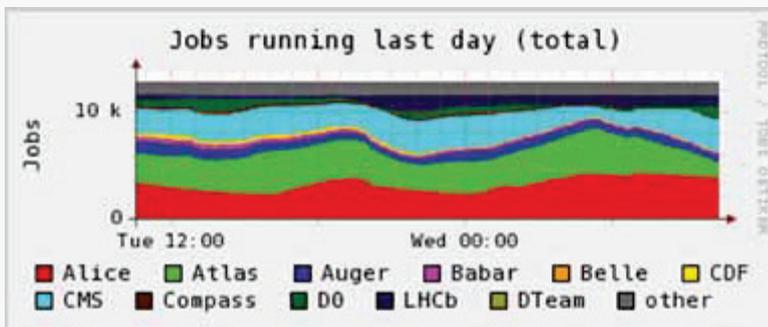


Abbildung 5: Am GridKa ausgeführte Benutzerjobs (Tagesübersicht).

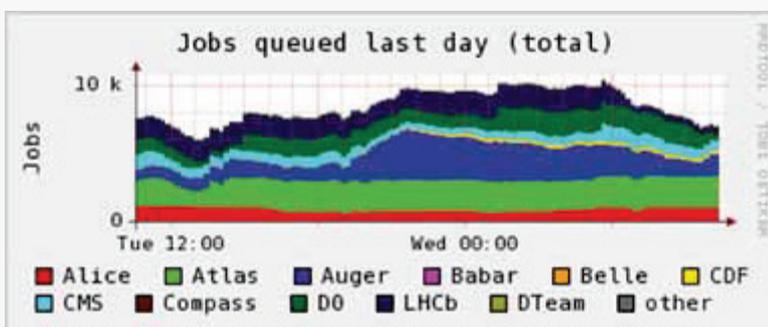


Abbildung 6: Am GridKa auf Ausführung wartende Benutzerjobs (Tagesübersicht).

Worldwide LHC Computing Grid (WLCG) ist GridKa eines der größten der insgesamt 11 Zentren der Tier-1-Schicht und als solches unter anderem für die Archivierung und regelmäßige Reprozessierung eines Teils der Rohdaten der vier LHC-Experimente zuständig. Die Experiment-Daten werden über eine 10 Gbit/s-Netzwerkverbindung vom CERN importiert, während die bei GridKa vorverarbeiteten Daten über mehrere solcher 10 Gbit/s-Verbindungen zu anderen Tier-1- und Tier-2-Zentren übertragen werden.

Um die immer weiter anwachsende Datenmenge der LHC-Experimente regelmäßig mit neuesten Algorithmen und Methoden zu (re-)prozessieren, wird die Rechenleistung von GridKa jährlich erhöht. Zur standardisierten Messung der bereitgestellten Rechenleistung wird von allen WLCG-Zentren der Benchmark HEP-SPEC06 verwendet, an dessen Entwicklung VSG maßgeblich beteiligt war. Derzeit laufen die Vorbereitungen für die nächste Erhöhungsstufe. Im kommenden Frühjahr werden alte und ineffiziente Systeme durch neue, energieeffizientere Rechner ersetzt und die Gesamtrechenleistung dabei von derzeit ca. 108.000 HEP-SPEC06 auf rund 130.000 HEP-SPEC06 erhöht.

Neben der Planung und dem Betrieb des GridKa-Clusters sowie der dafür notwendigen Resource Management- und Accounting-Systeme plant und betreibt die Abteilung VSG im Rahmen des GridKa-Projektes hochverfügbare Grid Services für die LHC- und andere User Communities. Über die Workload Management-Systeme (WMS) und Compute-Elemente (CE) werden die Rechenaufträge an geeignete Grid-Zentren, zum Beispiel GridKa, und dort weiter an deren lokales Resource Management-System geschickt.

Standardisierte Informationssysteme (Berkeley Database Information Index, BDII) liefern über Grid-Schnittstellen Informationen über Grid-Rechenzentren, beispielsweise welche Virtuellen Organisationen unterstützt werden und ob freie Rechen- und Speicher-Ressourcen vorhanden sind. Für das Datenmanagement werden File-Kataloge (LFC) sowie ein File Transfer Service (FTS) betrieben, der sämtliche Datentransfers zwischen GridKa und den angeschlossenen Tier-2- sowie Tier-1-Zentren steuert und jährlich viele Petabytes an Daten zwischen nationalen und internationalen Grid-Zentren und GridKa transferiert. Verschiedene Proxy-Dienste erlauben den effizienten Zugriff auf Datenbanksysteme am SCC und am CERN. Für Community-spezifische Dienste werden Vorrechner und sogenannte VO-Boxen bereitgestellt.



Services

D-Grid-Referenzinstallation

Das D-Grid bietet den Grid-Zugang über verschiedene Middlewares an – derzeit Globus Toolkit, gLite und UNICORE –, um verschiedene Nutzeranforderungen optimal zu bedienen. Eine prototypische Installation einer D-Grid Site ist die D-Grid-Referenzinstallation [4], die auf einem virtuellen Cluster eine vollwertige Site mit detaillierter Dokumentation bietet. Diese kann sowohl als Vorlage für das Aufsetzen neuer D-Grid Sites als auch für Funktions- und Kompatibilitätstests genutzt werden.

Monitoring für mehr als 25 Rechenzentren

Um die Funktion der verschiedenen Grid-Dienste zu überwachen, entwickelt VSG gemeinsam mit anderen Forschungseinrichtungen Monitoring-Systeme, die größtenteils auf der Open Source Software Nagios basieren. Nagios wurde speziell auf die Bedürfnisse des Grid Computing angepasst. Ein Beispiel solcher Monitoring-Systeme ist das interne Nagios-Monitoring für das GridKa-Tier-1-Zentrum, das inzwischen von der GridKa-Rufbereitschaft als Basiswerkzeug eingesetzt wird.

Seit dem Start von EGI betreibt VSG erfolgreich die regionale NGI-DE-Monitoring-Instanz für die Region Deutschland-Schweiz. Zurzeit werden Grid-Dienste von über 25 Rechenzentren kontinuierlich aus Karlsruhe überwacht.

Zusätzlich betreibt VSG in Zusammenarbeit mit dem Forschungszentrum Jülich und dem Leibniz-Rechenzentrum in München die zentralen Middleware-Monitoring-Systeme für das D-Grid. Zurzeit wird daran gearbeitet, die D-Grid-Rechenzentren in die NGI-DE-Nagios-Überwachung einzubinden.

GridKa-CA – Zertifikate für Benutzer, Rechner und Applikationen in ganz Deutschland

Die Zertifizierungsstelle GridKa-CA am KIT ist seit 2002 Mitglied der European Grid Policy Management Authority EuGridPMA, einer internationalen Organisation, der über 40 Länder angehören. Gemeinsam mit TAGPMA (Nord-, Mittel und Südamerika) und APGridPMA (Asien, Pazifik, Australien) entstand ein globaler Verbund der International Grid Trust Federation (IGTF), wodurch weltweit Vertrauen hergestellt wird.

