



Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

NACHRICHTEN



Jahrgang 36 • 3/2004

Grid-Computing

**Im Inhaltsverzeichnis verwendete
Abkürzungen:**

| | |
|-----|--|
| IAI | Institut für Angewandte Informatik |
| IPE | Institut für Prozess- datenverarbeitung und Elektronik |
| IWR | Institut für Wiissenschaftliches Rechnen |

Herausgeber:

Forschungszentrum Karlsruhe GmbH
Postfach 3640 · 76021 Karlsruhe
Telefon-Nr. 07247 82-0

Redaktion:

Dr. Joachim Hoffmann

Redaktionsbeirat:

Prof. Dr. J. Blümer, IK-IEKP
Dr. W. Breitung, IKET
Prof. Dr. E. Dinjus, ITC-CPV
Dr. K. Gompper, INE
Dr. J. Gspann, IMT
Dr. G. Metzsig, FTU
Dr. W. Pfleging, IMF-I
Prof. Dr. U. Schurath, IMK (Vorsitzender)
Dr. K.-F. Weibezahn, IMB

Grafik und Satz:

Stolz Grafisches Atelier · Karlsruhe

Layout:

Tassilo Schnitzer

Druck:

Baur GmbH · Kellern

Nachdruck mit Genehmigung des
Forschungszentrums Karlsruhe GmbH
unter Nennung der Gesellschaft und
des Autors gestattet. Beleg erbeten.

Die NACHRICHTEN
können kostenlos über die
Hauptabteilung Bibliothek und Medien
des Forschungszentrums bezogen
werden.

Printed in the Federal Republic of
Germany

Gedruckt auf chlorfrei gebleichtem
Papier



Die Vision des Grid-Computing:
Daten und Rechenleistung können
analog zum elektrischen Strom jeder-
zeit und überall auf der Welt nach Be-
darf „aus der Steckdose“ bezogen
werden.

Inhalt / Contents

- **Grid-Computing – Basis für institutsübergreifende virtuelle Organisationen** 133
Grid Computing – Basis of Multi-Institutional Virtual Organizations

H. Marten, IWR

A Grid is a worldwide communication infrastructure that allows seamless transparent access to instruments, data and computing power on demand. Numerous projects all over the world are developing software standards and technologies for this World Wide Grid. The article gives an introduction into the grid vision and outlines some of the international projects that are relevant to the work at Forschungszentrum Karlsruhe in more detail.

- **GridKa: Das Grid Computing Zentrum Karlsruhe für Anwendungen aus der Hochenergiephysik** 138
Building a Cluster Computer for the Computing Grid of Tomorrow

J. van Wezel, H. Marten, IWR

The Grid Computing Centre Karlsruhe takes part in the development, test and deployment of hardware and cluster infrastructure, grid computing middleware, and applications for particle physics. The construction of a large cluster computer with thousands of nodes and several PB data storage capacity is a major task and focus of research. CERN based accelerator experiments will use GridKa, one of only 8 world wide Tier-1 computing centers, for its huge compute demands. Computing and storage is provided already for several other running physics experiments on the exponentially expanding cluster.

- **Aufbau des großen GridKa-Clusters am Forschungszentrum Karlsruhe** 143
Construction of the Large GridKa Cluster at the Forschungszentrum Karlsruhe

M. Alef, B. Hoefft, H. Marten, J. van Wezel, IWR

Particle physicists from all over the world try hard to have the results of their measurements analysed by an extremely efficient and fast Data-Grid. The enormous amounts of data are the biggest challenge for the computer centres involved. The Grid Computing Centre Karlsruhe (GridKa) has successfully begun to build up a Grid-Computing Centre that is able to meet these expectations. In several areas like network connection, data storage, running and cooling of a computer cluster GridKa is one of the world's leading data processing centres.

■ **Unterstützung der Grid-Benutzer: Global Grid User Support (GGUS) – weltweit rund um die Uhr** 148

Global Grid User Support (GGUS) – Worldwide Around the Clock

H. Dres, G. Grein, R. Kupsch, R. Pietschmann, W. Thöne, IWR

Building up a World Wide Grid stands not only for providing a huge amount of computer power, data storage, and network capacity, this also means to establish a global grid user support – accessible by every user from any point on the globe at any time he wants. This paper describes the development of the Global Grid User Support, which has been promoted at GridKa since 2003, its concepts and the implementation, current status and future aspects.

■ **Verschlüsselung und Authentifizierung als Basis internationalen Vertrauens** 153

Encryption and Authentication as a Basis of International Trust

U. Epting, IWR

The idea of building a World Wide Grid is a particularly big challenge for security. The Globus Software Toolkit is widely used in international grid-projects. It provides with the Globus Security Infrastructure (GSI) reliable features based on international standards for encryption and secure authentication. Public key cryptography, X.509-certificates and the secure socket layer protocol (SSL) also known as transport layer protocol (TLS) are used to identify users and machines across specially built international trust domains. The Institute for Scientific Computing at FZK runs a certification authority – GridKa-CA – to provide certificates for scientists located in Germany. Some mathematical fundamentals as well as organizational and technical issues are presented in this article.

■ **Verteilte Simulationen auf einem Grid** 158

Distributed Simulation on a Grid

K.-U. Stucky, S. Halstenberg, W. Jakob, A. Quinte, W. Süß, IAI

Research and development increasingly leverage computer simulation of complex systems to describe their static and dynamic properties based on appropriate models. These calculations often require large amounts of computing power, data transfer, and storage capacity. Grids and their middleware infrastructure are capable to meet these requirements by providing access to widely distributed resources. We present a concept for a simulation grid and a resource broker with global optimisation based on evolutionary algorithms. The Virtual Eye, a system for planning in cornea surgery, is currently redesigned as a reference application on a grid. Implementation details for performing parameter studies are given.

■ Grid-Computing in Echtzeit

163

Realtime Grid Computing

W. Eppler, T. Müller, V. Hartmann, S. Tchilingarian, A. Augenstein, M. Naji, H. Gemmeke, IPE

The inclusion of realtime requirements into grid specifications is very important to integrate experiments with data acquisition and slow control components into the grid. Grid development and research is application driven. FZK has several applications in Particle Physics, Proteomics, Fusion Technology and Cancer Research that consist of realtime parts. They benefit from grid technologies in several aspects: from the access to temporary big demand of resources, the efficient management of distributed systems, and the urge to overall standardization of interfaces. The present paper shows the work and objectives towards a realtime grid.

■ Das CrossGrid-Projekt

169

The CrossGrid Project

M. Kunze, IWR

There are many large-scale problems which require new approaches to computing, such as earth observation, environmental management, biomedicine, industrial and scientific modelling. The CrossGrid project addresses those problems and is oriented towards specific end-users: Medical doctors, who could obtain new tools for correct diagnoses and to guide them during operations; flood crisis teams, which could predict the risk of a flood on the basis of historical records and actual hydrological and meteorological data; physicists, who could optimise the analysis of massive volumes of data distributed across countries and continents. Portals for specific Grid applications have been designed, which should allow for easy connection to the Grid and create a customised work and development environment.

■ Das Projekt CampusGrid

173

The Project CampusGrid

F. Schmitz, IWR

The central idea of Grid Computing is the virtualization of heterogeneous resources – CPUs, disks and archives as well as support (GGUS: Global Grid User Support) and even experimental facilities. In order to meet this challenge the Institute for Scientific Computing, IWR, has started the project CampusGrid in 2004. Its medium term goal is to provide a seamless IT environment supporting the on-site research activities in Physics, Bioinformatics, Nanotechnology and Meteorology. The environment will include all kinds of HPC resources: vector computers, shared memory SMP servers and clusters of commodity components as well as a shared high-performance storage solution. After introducing the general ideas the report will inform about the current project status and scheduled development tasks. This is associated with reports on other activities in the fields of Grid computing and high performance computing at IWR.

■ **Von der Vision zur Realität: Grid-Computing für kommerzielle Anwendungen**
From Vision to Reality: Grid Computing for Commercial Use

177

K.-P. Mickel, IWR

Grid computing is a new paradigm for the information technology. The well known WWW (World Wide Web) will be succeeded by the upcoming WWG (World Wide Grid). The futurologists are promising that it will be possible to get large IT-resources „from a plug in the wall“ without the necessity to know who provides the resources and where the resources are coming from. Initially grid computing was developed and used only for scientific communities. But nowadays more and more commercial companies are beginning to use grid computing. Their goal is to reduce their IT costs dramatically and to react better and quicker on the strong requirements of the global markets.

Grid-Computing – Basis für institutsübergreifende virtuelle Organisationen

H. Marten, IWR

Einleitung

Die rasante Entwicklung von Wissenschaft und Technik während der letzten Jahrhunderte hat sowohl unser tägliches Leben als auch unsere wissenschaftlichen Arbeitsmethoden stark beeinflusst. Erfindungen wissenschaftlicher Instrumente wie das Mikroskop oder das optische Teleskop im 17. Jahrhundert und deren Weiterentwicklungen haben zu revolutionären und immer detaillierteren Einblicken in die Biologie, Medizin, Astronomie etc. geführt. Heute sind wir in der Lage, mit Rasterelektronenmikroskopen atomare Strukturen aufzulösen, mit Computertomographen dreidimensionale Aufnahmen vom Inneren eines Patienten zu machen, oder Teleskope in den Weltraum zu senden, um ganze Himmelsregionen in mehreren Wellenlängenbereichen gleichzeitig zu beobachten.

Entwicklung und Bau derartiger „Instrumente“ sind heute im Allgemeinen interdisziplinäre Projekte mit Beteiligung vieler internationaler Partner und erfordern einen effektiven Informationsaustausch über ausgereifte Kommunikationsstrukturen. Aber auch die Datenmengen, die diese Geräte im anschließenden Forschungsbetrieb liefern, haben sich in den letzten Jahrzehnten dramatisch entwickelt; in Einzelfällen reden wir heute von hunderten von Terabyte oder gar mehreren Petabyte pro Jahr (1 Petabyte = 1 Million Gigabyte). Riesige Rechenkapazitäten sind notwendig, um diese enormen Datenfluten auszuwerten. Die Vision der Wissenschaftler ist nun, sämtliche dazu not-

wendigen Ressourcen zusammen zu schließen und so eine weltweite Infrastruktur zu schaffen, die über alle Grenzen hinweg den transparenten Zugriff auf Instrumente, Daten und Rechenleistung erlaubt.

Die Vision

Die primären Datenstrukturen der einzelnen Komponenten, also Messdaten, Übertragungsprotokolle, Datenbanken usw., sind durch den jeweiligen Einsatzzweck oder die verwendete Technologie bestimmt, und damit zwangsläufig sehr heterogen. Ferner ist davon auszugehen, dass die technischen Einzelkomponenten wie Satelliten, Netzwerke, Rechencluster etc. auch künftig von unabhängigen Institutionen betrieben werden. Damit wird eine weltweite Infrastruktur nicht nur heterogen, sondern kann sich zudem beliebig dynamisch verändern. So könnte z.B. ein Rechenzentrum seine Kapazitäten heute der Meteorologie und morgen einer ganz anderen

Wissenschaftlergruppe zur Verfügung stellen. Den Endbenutzer interessiert jedoch letztlich nur, dass sein Rechenauftrag ausgeführt wird, nicht aber wo. Und er sollte sich auch nicht mit unterschiedlichen Übertragungsprotokollen oder Datenformaten auseinandersetzen müssen. Gefordert ist also eine Infrastruktur, die Rechenleistung, Software-Produkte, Daten, Sensoren, Messinstrumente usw. in Form kompletter Dienste anbietet, die von jedem Arbeitsplatz der Welt zugänglich sind. Zusätzlich muss es (Software-)Dienste geben, die diese Ressourcen selbstständig verwalten und benutzerfreundlich, sicher und zuverlässig anbieten können (so genannte Resource Broker). Daten und Rechenleistung sollen also für jedermann nach Bedarf „aus der Steckdose“ abrufbar sein, und man bezeichnet diese Infrastruktur in Analogie zum elektrischen Stromnetz (engl.: *power grid*) als Grid [1]. Abb. 1 zeigt das Schema einer solchen dynamischen Grid-Welt.