Digitale Zertifikate enthalten einen öffentlichen Schlüssel, Zusatzinformationen und die Signatur der Zertifizierungsstelle. Die Zertifizierungsstelle beglaubigt mit ihrer Signatur die Zugehörigkeit eines öffentlichen Schlüssels zu einer bestimmten Person. Die Zertifikate dienen der Authentifizierung und der Ver- und Entschlüsselung vertraulicher Daten, die über das Internet und andere Netze verbreitet werden.

Die GridKa-CA [5] stellt Standard X.509-Zertifikate für Benutzer, Rechner und Applikationen in ganz Deutschland aus. Zur langen Liste der Organisationen, die GridKa-CA-Zertifikate verwenden, gehören sowohl Forschungseinrichtungen und Universitäten als auch Industrieunternehmen.

Support

GGUS – weltweit verteiltes Support-System

Global Grid User Support [6], abgekürzt GGUS, heißt das weltweit verteilte Support-System, das von einer Arbeitsgruppe innerhalb der Abteilung VSG geleitet und betrieben wird. Wissenschaftler von internationalen Grid-Projekten (EGI [7]) oder von sogenannten Virtuellen Organisationen wie zum Beispiel der Hochenergiephysik (LHC-Experimente am CERN [1]) nutzen dieses System, um Störungen im Zusammenhang mit dem Computing Grid zu melden, sich selbst Tipps und Tricks aus der immer umfangreicher werdenden Ticketdatenbank zu holen oder sich mit wichtigen News zu versorgen.

Herzstück von GGUS ist eine komplexe Webapplikation, die durch die Bereitstellung von Webservice-Schnittstellen die Integration von weiteren Helpdesk-Systemen nach föderalischem Prinzip ermöglicht.

Derzeit sind 15 externe Systeme mit GGUS über diese Schnittstellen verknüpft. Bislang konnten über 100 Expertengruppen in allen Kontinenten der Erde gewonnen werden, die rund um die Uhr Support für die ständig wachsende Gemeinschaft der Grid-Benutzer leisten.

xGUS – Helpdesk-Portal für erfolgreichen User Support

Mit dem Helpdesk Template xGUS wird Organisationen, deren Mitglieder das Grid nutzen, ein per Web zugängliches Helpdesk-Portal mit allen notwendigen Funktionen für erfolgreichen User Support zur Verfügung gestellt. Diese Organisationen können nationale Grid-Initiativen, User Communities oder Ähnliches sein. Ein xGUS-Portal ist sinnvoll, sobald eine zweistufige Support-Struktur mit einem 1st-Level Support und mehreren 2nd-Level Support Units abgebildet werden soll. Der Vorteil von xGUS gegenüber anderen Ticketsystemen wird im Umfeld des EU-Projekts EGI-InSPIRE besonders deutlich, da xGUS eine direkte Schnittstelle zum zentralen EGI-Helpdesk GGUS (Global Grid User Support) bereitstellt, über die Tickets untereinander ausgetauscht und synchronisiert werden können. Aber auch als unabhängiges Helpdesk ist xGUS genau auf die Bedürfnisse von Grid-Nutzern und Administratoren abgestimmt. Betrieb, Wartung und Weiterentwicklung aller xGUS-Instanzen findet am KIT statt. Zurzeit sind fünf xGUS-Instanzen für die Grid-Initiativen folgender Länder/Regionen in Betrieb: Afrika, Lateinamerika, Serbien, Schweiz und Deutschland.

HEPiX

Die High Energy Physics UNIX Interest Group [2] veranstaltet zweimal im Jahr einen Workshop, an dem sich Vertreter der Rechenzentren aller Hochenergiephysik-Laboratorien zu einem intensiven Gedanken- und Erfahrungsaustausch treffen. Bei Bedarf werden Arbeitsgruppen eingerichtet, die sich konkret mit einer Aufgabe beschäftigen. So hat die HEPiX Benchmark Working Group den Benchmark HEP-SPEC06 (auf der Basis des Industriestandard-Benchmarks SPEC CPU2006 [3]) zur Messung der CPU Performance entwickelt, der für alle am WLCG beteiligten Rechenzentren verbindlich ist. Die HEPiX Virtualization Working Group beschäftigt sich mit Fragen der Virtualisierung im Umfeld von Clustersystemen. Die Abteilung VSG ist an beiden Arbeitsgruppen beteiligt.

NGI-DE-Helpdesk

Das auf xGUS basierende NGI-DE-Helpdesk dient als Support-Schnittstelle für alle Grid-Nutzer in Deutschland, seien es Nutzer des bwGRiD, des D-Grid oder von NGI-DE. Es dient als einziger Anlaufpunkt für alle Probleme mit dem Grid und stellt über eine Schnittstelle zu GGUS auch den Kontakt zu internationalen Support-Gruppen her.

Kontinuierliche Prozessoptimierung mit ITIL

Beim Grid Computing erschwert die weltweite Verteilung der Diensteanbieter und -nutzer den Betrieb der angebotenen IT-Dienste. Auch die Koordination des weltweiten Supports bedarf gut aufeinander abgestimmter Prozesse und Workflows. Das SCC stand vor einer ähnlichen Herausforderung. Bei der Zusammenführung des ehemaligen Rechenzentrums der Universität Karlsruhe und des Instituts für Wissenschaftliches Rechnen des Forschungszentrums Karlsruhe zum SCC wurde deutlich, dass ein Rechenzentrum an zwei Standorten besondere koordinatorische Maßnahmen erfordert.

Daher wurde der Beschluss gefasst, im SCC das Regelwerk ITIL mit den Lebenszyklus-Phasen „Service Strategy“, „Service Design“, „Service Transition“ und „Service Operation“ zur kontinuierlichen Prozessoptimierung anzuwenden. Es wurde ein SCC-internes ITIL-Projekt unter Federführung der Abteilung „ISM - IT-Security und Service-Management“ ins Leben gerufen, welches auch von der Abteilung VSG unterstützt wurde. Der Fokus im Grid-Umfeld lag auf dem Service-Prozess „Service Catalogue Management“, der Erstellung von Service-Beschreibungen und der Zerlegung der Grid Services in so genannte Service-Bausteine und dem Service-Prozess „Incident Management“, insbesondere dem reibungslosen 7*24-Stunden-Betrieb von GridKa. Auch Wartungen des deutschen Tier-1-Rechenzentrums GridKa wurden ITIL-konform, gemäß den Vorgaben des „Change Managements“ geplant und durchgeführt. Es wird nun versucht, aus den SCC-intern gewonnen Erkenntnissen Verbesserungen für das nationale Grid Computing abzuleiten, um Verbesserungen hinsichtlich des Grid-Betriebes und -Supports in Deutschland zu erreichen.