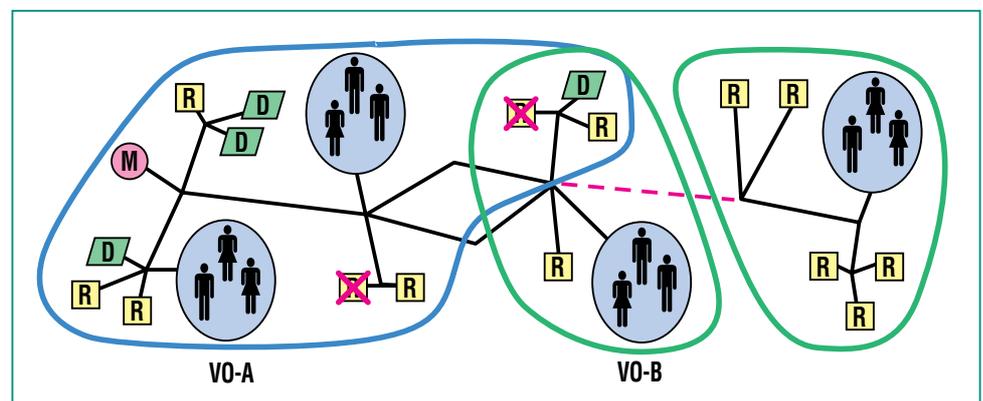


Abb. 1: Benutzergruppen, Rechenzentren (R), Datenzentren (D) und Messeinrichtungen (M) bilden im Grid verwaltungstechnisch virtuelle Organisationen (VOs). Einzelne Zentren können ihre Dienste temporär einer oder mehreren VOs anbieten, und VOs können bei Bedarf weitere Benutzergruppen und Ressourcen aufnehmen oder entkoppeln.

Eine solche Infrastruktur bietet unzählige, qualitativ neue Anwendungsmöglichkeiten. So könnte z.B. ein Arzt das Röntgenbild seines Krebspatienten über einen Grid-Auftrag weltweit mit ähnlichen Bildern anderer Patienten vergleichen lassen, um über die Patientenberichte ähnlich gelagerter Fälle eine optimale Behandlungsmethode zu finden. Oder nehmen wir an, die Analysesoftware eines Wettersatelliten entdeckt einen Hurrikan über dem Pazifik, initiiert eine Grid-Simulation um dessen Geschwindigkeit und Richtung vorherzusagen, und sendet Warnungen an die betroffenen bewohnten Regionen. Zur gleichen Zeit hat eine kleine Firma einen neuen Impfstoff gegen ein Virus entwickelt, und verwendet Grid-Ressourcen, um dessen Einfluss in Kombination mit einigen tausend weiteren Wirkstoffen zu simulieren. Irgendwo anders beginnt aufgrund starker Regenfälle ein Fluss über die Ufer zu treten, und ein Krisenstab startet einen Grid-Auftrag zur Flutsimulation unter Berücksichtigung von aktuellen Messungen der lokalen Wetterstationen um Umkreis von 100 Kilometern...

Offensichtlich bietet das Grid insbesondere immer dann große Vorteile, wenn kurzfristig sehr große Rechenkapazitäten benötigt, aus ökonomischen Gründen aber nicht dauerhaft lokal vorgehalten werden können, oder wenn sehr große Datensätze ausgewertet oder korreliert werden müssen. In der Literatur findet man die Begriffe „Computing Grid“ und „Data Grid“ für diese beiden rechenintensiven

oder datenintensiven „Extremfälle“.

Grid-Bausteine

Eine notwendige Voraussetzung für den Erfolg eines Grid ist die Existenz von Netzwerken mit ausreichend hoher Bandbreite und gegebenenfalls garantierten Service-Qualitäten. In Europa verbindet das gesamteuropäische Netzwerk Géant [2] nationale Forschungsnetze mit Bandbreiten bis zu 10 Gbit/s. Spezielle Knotenpunkte verbinden Géant mit Japan und den USA. In Amerika bildet TeraGrid [3] eine Initiative, fünf Hochleistungsrechenzentren über ein Netzwerk mit 40 Gbit/s zu verbinden. TeraGrid wird mit hoch auflösenden Grafikeinheiten und Werkzeugen für Grid-Computing eine Rechen- und Speicherkapazität aufbauen, die 10.000-15.000 modernen PCs entspricht.

Die Komplexität der physikalischen Infrastruktur sowie einer heterogenen Datenhaltung wird vor dem Benutzer durch eine Virtualisierungssoftware, die so genannte Grid Middleware, versteckt. Was übrig bleibt sind Benutzerschnittstellen ähnlich den Befehlen oder der Oberfläche eines Betriebssystems, die den Zugang zum Grid oder zu einzelnen Diensten von einem stationären Arbeitsplatz oder auch über mobile Geräte wie Handys oder PDAs erlauben. Eines der zur Zeit am häufigsten verwendeten Basis-Pakete für Middleware ist das Globus Toolkit [4]. Es erlaubt, die Identität eines Benutzers zu überprüfen, einzelne Prozesse auszuführen, die zur Verfü-

gung stehenden Grid-Ressourcen zu überwachen und Daten in einer verteilten Umgebung zu lokalisieren bzw. die notwendigen Datenströme zu organisieren. Das Globus Projekt wurde 1995 vom Argonne National Laboratory, der Universität von Süd-Kalifornien und der Universität von Chicago gegründet, und Ende 2003 unter Einbeziehung der Universität von Edinburgh und dem schwedischen Zentrum für Parallelrechner in die Globus Alliance gewandelt.

Globales, virtualisiertes Grid Computing hätte keine Überlebensmöglichkeit ohne klare Definitionen von Software-Schnittstellen und Standards. Das Global Grid Forum, GGF [5], arbeitet als Standardisierungsgremium, ähnlich der Internet Engineering Task Force (IETF) für das World Wide Web. Es wurde gegründet als Zusammenschluss der amerikanischen und europäischen Grid Foren sowie der Interessengemeinschaften im asiatischen Raum unter starker industrieller Beteiligung und organisiert pro Jahr drei große Arbeitstreffen von etwa 50 internationalen Arbeitsgruppen.

Dutzende, wenn nicht sogar hunderte wissenschaftlicher und industrieller Projekte weltweit tragen zur Einführung und weiteren Entwicklung dieser Standards bei. In Europa wurden innerhalb des 5. Rahmenprogramms der EU 24 Grid-Projekte mit insgesamt 58 Millionen Euro gefördert. Zwei dieser Projekte sind für die Arbeiten im Forschungszentrum von besonderer Bedeutung: Das europäische DataGrid Projekt, EDG [6], das im März 2004 ende-

te, hat ein Grid entwickelt und aufgebaut, das Forschungsbereiche mit großen Datenmengen und gleichzeitig großem Rechenzeitbedarf unterstützt. Die Grid-Anwendungen kamen in diesem Fall aus der Hochenergiephysik, der satellitengestützten Erdbeobachtung und der Bioinformatik. Das Projekt CrossGrid [7] hat die Infrastruktur des DataGrid Projektes durch Einbindung weiterer Rechencluster erweitert, und es entwickelt auf der Basis der EDG und Globus Software Werkzeuge und Methoden zur Ausführung von parallelen und interaktiven Anwendungen auf dem Grid.

Elementarteilchenphysik als Triebfeder für Grid-Entwicklungen

Bei genauerer Betrachtung stellt man fest, dass sehr viele weitere Grid-Initiativen von der Teilchen- oder Hochenergiephysik ausgehen. Dies liegt daran, dass die beteiligten Gruppen bereits traditionell in großen internationalen Kollaborationen organisiert sind und mit Teilchenbeschleunigern (CERN, SLAC, FermiLab, DESY, ...) gigantische Datenmengen produzieren, deren Auswertung einen entsprechend hohen Bedarf an Rechenleistung nach sich zieht. Eines der größten Projekte der Hochenergiephysik in den

nächsten Jahren wird der Bau und Betrieb des Large Hadron Collider, LHC [8], am europäischen Teilchenphysiklabor CERN sein. An diesem Ringbeschleuniger mit einem Umfang von 27 Kilometern wird man ab 2007 mit vier riesigen Detektoren die Struktur der Materie anhand der Bruchstücke von Teilchenkollisionen erforschen. Das geplante Datenaufkommen einschließlich Simulationen liegt bei etwa 12 Petabyte pro Jahr und erfordert circa 70.000 der heute schnellsten PCs. Da diese Datenmengen nicht mehr von einem einzelnen Rechenzentrum bewältigt werden können, wurde 2001 das LHC Computing Grid Project,



Abb. 2: Zu jedem Zeitpunkt wird die Verfügbarkeit der im Resource Broker des LCG Projektes eingetragenen Rechencluster überprüft [11].

LCG [9], als fünftes Teilprojekt neben der Entwicklung der vier LHC-Detektoren gegründet. LCG soll Ressourcen aus Rechenzentren in Europa, Amerika und Asien in mehreren virtuellen Organisationen zusammenfassen und Grid Services für etwa 6.000 Wissenschaftler weltweit zur Verfügung stellen. Das derzeit am IWR entstehende Grid Computing Zentrum Karlsruhe, GridKa [10], wird eines der acht leistungsfähigsten Rechenzentren in diesem Netzwerk sein.

Die Entwicklung des LCG schreitet zügig voran: zwecks Integration und künftiger Interoperabilität mit weiteren Grid-Projekten wur-

den die Middleware-Komponenten aus verschiedenen europäischen und amerikanischen Projekten neu zusammengestellt und aneinander angepasst, so dass im September 2003 mit LCG-1 auf 12 großen Rechenclustern erste Tests durchgeführt werden konnten. Mit einer verbesserten Version LCG-2 wurde dann im April 2004 in so genannten weltweiten „Data Challenges“ begonnen, die Funktionsfähigkeit, Stabilität und Skalierbarkeit der Software zur Simulation und Datenauswertung der LHC-Experimente auf dem Grid zu testen. Anfang Juni 2004 war die LCG-2 Middleware bereits auf 58 Rechenclustern in 22 Ländern mit einer Ge-

samtkapazität von 3.800 Prozessoren installiert. Untersucht wurde in diesen Data Challenges insbesondere auch die Interoperabilität mit dem amerikanischen Grid3 und dem skandinavischen NorduGrid. Abb. 2 gibt eine Übersicht über die Institute, die ihre Rechnersysteme im Juni 2004 mittels LCG-2 zu einem Grid zusammengeschlossen haben.