Manfred Alef, Dr. Torsten Antoni, Stefan Bozic, Wilhelm Bühler, Helmut Dres, Ursula Epting, Günter Grein, Dr. Andreas Heiss, Dr. Sven Hermann, Foued Jrad, Dr. Christopher Jung, Tobias König, Dr. Holger Marten, Andreas Motzke, Sabine Reißer, Dr. Jie Tao

Verweise

1. WLCG. <http://lcg.web.cern.ch/lcg/>
2. HEPiX. <http://www.hepix.org>
3. SPEC. <http://www.spec.org>
4. Referenzinstallation. <http://dgiref.d-grid.de/>
5. GridKA-CA. <http://www.gridka.de/ca>
6. GGUS. <http://ggus.eu>
7. EGI. <http://egi.eu>

The SCC Department Distributed Systems and Grid (VSG)

Grid Computing is a key technology that provides scientists and engineers from research and industry with worldwide distributed data and IT resources. Via high performance wide area networks grids integrate today's compute resources from desktops to cluster and visualization systems up to supercomputers. Also distributed data storage and archives from the terabyte to the multi-petabyte range as well as data of various types and disciplines, like genome data bases, climate data and simulation results or measurements with synchrotron radiation sources, satellites, telescopes, particle detectors and other large scale scientific instruments are included. Transparent access from any location to such a distributed infrastructure allows to work on complex scientific and social problems and to collaborate in new interdisciplinary ways. Since many years the SCC department "Distributed Systems and Grid" (VSG) contributes within national and international projects to the development and operations of this infrastructure, and works as a service and resource provider at the interface to the users.



Abbildung 7: Die Webseiten der aktuell genutzten xGUS-Instanzen. (Von hinten nach vorne): Die Grid-Initiativen von Serbien, Lateinamerika und Karibik, Schweiz, Afrika, Deutschland.

Peer Energy Cloud – Teilvorhaben am KIT

Um die wachsenden Herausforderungen einer effizienteren und dynamischeren Organisation des Energiemarkts in Deutschland annehmen zu können, befasst sich das Projekt Peer Energy Cloud (s. Seite 12), an dem auch das KIT beteiligt ist, mit der Entwicklung innovativer Cloud-basierter Methoden im Energiemanagement. Dabei spielen dezentrale Energieversorgungsleistungen beispielsweise über private Photovoltaik-Anlagen sowie digitale Marktplätze für den Stromhandel eine Schlüsselrolle.

Anwendungsszenario Smart Micro Grid

Zur Realisierung eines hohen Anteils regenerativer Energien im Energie-Mix der Bundesrepublik ist im modernen Energiemanagement die so genannte „Lastentkopplung“ der Energienetze erforderlich. Die Strombereitstellung wird dabei nicht mehr allein zentral durch den aktuellen Energieverbrauch gesteuert, sondern auch über so genannte Smart Micro Grids, in denen sich Verbraucher und lokale Erzeuger zusammenschließen. Dieses Szenario beinhaltet sowohl die Etablierung eines virtuellen Marktplatzes für den lokalen Stromhandel wie auch die Konzeption und Entwicklung innovativer Erfassungs- und Prognoseverfahren.

für die im Projekt genutzten Softwarekomponenten (am SCC in der Forschungsgruppe Cloud Computing unter der Leitung von Dr. Marcel Kunze)

- Realisierung von vertrauensfördernden Sicherheitsmechanismen bei allen sicherheitsrelevanten Komponenten (am Institut für Programmstrukturen und Datenorganisation in der Forschungsgruppe Zertifizierbare Vertrauenswürdige Informatiksysteme unter der Leitung von Prof. Dr. Alexander Pretschner).

Der virtuelle Marktplatz als Einstiegs- und Kontrollpunkt für den agentenbasierten Stromhandel ermöglicht einen lokalen Ausgleich zwischen Erzeugung und Verbrauch von Elektrizität

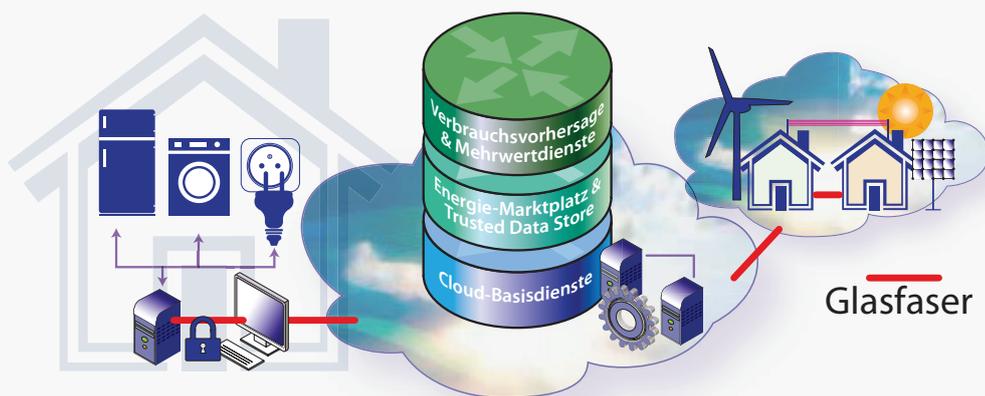


Abbildung 1: Anwendungsszenario Smart Micro Grid mit Cloud-basiertem Datenmanagement.

Die Einbindung lokaler Sensorik und Aktuatorik in intelligenten digitalen Haushalten (Smart Homes) über eine gesondert gesicherte Glasfaserleitung ermöglicht dabei die Verarbeitung von Nutzungsdaten in Echtzeit, beispielsweise zu Prognosezwecken. Werden ausgewählte Daten vom Verbraucher freiwillig für eine weitere Verwendung freigegeben, können dem Nutzer zusätzliche energiebezogene Mehrwertdienste (zum Beispiel Energie-Auditing, Objektschutz oder Gebäudeüberwachung) durch Drittanbieter angeboten werden (s. Abbildung 1). Als konkreter Anwendungsfall wird im Projekt ein Modellgebiet in Saarlouis betrachtet, das aus ca. 100 Wohneinheiten und mehreren Photovoltaik-Anlagen besteht. Die Teilnehmer haben den Vorteil, dass sie ihren Strombedarf günstiger und flexibler auf einem Marktplatz decken können.

innerhalb eines Subnetzes und führt so zu einer Entlastung der übergeordneten Netzebenen, indem Lastschwankungen ausgeglichen oder schon im Vorfeld vermieden werden. Der Marktplatz wird auf einer Cloud-Plattform im Internet realisiert, die sowohl private (zum Beispiel bei den Stadtwerken Saarlouis angesiedelte) als auch öffentlich verfügbare Ressourcen verwendet. Die dort angebotenen Dienstleistungen sind durch die dynamischen Eigenschaften der Cloud besonders kostengünstig und skalieren mit der Nachfrage. Insbesondere große Smart Grids, in denen viele Haushalte Sensordaten liefern, profitieren davon, denn dort sind entsprechend große Mengen an Ressourcen zum Erbringen der Dienstleistungen erforderlich. Die Daten werden dabei in besonders leistungsstarken und sicheren Datenspeichern abgelegt.

Das KIT verfolgt im Rahmen eines Teilvorhabens zwei Hauptaufgaben:

Die wesentlichen Innovationen und Arbeitsschwerpunkte des Projekts sind (s. Abbildung 2):

- Realisierung einer Kern-Infrastruktur bzw. Ablaufumgebung
- Ein offener Marktplatz mit dynamischem Multi-Agenten-

Handel und Mehrwertdiensten

- Echtzeitfähige, skalierbare Massendatenspeicher mit Zugangskontrolle und Nutzungsprofilen sowie
- sichere Sensorik und Aktuatorik in der Cloud.