Europa auf dem Weg zur e-Science

Auch wenn in Einzelprojekten wie dem LCG bereits beachtliche Erfolge zu verzeichnen sind, so sind viele Detailfragen noch Gegenstand der aktuellen Forschung und Entwicklung: Wie organisiert und optimiert man große verteilte Datenströme in einem weltweiten Namensraum? Wie gut skalieren die Netzwerktechnologien, Prozessorleistungen, Datenspeicher? Wie administriert man die Accounts und globalen Aktivitäten von tausenden von Benutzern und wie rechnet man sie ab? Wie garantiert man den sicheren Zugriff auf die Daten und wie schützt man die Rechnersysteme in einer prinzipiell offenen Grid-Welt vor Angriffen? Und schließlich – wie organisiert man eine weltweite, qualifizierte Benutzerunterstützung über mehrere Zeitzonen?

Eines der vielleicht wichtigsten europäischen Projekte, das in den nächsten Jahren Antworten auf diese Fragen finden und *demonstrieren* soll, ist das EGEE („Enabling Grids and E-science in Europe“) [12]. Gefördert mit einem Gesamtvolumen von 32 Millionen Euro für zunächst 2 Jahre arbeiten seit April 2004 70 (!) Part-

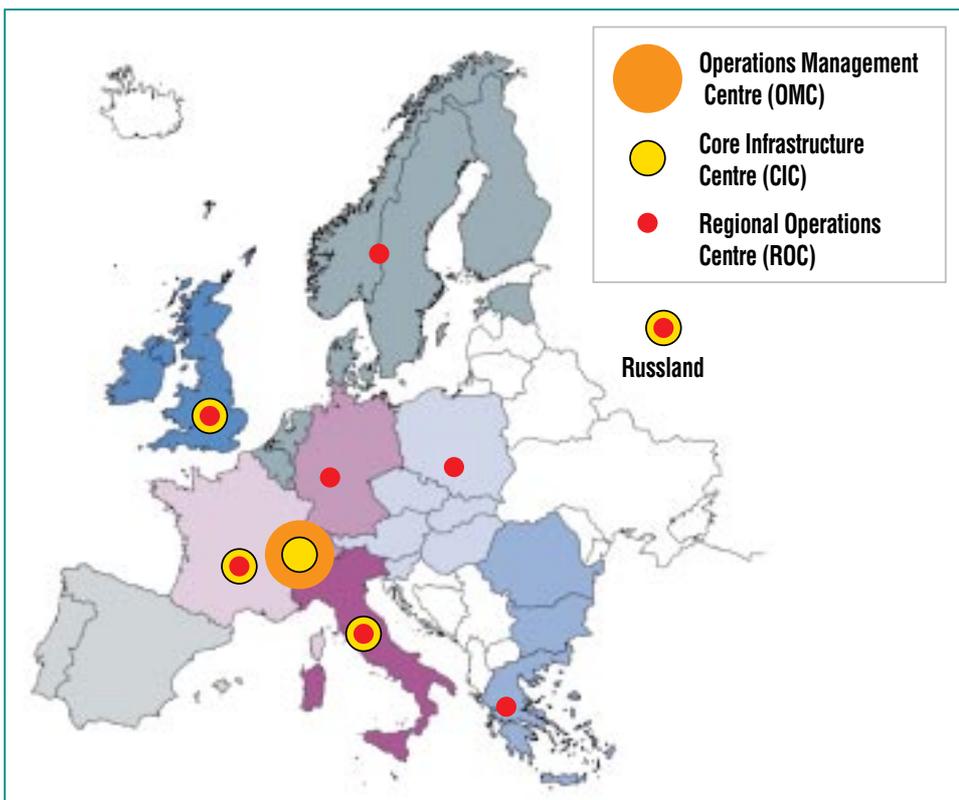


Abb. 3: Das Projekt EGEE unterteilt Europa in Föderationen. Eine Hierarchie von Zentren übernimmt das Gesamtmanagement (OMC), betreibt zentrale Dienste der Grid-Infrastruktur (CIC) bzw. übernimmt bestimmte Aufgaben für Management und Betrieb innerhalb einer Föderation (ROC).

ner in 27 Ländern daran, Rechnersysteme unterschiedlichster Wissenschaftszweige und bereits existierende Grid-Initiativen zu einer gesamteuropäischen Grid-Infrastruktur zusammen zu schließen. Abb. 3 zeigt die organisatorische Zusammenfassung der EGEE-Partnerländer zu Föderationen. EGEE wird Werkzeuge, Verfahrensweisen und Musterverträge für ein europäisches Grid-Management entwickeln, die es erlauben, auch in Zukunft einen gemeinsamen Markt für Rechenzeit zu betreiben und zu erweitern. Aber auch die existierenden Middleware-Technologien werden in Richtung Web-

Services nach den GGF-Standards weiterentwickelt. Neben den deutschen Partnern DESY, DKRZ, FhG, GSI ist auch das Forschungszentrum in diesem Projekt insbesondere in den Bereichen Dissemination, Benutzertraining und Grid-Support, -Betrieb und -Management beteiligt.

Zusammenfassung

Vor etwas mehr als zehn Jahren hat Tim Berners-Lee am CERN das World Wide Web als Medium für den Austausch zwischen Wissenschaftlern entwickelt, und dessen Öffnung für Firmen und für die Öffentlichkeit hat zu völlig

neuen Kommunikations- und Geschäftsmodellen geführt. Die wachsende Anzahl und Größe von Grid-Projekten in aller Welt sowie die starke industrielle Beteiligung an Standardisierungsprozessen deutet darauf hin, dass wir auf gutem Wege sind, das World Wide Web (WWW) in ein World Wide Grid (WWG) zu verwandeln, in dem jeder Daten und Rechenleistung wie Strom aus der Steckdose beziehen kann.