Datenschutz und IT-Sicherheit

Das Anwendungsszenario Smart Grid wie auch die Speicherung von Haushaltsdaten in der Cloud stellen besondere Herausforderungen an die Sicherheit der Hard- und Software. Über Sensoren für den Energieverbrauch können unter Umständen bereits die in einem Haushalt eingeschalteten Fernsehprogramme erkannt werden. Vor diesem Hintergrund müssen spezielle Maßnahmen im Hinblick auf Datenschutz und IT-Sicherheit getroffen werden. Ein besonderer Fokus des Projekts liegt daher auf dem Datenschutz: Über Datenschutzprofile, Datennutzungskontrollmechanismen, Verschlüsselung und Anonymisierung werden die persönlichen Daten der Teilnehmer vor unberechtigten Zugriffen geschützt.

In diesem Zusammenhang werden in der aktuell laufenden Projektphase die Anforderungen eines auf Cloud-Technologien basierenden Marktplatzes für Mehrwertdienste, der in einem Smart Grid Anwendung findet, analysiert. Insbesondere werden die Aspekte der Sicherheit und des Datenschutzes im Hinblick auf die Erhebung von Sensordaten in realen Umgebungen und deren Verarbeitung in Cloud-Diensten untersucht, womit wichtige Grundlagen für einen Trusted Smart Grid Datastore erarbeitet werden.

Da es sich bei den erhobenen Energieverbrauchsdaten um äußerst sensible Daten handelt, wird ein entsprechend umsichtiger Umgang mit diesen Daten bei allen Beteiligten zu gewährleisten sein. Hierbei stellt sich insbesondere das Problem des Vertrauens, denn Benutzer haben nicht notwendigerweise Vertrauen in den korrekten Umgang mit ihren Daten durch den Marktplatz und/oder die Basis- und Mehrwertdienste, vor allem wenn datenschutzrechtlich brisante Finanzdaten von Marktplätzen oder persönliche Messwerte von Sensoren aus dem privaten Umfeld des Nutzers in Cloud-Umgebungen gespeichert werden. Das Vertrauen der Benutzer zu Markt-

plätzen und sozialen Netzwerken, die auf Cloud-Technologien basieren, ist daher noch geringer als dies bei klassischen Systemen der Fall ist. Bei Sensordaten kommt zusätzlich die Sorge der Nutzer hinzu, nicht zu wissen, welche Daten in der Cloud gespeichert werden.

Die Sicherheits- und Vertrauensverhältnisse der in der Cloud abgelegten Benutzerdaten sollen bei der Peer Energy Cloud durch neuartige Verfahren zur Verteilung, Speicherung, Zugriffsrechteverteilung und Verschlüsselung der Daten geklärt werden. Die zentrale Fragestellung, die das KIT in diesem Zusammenhang bearbeitet, lautet: Welche neuartigen Verfahren zur Zugangskontrolle können genutzt werden und wie können Daten so in der Cloud abgelegt werden, dass zum einen Mehrwertdienste effizient erbracht werden können, zum anderen aber die Daten so verschlüsselt und verteilt gespeichert werden, dass ein einzelner Dienstanbieter keinen Missbrauch betreiben kann? Mithilfe sogenannter Privacy-Profile soll es jedem Benutzer individuell ermöglicht werden, die konkrete Verwendung und Nutzung seiner Daten festzulegen und diese gegebenenfalls zu einem späteren Zeitpunkt zu ändern. So soll es beispielsweise möglich sein, die Weitergabe und Nutzung persönlicher Daten exakt zu bestimmen.

Das langfristige und über dieses Projekt hinausreichende Ziel liegt dabei im Abbau von Nutzer- und Nutzungsvorbehalten gegenüber der Datenspeicherung und Ausführung von Geschäftsprozessen jeglicher Art in der Cloud.

Marktplatz, Basisdienste und Cloud-Plattform

Die Ziele des zu entwickelnden virtuellen Marktplatzes sind:

- Einfache Handhabung aus Nutzersicht („Usability“)
- Flexibilität und Offenheit in der Unterstützung von Abrechnungsmodellen und unterschiedlichen Erzeugerquellen sowie
- Unterstützung einer unabhängigen Stelle zur Garantie von Vertrauenswürdigkeit.

Im Rahmen der Arbeiten zum Marktplatz und zu den Basis-



Abbildung 2: Zentrale Schwerpunkte des Projekts.

diensten liegt daher ein großer Schwerpunkt auf der Erforschung und teilweisen Realisierung von Benutzerschnittstellen sowie der Erstellung von Cloud-Basisdiensten, die den Markt vertrauenswürdig, skalierbar und leistungsfähig machen. Dazu gehören Verfahren der Datenverteilung, Kryptographie und die Zugangskontrolle, aber auch die automatisierte Kopplung von replikationsfähigen Datenspeichern und Portalen hinter Lastverteilern sowie deren Konfiguration.

Ein hohes Maß an Sicherheit für die Daten des Projekts garantiert der Trusted Smart Grid Datastore, der die Skalierbarkeit unterstützt und die Datenverteilung nach den beteiligten Akteuren dezentral organisieren soll. Basis dieses Datenspeichers wird eine Cloud-typische NoSQL-Datenbank, wie beispielsweise Cassandra, sein. Ein weiterer Schwerpunkt liegt auf der Anbindung verschiedener anwendungsfallabhängiger Dienste. Diese können über den Datenspeicher auf Informationen externer Sensoren, beispielsweise aus dem Lebensumfeld der Endanwender, zugreifen. Auf dieser Basis kann auch der agentenbasierte vollautomatische Handel mit Energiekontingenten realisiert werden. Drittanbieter können in diesem Umfeld innovative Dienste entwickeln und in den Markt einbringen, wo sie unter Nutzung der Cloud-Basisdienste betrieben werden können.

Die Cloud Core-Infrastruktur/Plattform stellt das zentrale Arbeitsgebiet des SCC in der Peer Energy Cloud dar. Sie stellt Soft- und gegebenenfalls Hardwarekomponenten bzw. Ressourcen zur Verfügung, mit deren Hilfe die im Projekt entwickelten Dienste getestet und ausgeführt werden können. Gesamtziel ist die Spezifikation und prototypische Realisierung einer Ablaufumgebung, die – losgelöst von spezifischen Virtualisierungsinfrastrukturen wie VMware vSphere oder Amazon EC2 – eine Basis für den Betrieb virtueller Maschinen und darauf aufbauender Dienste darstellt. Die Innovation besteht vor allem darin, einen dynamischen und flexiblen Dienst in der Cloud zu etablieren, der von allen Partnern und Diensten eigenständig genutzt und kontrolliert werden kann. Auf diesem „Core Service“ werden die benötigten Komponenten der Ge-

schäftsprozesse transparent und automatisierbar orchestriert, überwacht, skaliert und abgerechnet (s. Abbildung 3).

Über der Infrastrukturebene wird dafür zusätzlich eine Plattformebene etabliert, auf der die Anwendungen eingerichtet werden. Von der Nutzerseite besteht Zugang über standardisierte interoperable Web-Protokolle, wodurch die Nutzung aller Dienste prinzipiell mit jeder Art von Client auf jeder Art von Plattform möglich wird. Als Basistechnologie kommen dabei auf allen Ebenen virtuelle Maschinen zum Einsatz. Dadurch ist der vorgestellte Cloud-Ansatz durch die Isolation der virtuellen Maschinen voneinander prinzipiell mehrmandantenfähig. Des Weiteren realisiert die Cloud Core-Plattform Monitoring-Komponenten, die den Ressourcenverbrauch detailliert ermitteln bzw. verwalten und so zur Erkennung von Lastspitzen, Ressourcenengpässen und deren Vorhersage sowie zur Rechnungsstellung genutzt werden können.

Mehrwertdienste

Im Rahmen dieses Arbeitspakets entstehen die zentralen wissenschaftlichen und technologischen Grundlagen für die Entwicklung Cloud-basierter Prognoseverfahren, der Verfahren zum Handel von Kontingenten, sowie der Etablierung lokaler Cloud-basierter Peer-2-Peer Netzwerke zum Ressourcenausgleich. Die Arbeiten umfassen:

- Die Anwendung maschineller Lernverfahren auf diesen Lastprofilen zur Erstellung von Verbrauchs- bzw. Erzeugungsprognosen
- Die Entwicklung und Realisierung eines dynamisch orchestrierten Agentensystems für den Handel von Ressourcenkontingenten innerhalb des Micro Grids
- Die Entwicklung von Anreizmodellen zur Nutzung Cloud-basierter Dienste
- Die Realisierung von Privacy-Profilen, die die Vertrauenswürdigkeit der Cloud-basierten Dienste stärken und sicherstellen.

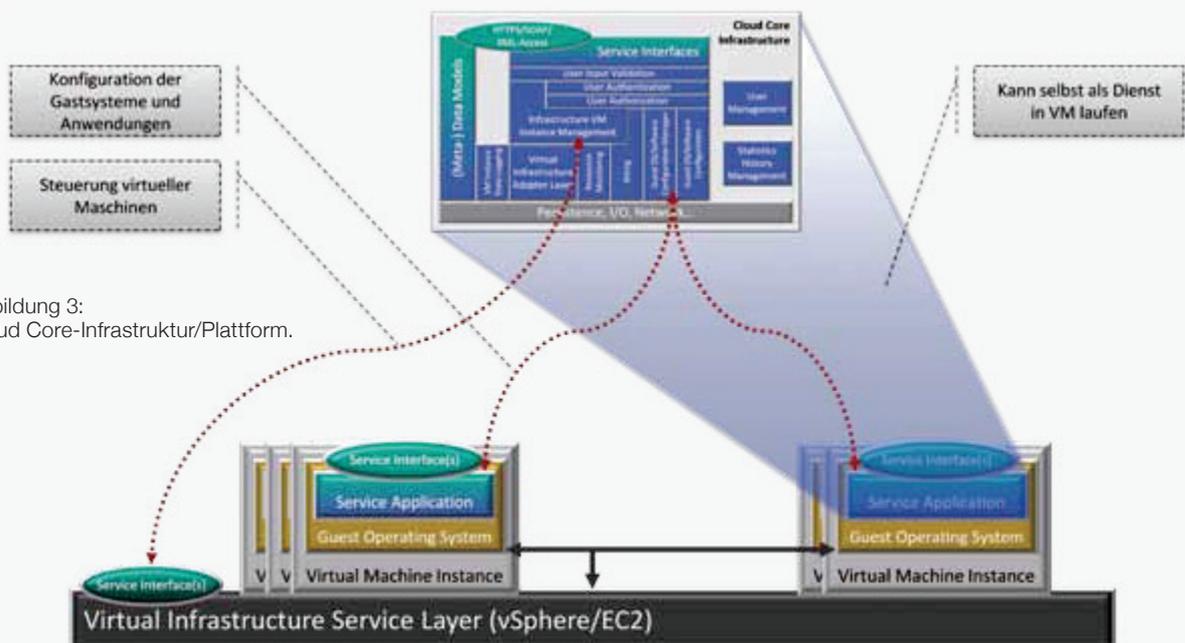


Abbildung 3: Cloud Core-Infrastruktur/Plattform.

Mit den entwickelten Diensten werden verschiedene strategische Ziele verfolgt. Zum einen soll durch das entwickelte Verfahren zur Prognose und zum Handel von Kontingenten die automatische Lastflussoptimierung innerhalb des Micro Grids ohne Eingriffe der Netzbetreiber verfolgt werden, damit ein kostspieliger Ausbau des 0,4 KV-Netzes umgangen werden kann. Da die urbane Netzstruktur des Modellgebiets in Saarlouis optimal ausgenutzt wird, wird die Einspeisung in überregionale Netze vermieden und dient somit auch deren Entlastung. Sollten sich die entwickelten Verfahren, Methoden und Dienste für lokale Netze bewähren, ist langfristig eine Übertragung und Adaption auf große Energienetze und andere Micro Grids denkbar. Zum anderen wird mit der Etablierung und Akzeptanz einer Cloud-basierten Handelsplattform für KMUs die Möglichkeit geschaffen, neuartige Geschäftsfelder zu erschließen. Beispielhaft seien an dieser Stelle Rechnungsstellung, Energieberatung und Sicherheitsdienste genannt, die sich unter dem Begriff Urban Management zusammenfassen lassen.

Sensoren, Aktuatoren und Feldversuche

Das SCC beteiligt sich umfassend am Aufbau des Systemdemonstrators im Micro Grid der Stadtwerke Saarlouis, insbesondere an der Integration der dort angesiedelten auf VMware vSphere aufbauenden Private Cloud. Im geplanten Modellversuch sollen die entwickelten Methoden, Algorithmen und Verfahren in einer realen Umgebung unter Echtzeitbedingungen getestet und erprobt werden. Neben der Installation von intelligenten Stromzählern, Sensorik und Aktuatorik in den einzelnen Haushalten sowie deren Anbindung an die verschiedenen Dienste umfassen die Arbeiten die Bereitstellung aller Komponenten des entwickelten Demonstrators in der Public Cloud.

Fazit

Die im Projekt schwerpunktmäßig vom SCC konzipierte Cloud-Infrastruktur/Plattform sieht vor, dass verschiedene Mehrwertdienste unter Wahrung der Sicherheit und des Datenschutzes in einem offenen Markt integriert werden können, wie zum Beispiel skalierbare Datenspeicher, Rechenressourcen für Prognoseverfahren oder agentenbasierte P2P-Systeme. Dies kann vor allem auch im Bereich der wissenschaftlichen Anwendungen interessant sein, da auf der Basis der im Projekt entwickelten Technologien bisher unabhängig betriebene Softwareprodukte und Anwendungen in der neuen Landschaft zu Systemen gekoppelt werden können. Darüber hinaus kann die neue Umgebung auch wissenschaftlichen Instituten mittelfristig ermöglichen, selbst entwickelte Anwendungen als SaaS-Angebote im Internet anzubieten und zu vermarkten.