Literatur und Abkürzungen

- | | | |
|--|---|---|
| <p>[1] I. Foster, C. Kesselman, „The Grid – Blueprint for a new computing infrastructure“, Morgan Kaufmann Publ. 1999</p> <p>[2] Géant – Das gesamteuropäische Forschungsnetz, http://www.dante.net</p> <p>[3] TeraGrid – Ein amerikanisches Grid von Hochleistungsrechenzentren, http://www.teragrid.org</p> <p>[4] Globus – Globus Alliance und das Globus Toolkit, http://www.globus.org</p> | <p>[5] GGF – Global Grid Forum, http://www.gridforum.org</p> <p>[6] EDG – Das europäische DataGrid Projekt, http://www.eu-datagrid.org</p> <p>[7] CrossGrid – Das europäische CrossGrid Projekt, http://www.eu-crossgrid.org</p> <p>[8] LHC – Large Hadron Collider am CERN, http://www.cern.ch/lhc</p> <p>[9] LCG – LHC Computing Grid Projekt, http://www.cern.ch/lcg</p> | <p>[10] GridKa – Grid Computing Zentrum Karlsruhe, http://www.gridka.de</p> <p>[11] LCG Mapcenter – http://goc.grid-support.ac.uk/mapcenter</p> <p>[12] EGEE – Das europäische Projekt „Enabling Grids and E-Science in Europe“, http://www.eu-egee.org</p> |
|--|---|---|

GridKa: Das Grid Computing Zentrum Karlsruhe für Anwendungen aus der Hochenergiephysik

J. van Wezel, H. Marten, IWR

Einleitung

Die deutschen Elementarteilchenphysiker veröffentlichten im Jahr 2001 ihre Anforderungen an ein Regionales Daten- und Rechenzentrum in einem Memorandum. Der Vorschlag, Rechenkapazitäten in einem Zentrum zu bündeln, um bevorstehende Anforderungen des *Large Hadron Colliders* (LHC) bei CERN zu meistern, wurde vom Forschungszentrum Karlsruhe schon im Dezember 2001 mit einem Plan und der Zusage zum Bau und Betrieb eines Groß-Rechenclusters beantwortet [1]. Aus den Anfängen als „*Regional Data and Computing Centre*“ Karlsruhe ging im Oktober 2002 das *Grid Computing Centre* Karlsruhe mit dem Namen GridKa hervor (Abb. 1). Das GridKa-Zentrum ist verbunden mit dem Institut für Wissenschaftliches Rechnen (IWR) des Forschungszentrums und ist in den Helmholtz-Forschungsbereich „Struktur der Materie“ eingebettet.

Im GridKa arbeiten viele Forschungseinrichtungen zusammen

Der Aufbau eines der größten Computercluster in Deutschland ist ein langjähriges und komplexes Projekt. Nach der Bau- und Testphase wird der eigentliche Betrieb mindestens 15 Jahre dauern und erfordert ein ständiges Anpassen und Modernisieren der Hard- und Software und gegebenenfalls Erweitern der Komponenten für viele Jahre im Voraus. Die Erfahrung des IWR wird hierbei unterstützt von verschiedenen Experten. Im *Projekt Overview Board* begleiten und steuern Vertreter aus Physikausschüssen, BMBF und der Helmholtz-Gemeinschaft den Aufbau. Ein zweites Gremium mit Vertretern aus der Physik von allen beteiligten Universitäten und des Deutschen Elektronen-Synchrotrons DESY stärkt die technische Abstimmung zwischen Benutzerwünschen und Realisation.

Die Benutzer des GridKa-Zentrums sind in erster Linie Wissenschaftler der Deutschen Hochenergiephysik. Über Partnerschaften mit Institutionen wie den Helmholtz-Zentren DESY, GSI und Forschungszentrum Jülich oder CERN in Genf, RAL in Oxford, aber auch SLAC und FNAL in den USA, werden Erfahrungen ausgetauscht. So bekommen auch ausländische Benutzer die Möglichkeit, bei GridKa zu rechnen.

Die Rechenleistung des GridKa-Clusters wird vorerst für die Hochenergiephysik eingesetzt. Es wird aber ausdrücklich vorgesehen, künftig auch Kapazitäten für Disziplinen außerhalb der Teilchenphysik bereitzustellen, z.B. für die Astronomie, Astroteilchenphysik, Klimaforschung und Genetik.

Messungen am Large Hadron Collider (LHC)

Der Large Hadron Collider ist ein Teilchenbeschleuniger, mit dem es möglich sein wird, die Bausteine der Materie weiter zu entschlüsseln und fundamentale Fragen aus der Physik zu beantworten. Der LHC wird am Europäischen Laboratorium für Teilchenphysik CERN in Genf in der Schweiz gebaut und benutzt den bestehenden Tunnel des früheren LEP-Ringes. Der unterirdische Beschleuniger hat einen Kreisumfang von 27 km und beschleunigt zwei Strahlen in entgegengesetzter Richtung. Die starken Magnetfelder, die die Teilchen auf ihrer Bahn halten, werden von supraleitenden Spulen bei Tempe-



Abb. 1: Seit der Einweihung schmückt sich das Grid Computing Centre Karlsruhe mit dem GridKa-Logo.

raturen um -270°C erzeugt. Ab 2007 werden im LHC alle 25 ns Protonen mit einer Energie von 14 TeV aufeinander treffen. Daneben werden auch Strahlen von Bleikernen beschleunigt und kollidieren mit einer Aufprallenergie von 1150 TeV. An den Kollisionspunkten werden Detektoren aufgestellt, um die Reaktionen beim Aufeinandertreffen der Protonenstrahlen aufzuzeichnen. Die Experimente werden in bislang unerschlossenen Energiebereichen nach neuen Teilchen suchen. Das so genannte Higgs-Teilchen und der Ursprung der Masse sind grundlegende Fragen, die von den LHC-Experimenten erforscht werden.