Die im Projekt entwickelte Cloud-Plattform kann darüber hinaus auch eine exzellente Grundlage zur Versorgung der IT-Landschaften mit allen Arten von Cloud-Diensten bilden, sowohl im Bereich der wissenschaftlichen Entwicklungsumgebungen (beispielsweise monetär abrechenbare ad-hoc-IT-Ressourcen in Selbstbedienung) als auch bei der

Versorgung mit Basisdiensten. Insbesondere die Bearbeitung von Forschungs- und Entwicklungsfragen im Bereich von Hybrid-Clouds kann zu großen Synergieeffekten führen, da die Ressourcen einfacher als bisher instituts- oder standortübergreifend genutzt werden können.

Dr. Marcel Kunze, Dr. Matthias Bonn

Peer Energy Cloud – Intelligent use of renewable energies

Being one of the 12 winners of the Trusted Cloud Initiative funded by the Federal Ministry of Economics and Technology, the project aims at the development of secure cloud-based solutions for the intelligent use of renewable energies. Central objectives are to organize the energy market more efficiently and dynamically and a trustworthy data management, e. g. of private energy usage profiles. In order to meet these challenges, Peer Energy Cloud addresses the development of innovative energy management methods. In this context SCC mainly focuses on the cloud core platform, which delivers software components to run all the services being developed in the project. The central goal is the specification and prototyped realization of an environment serving as a basis to run virtual machines and services built on top of these machines. Unlike pure virtual environments, the project establishes a flexible and dynamic service which can be used and controlled from all associated partners. All necessary components and business processes – as storage or portal services – may be orchestrated, monitored, scaled and billed automatically.

HadoOptimizer – Lösung komplexer Optimierungsprobleme

In der parametrischen Optimierung wird versucht, eine möglichst gute Lösung eines Optimierungsproblems zu finden, indem ausgewählte Parameter eines Problems gezielt variiert werden. Der HadoOptimizer ist ein Framework, mit dem sich Evolutionäre Algorithmen unter Zuhilfenahme des Cloud Computing berechnen und die parallele Nutzung vieler Recheneinheiten massiv beschleunigen lassen. Der HadoOptimizer steht als Open Source unter der Apache-Lizenz Version 2.0 frei zur Verfügung und wurde am SCC von Christian Kumpke im Rahmen seiner Diplomarbeit entwickelt.

Der lateinische Begriff „Optimum“ bedeutet wörtlich übersetzt „das Beste“. Um, ausgehend von einem Startwert, jeweils bessere Lösungen iterativ identifizieren zu können, ist die wichtigste Komponente jeder Optimierungsrechnung die Zuordnung von einem oder mehreren Bewertungen zu einem gegebenen Systemzustand. Ein Systemzustand wird hierbei durch eine Zahl an Variablen oder Parameter beschrieben, die sich hinsichtlich ihrer Typen (z.B. boolesche Werte oder reelle Zahlen), den erlaubten Wertebereichen und natürlich ihrer Interpretation unterscheiden. Handelt es sich um nur ein einzelnes, numerisches Qualitätskriterium, so werden verschiedene Systemzustände miteinander vergleichbar, man spricht von einer Metrik.

Als Beispiele für eine solche Metrik kann etwa der Luftwiderstand eines Fahrzeugs oder der Emissionsausstoß eines Motors gelten. Diese werden von verschiedenen Geometrieparametern, aber auch Variablen wie dem Zündzeitpunkt oder dem Einspritzdruck abhängen. Einsatzfelder der parametrischen Optimierung lassen sich in allen Bereichen der Technik und Wirtschaft finden.

Für viele Problemstellungen im Feld der parametrischen Optimierung lässt sich „die beste“ Lösung jedoch nicht effizient

berechnen: so ergeben sich bereits bei einem Problem mit nur 10 Parametern, die jeweils einen von 10 Werten annehmen können, bereits 10^{10} unterschiedliche Lösungsmöglichkeiten. Nimmt man – für komplexe Problemstellungen eher konservativ – 1 Sekunde pro Bewertung an, so würde die Berechnung aller Lösungsmöglichkeiten auf einem einzelnen Prozessorkern rund 317 Jahre in Anspruch nehmen. Selbst mit der Vielzahl der beispielsweise am Grid Computing Centre Karlsruhe (GridKa) verfügbaren Ressourcen – aktuell arbeiten dort rund 12.000 Rechenkerne parallel an der Lösung wissenschaftlicher Problemstellungen – würde die Berechnung immer noch rund 10 Tage dauern. Daher wird in vielen Fällen auf Näherungsverfahren zurückgegriffen, die es erlauben, unter vertretbarem Einsatz von Ressourcen eine ausreichend „gute“ Lösung zu finden. Die Klasse der Evolutionären Algorithmen ist ein oft eingesetzter Vertreter solcher Näherungsverfahren.

Evolutionäre Algorithmen

Evolutionären Algorithmen orientieren sich am Vorbild der biologischen Evolution und folgen einem Zyklus aus Duplikation/Rekombination, Mutation und Selektion. Mögliche Lösungen werden hierbei iterativ zuerst vervielfältigt, dann zufällig geringfügig variiert (biologisches Vorbild: Mutation)

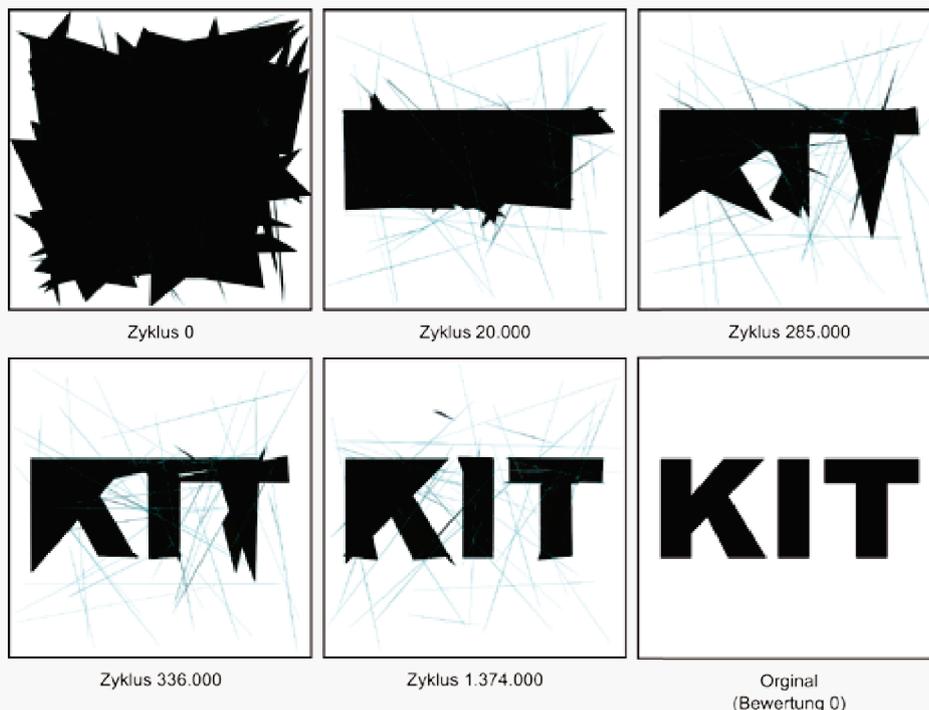


Abbildung 1: Evolutionäre Algorithmen erreichen oft bereits in den ersten Zyklen eine gute Approximation an das Optimum. Gezeigt ist der Versuch, mit Hilfe des HadoOptimizers den Schriftzug KIT mit Dreiecken zu approximieren.

und schließlich hinsichtlich ihrer Güte bewertet, wobei schlecht bewertete Lösungen verworfen werden (biologisches Vorbild: Selektion). Jeder Zyklus entspricht einer Generation, innerhalb derer Eltern- und Kind-Individuen identifiziert werden können. Eltern stellen die besten Lösungen der letzten Generation dar, Kinder entstehen aus den Eltern durch die Rekombination oder schlichte Vervielfältigung. Die zwei Hauptvertreter Evolutionärer Algorithmen sind Genetische Algorithmen („GA“) und Evolutionsstrategien („ES“).

Genetische Algorithmen beschreiben Lösungsvorschläge durch Ketten von Einsen und Nullen. Mutationen entstehen, in dem zufällig Einsen in Nullen und Nullen in Einsen umgewandelt werden. Daneben kann es wie bei Chromosomen in der Natur zum „Cross-Over“ kommen. Hierbei werden jeweils Teile der Ketten verschiedener Elternteile zu einem neuen Lösungsvorschlag kombiniert.

Anders als Genetische Algorithmen verwenden Evolutionsstrategien Ketten reeller Zahlen für die Parameterbeschreibung. Auch hier ist ein Cross-Over möglich. Häufiger werden Kinder jedoch einfach mutiert, in dem auf jede reelle Zahl eine Zufallszahl aufaddiert wird. Dabei sind diese Zufallszahlen so beschaffen, dass kleine Zahlen häufiger auftreten als große Zahlen. So werden kleine Änderungen bevorzugt, mit einer niedrigeren Wahrscheinlichkeit kann eine Evolutionsstrategie aber auch große Sprünge im Suchraum vollziehen.

Sowohl Genetische Algorithmen als auch Evolutionsstrategien haben eine lange Geschichte, die bis in die 1960er Jahre zurückreicht. Teilweise wurden hier sogar Optimierungen mit einfachen Holzmodellen durchgeführt, etwa durch Variation eines Flügelprofils an einem realen Modell.

Evolutionäre Algorithmen erfahren jedoch heute eine Renaissance, da sie viele Vorteile zum Beispiel hinsichtlich ihrer Robustheit und der Ausnutzung der vorhandenen Infrastruktur aufweisen. Häufig beobachtet man bei Optimierungen mit Evolutionären Algorithmen auch eine erhebliche Verbesserung der Zielfunktion in den ersten Optimierungszyklen. Abbildung 1 demonstriert dies an einem Beispiel, bei dem die 6 Koordinaten einer Reihe schwarzer Dreiecke so verändert werden mussten, dass deren Überlagerung dem Schriftzug „KIT“ möglichst ähnlich wurde. Das Beispiel wurde mit dem HadoOptimizer berechnet.

Ist es für ein Optimierungsproblem wichtig, möglichst nahe an die theoretisch beste Lösung heranzukommen, so bedeutet dies umgekehrt jedoch auch, dass teilweise sehr viele Bewertungszyklen durchlaufen werden müssen. Aufgrund der hohen Anzahl dann zu bewertender alternativer Lösungen können Evolutionäre Algorithmen auch besonders rechenaufwändig sein.

Cloud Computing und Hadoop

Praktischerweise können aber viele Rechenschritte (insbesondere die Mutation und die Bewertung) in Evolutionären Algorithmen unabhängig voneinander durchgeführt werden. Daher kann dieses Verfahren besonders von Ansätzen wie dem Cloud Computing profitieren, die die massive Parallelisierung von Berechnungen ermöglichen.

Beim Cloud Computing werden IT-Dienstleistungen wie die Berechnung von Diensten oder Speicherung von Daten meist kurzfristig und automatisiert von einem externen Dienstleister zu bezogen. Der externe Dienstleister wird oft als Cloud Provider bezeichnet, während die Cloud das betriebene Rechenzentrum inklusive der angebotenen Dienstleistungen beschreibt. Aufgrund des großen öffentlichen Interesses am Cloud Computing ist eine Abgrenzung gegenüber anderen Ansätzen verteilten Rechnens nicht immer leicht zu ziehen. Das amerikanische National Institute of Standards and Technology (NIST) nennt hierfür als essentielle Charakteristiken insbesondere folgende Punkte [1]:

- Der Anwender kann Ressourcen wie Speicher oder Rechenzeit automatisiert anfordern, ohne dass Interaktionen mit Mitarbeitern des Betreibers der Cloud erforderlich wären.
- Der Cloud Provider bietet seine Dienstleistungen einer großen Anzahl von Kunden gleichzeitig an. Die für die unterschiedlichen Kunden auszuführenden Rechenaufgaben beeinflussen sich nicht und sind durch geeignete technische Maßnahmen (insbesondere Virtualisierungsmechanismen) voneinander getrennt.
- Zusätzliche Rechenressourcen können durch den Kunden sehr schnell angefordert und wieder freigegeben werden („Elastizität“). Oft erscheinen die verfügbaren Rechenkapazitäten aus Sicht des Kunden praktisch unbeschränkt.

Hadoop¹ ist ein oft im Bereich des Cloud Computing eingesetztes Software Framework, das die Implementierung verteilter Software erleichtert. Hadoop beinhaltet unter anderem Komponenten zur verteilten Speicherung von Daten, zur Datenanalyse und zur Überwachung verteilter Systeme. Kernbestandteil von Hadoop ist eine Implementierung des von Google eingeführten MapReduce-Algorithmus [2]. MapReduce ermöglicht es, nebenläufige Berechnungen auf zwei Phasen (die Map- und die Reduce-Phase) aufzuteilen. Während jeder der beiden Phasen wird der entsprechende Teil der zu berechnenden Aufgabe auf einer hohen Anzahl von Rechnern (den so genannten Mappern bzw. Reducern) parallel ausgeführt, wodurch eine hohe Beschleunigung gegenüber einer sequentiellen Berechnung erreicht werden kann, die im Wesentlichen durch die Parallelisierbarkeit der Aufgabenstellung und der Anzahl der verfügbaren Rechenknoten beschränkt wird. Allerdings muss in der Implementierungsphase die jeweilige Aufgabenstellung zunächst auf das zweiphasige MapReduce-Modell abgebildet werden. Dies bringt nicht nur einen hohen zeitlichen Aufwand mit sich, sondern kann auch eine vermeidbare Fehlerquelle darstellen. Das im Rahmen der Diplomarbeit von Christian Kumppe am SCC entwickelte Framework HadoOptimizer löst diese Aufgabe mit Hilfe von Hadoop für viele Spielarten Evolutionärer Algorithmen und hilft dem Nutzer so, seine Optimierungsrechnungen massiv zu beschleunigen.