Jeweils optimiert für spezifische Messungen kommen vier riesige Detektoren zum Einsatz. Neben der Konstruktion und dem Aufbau der Detektoren ATLAS, CMS, ALICE and LHCb werden auch Entwürfe für Software und Computeranlagen zur Speicherung und späteren Auswertung der Messdaten erarbeitet. Die Aufzeichnung und Auswertung der Daten stellt eine gewaltige Herausforderung dar: im LHC werden in jeder Sekunde rund 800 Millionen Teilchenkollisionen stattfinden. Das bedeutet, dass von Zehntausenden von Rechnern viele Petabyte Daten pro Jahr verarbeitet werden müssen. An dieser Verarbeitung ist das GridKa-Zentrum des Forschungszentrums Karlsruhe maßgeblich beteiligt und wird zusammen mit weltweit 7 bis 8 weiteren Rechenzentren den Kern des LHC Computing Grid bilden.

Regionale Zentren

Ein einziges Rechenzentrum am CERN wäre logistisch und finanziell nicht in der Lage, die Aufgaben des LHC-Computing zu erfüllen. Wie Satelliten werden deshalb regionale Zentren jeweils einen Teil der Rechenleistung übernehmen. Der Report „Regional Centre Category and Service Definition“ [6] beschreibt Kategorien, basierend auf notwendigen Dienstleistungen für Rechenzentren, die am LHC-Computing teilnehmen. GridKa steht als Kategorie-1-Zentrum in einer Rechner- und Speicherkette, die bei CERN (Kategorie 0) anfängt und bei Kategorie 4, dem Arbeitsplatzrechner der Teilchenphysiker, endet. Zu der Gruppe von Kategorie-1-Zentren

gehören unter anderem IN2P3 in Lyon (F), RAL (GB) und FNAL in Batavia/Chicago (USA). Ein Kategorie-1-Zentrum (oder auf Englisch: Tier-1) wird neben Rechen vor allem Speicherkapazität anbieten und über eine direkte Hochgeschwindigkeitsnetzwerk-anbindung zum CERN die Messdaten entgegennehmen und speichern. Weiterhin wird von einem Kategorie-1-Zentrum eine sehr große Zuverlässigkeit und Betreuung rund um die Uhr erwartet. Die Kategorie-2-Ebene werden Zentren an Universitäten, Hochschulen und Forschungseinrichtungen bieten. Die Anforderungen an Benutzerunterstützung und Verfügbarkeit sind für Kategorie-2-Zentren geringer als für Zentren der Kategorie 1 (Abb. 2).

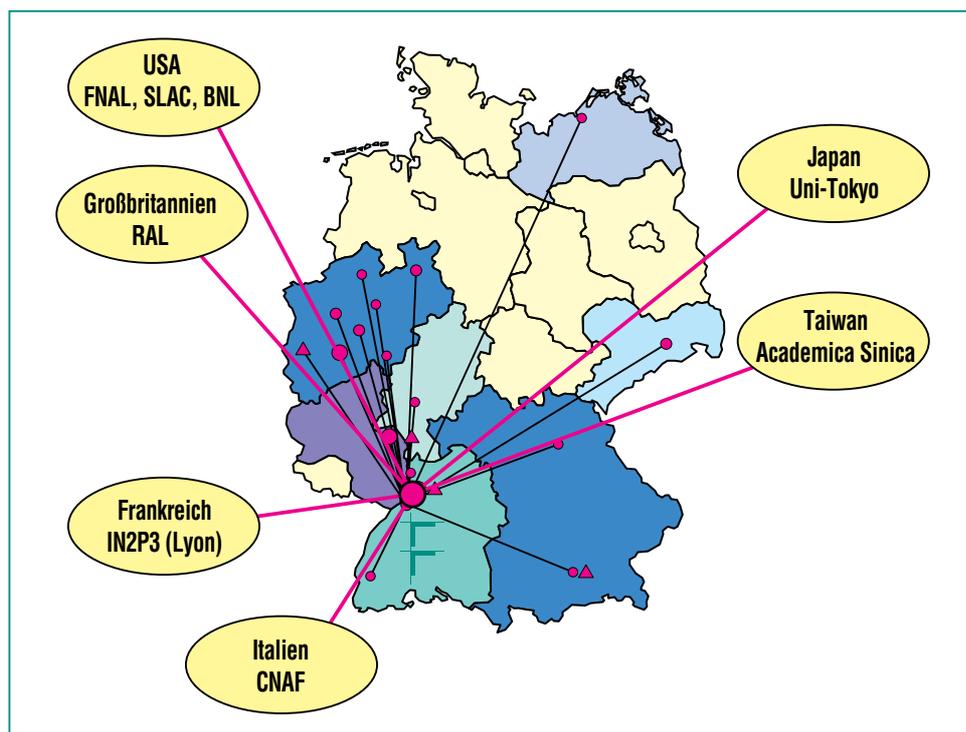


Abb. 2: Der Clusterrechner am GridKa wird betrieben für Universitäten und Forschungseinrichtungen in Deutschland. Innerhalb des LHC-Computing-Grid-Projektes arbeitet GridKa weltweit zusammen mit verschiedenen Kategorie-1-Zentren.

Rechen- und Speicheranforderungen des LHC

Insgesamt vier große Detektoren sind im Bau zur Messdatenerfassung. Beispielsweise hat der ATLAS-Detektor etwa 200 Millionen Lesekanäle. Im Betrieb liefern diese 40 TB Messdaten pro Sekunde, die online über eingebaute Prozessoren und eine Rechneranlage auf 100 MB/s reduziert werden. Zusätzlich werden noch *Condition Data* gespeichert, die den Zustand des Detektors im Moment der Messung beschreiben. Man schätzt, dass der ATLAS-Detektor pro Jahr ca. 3 Petabyte Daten liefert, die auf Band gespeichert und an die Kategorie-1-Zentren verteilt werden müssen. Dies bedeutet, dass das GridKa-Cluster über lange Zeiträume Daten mit über 250 MegaByte pro Sekunde entgegen nehmen muss.