HadoOptimizer

Das Ziel bei der Entwicklung des HadoOptimizers [3] war es, Evolutionäre Algorithmen unter möglichst geringem Aufwand für einen Einsatz auf einem Hadoop-Cluster anzupassen und schlussendlich ausführen zu können. Der in Java geschriebene

¹<http://hadoop.apache.org/>

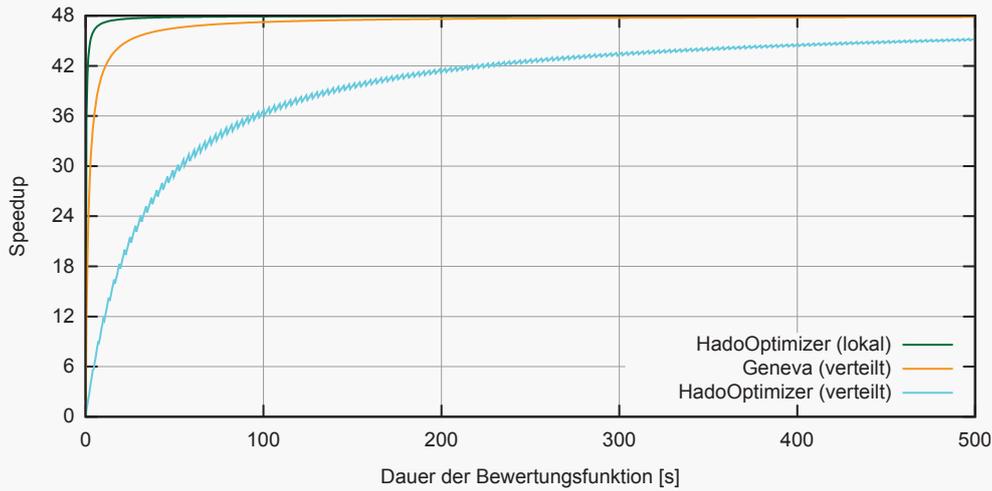


Abbildung 2: Je nach der Dauer jeder Einzelbewertung arbeitet der HadoopOptimizer unterschiedlich effizient. Weniger allgemeine, dafür jedoch effizientere Optimierungsumgebungen wie die Geneva-Bibliothek können eine höhere Performance erreichen.

HadoopOptimizer erreicht dies, indem er dem Anwender eine für die Definition von Evolutionären Algorithmen optimierte Schnittstelle anbietet, Interna des Hadoop-Frameworks aber vor dem Anwender verbirgt. So ist es für den Nutzer insbesondere nicht erforderlich, sich mit der Funktionsweise des MapReduce-Algorithmus auseinanderzusetzen oder auch nur zu wissen, dass schlussendlich der MapReduce-Algorithmus zur Parallelisierung eingesetzt wird. Dies kann den Einsatz des HadoopOptimizers für den Anwender beträchtlich erleichtern. Weiterhin ermöglicht es der HadoopOptimizer, Evolutionäre Algorithmen auch auf öffentlichen Cloud-Plattformen (insbesondere Amazon Elastic MapReduce²) auszuführen, wodurch eine praktisch unbeschränkt Anzahl von Rechenknoten für die Berechnung zur Verfügung steht. Hierbei ist allerdings zu beachten, dass die Anzahl der sinnvoll einsetzbaren Rechenknoten durch die Anzahl der Individuen einer Generation begrenzt wird.

Diesen Vorteilen steht ein erhöhter Overhead bei der Berechnung einzelner Rechenschritte und der Kommunikation innerhalb des Hadoop-Frameworks gegenüber. Da es sich bei Hadoop um ein allgemeines MapReduce-Framework handelt, stehen keine speziellen Möglichkeiten zur Optimierung hinsichtlich der Anforderungen Evolutionärer Algorithmen zur Verfügung. Abhängig vom jeweiligen Optimierungsproblem bietet der HadoopOptimizer daher im Vergleich zu anderen, höher optimierten Optimierungsumgebungen bei gleicher Anzahl eingesetzter Rechenknoten eine geringere Leistung. Im Folgenden wird ein Vergleich mit der Optimierungsumgebung Geneva vorgestellt.

Geneva und Performancevergleich

Geneva (Grid-enabled evolutionary algorithms) ist eine Optimierungsumgebung der KIT-Ausgründung Gemfony scientific [4]. Geneva implementiert neben Evolutionären Algorithmen auch Schwarmalgorithmen, Simulated Annealing und Gradientenverfahren. Wie der HadoopOptimizer steht Geneva unter einer Open Source Lizenz (der Affero GPL v3) und ist damit in der Forschung barrierefrei einsetzbar. Geneva erlaubt neben Mehrkernsystemen die Verwendung von Clustern, Grids und Clouds.

Abbildung 2 zeigt die Leistungsunterschiede zwischen dem HadoopOptimizer und Geneva. Zum Vergleich beider Umgebungen diente dabei ein konstruiertes Optimierungsproblem, bei dem die Bewertung einzelner Kandidatenlösungen ersetzt wurde durch eine Wartezeit konfigurierbarer Länge. Dieser Trick macht die ermittelte Berechnungsdauer weitestgehend unabhängig von der verwendeten Recheneinheit. Hierdurch lässt sich besonders gut der von Hadoop und dem HadoopOptimizer eingeführte Overhead bewerten und mit anderen Umgebungen vergleichen. So wird der ermittelte Overhead insbesondere durch die Kommunikation zwischen den Knoten, aber auch Aspekten wie der Serialisierung von Objekten oder der verwendeten Programmiersprache erzeugt. Wie die Grafik zeigt, erreicht der HadoopOptimizer für Problemstellungen mit lang andauernder Bewertungsfunktion einen höheren Speedup, der sich immer mehr dem maximalen Wert annähert. Im Vergleich nähert sich die in C++ implementierte Geneva-Bibliothek deutlich schneller dem Optimum an. Ein Grund hierfür liegt in dem Overhead, den Hadoop beim Kopieren vor jeder Bewertungsfunktion erzeugt: ist die Bewertungsfunktion rechenaufwändiger, so fällt dieser Overhead weniger stark ins Gewicht. Die Geneva-Bibliothek führt die Kommunikation

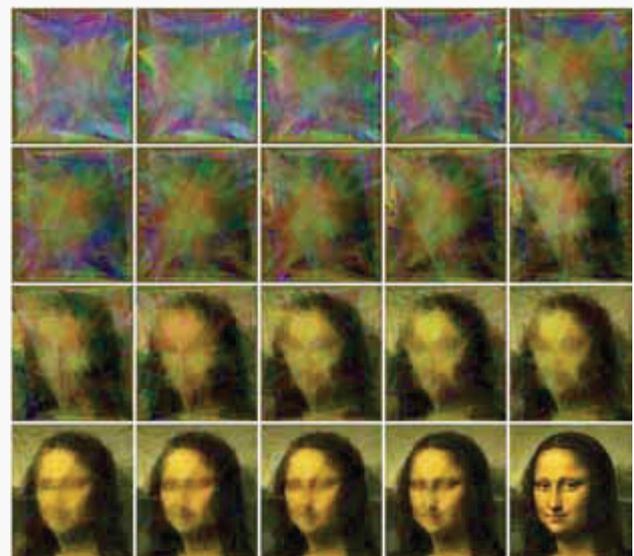


Abbildung 3: Ein Bild der Mona Lisa wird mit Hilfe der Geneva-Bibliothek approximiert, indem die Parameter von 300 semitransparenten, farbigen und einander teilweise überlappenden Dreiecken so verändert werden, dass die Unterschiede zum Zielbild minimal werden.

² <http://aws.amazon.com/de/elasticmapreduce/>