Die Rohdaten aus den Detektoren werden in erster Linie zur Rekonstruktion der Kollisionen verwendet, eine sehr rechenintensive Aufgabe. Ergebnis dieser Rekonstruktion sind die *Event Summary Daten* (ESD), die eine möglichst genaue Wiedergabe der Kollision enthalten. Die ESD müssen ständig online zur Verfügung stehen. Eine weitere Rechenaufgabe ist die Simulation der Messungen in einem Monte-Carlo-Verfahren. Die simulierten Messdaten werden später zur Auswertung der wirklichen Daten gebraucht.

Für alle Experimente liegen Planungen der Datenmengen und der Anforderungen an Rechenkapazität vor. Für die Datenspeicherung wird unterschieden zwi-

schen on-line auf Platte und offline auf Band gespeicherten Daten. Die meisten Daten werden nach der ersten Analyse nur noch selten gebraucht und müssen darum nicht auf teuren Platten gehalten werden. Die Aufbewahrung auf Band ist um ein Vielfaches billiger. Zu erwarten ist allerdings, dass sich in der Praxis die Anforderungen an Rechenzeit und Datenspeicher aufgrund neuer technologischer Entwicklungen in der Hardware und neuer Erkenntnisse in der Datenauswertung im Grid noch verschieben werden.

Andere Experimente

Der LHC wird erst im Jahr 2007 in Betrieb gehen. Bis dahin arbeiten die vier LHC-Experimente zur Vorbereitung und zu Testzwecken mit simulierten Daten. Die deutsche Teilchenphysik ist allerdings nicht nur am LHC beteiligt, sondern auch an drei Experimenten in den USA engagiert: das sind BaBar am Stanford Linear Accelerator (SLAC) [3], sowie D0 und CDF am Fermi National Accelerator

Laboratory (FNAL) in Batavia/Chicago [4]. Ein weiteres Experiment außerhalb des LHC-Komplexes ist COMPASS am Super Protonen Synchrotron SPS des CERN [5], das als Vorbeschleuniger zum LHC weiter in Betrieb bleibt. Die meisten Experimente haben eine eigene Software entwickelt, die es erlaubt, weltweit verteilte Rechenzentren für Datenspeicherung und Analyse zu benutzen. Seit seiner Gründung hat GridKa einen signifikanten Anteil seiner stetig wachsenden Rechenkapazität für diese Nicht-LHC-Experimente zur Verfügung gestellt. Im Laufe der kommenden Jahre wird dieser Rechenzeitanteil von dem der LHC-Experimente übertroffen werden (Abb. 3).

Grid und LCG

Über mehrere Jahre hatte das „Models of Networked Analysis at Regional Centres“ (MONARC) Projekt Basismodelle für das zukünftige LHC-Computerverfahren geliefert und ein Toolset zur Simulation von Daten produ-

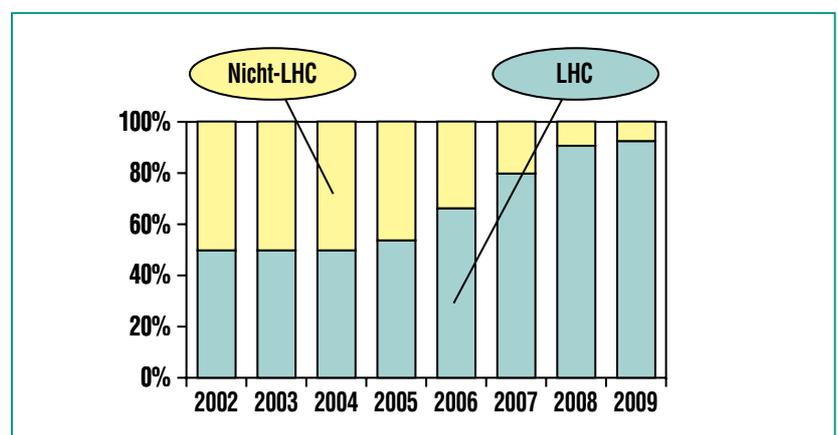


Abb. 3: Der Rechenzeitanteil der LHC-Experimente wird bei GridKa in den nächsten Jahren gegenüber demjenigen der Nicht-LHC-Experimente stetig anwachsen.

ziert, mit dem die Methoden für verteiltes Rechnen und Datenspeicherung ausprobiert werden konnten. Die MONARC-Studie hat zugleich Richtlinien für die technische Ausrüstung, Konfiguration und Dienstleistung von regionalen Computer-Zentren vorgeschlagen und damit die Basis für die Ausrüstung des GridKa-Zentrums gelegt.

Die Möglichkeit, Rechen- und Speicheraufgaben weltweit über Zentren zu verteilen und damit ein Computing-Grid zu bilden, ist die Aufgabe des LHC-Computing-Grid-(LCG)-Projektes. Innerhalb LCG werden die Entwicklungen der so genannten *Grid Middleware* zusammengeführt. Die *Middleware*-Software-Schicht bildet, oberhalb der Betriebssysteme der individuellen Rechner, mit Bibliotheken und Werkzeugen den Rahmen für die Anwendersoftware der LHC-Experimente. Die LCG-Software enthält Komponenten zur Benutzer-Authentifizierung über Virtuelle Organisationen, für eine Benutzerschnittstelle zum Abschicken und Verwalten von Programmen und zur Datenverwaltung und zum Datenaustausch über *wide area* Netzwerke (Abb. 4).

LCG und GridKa

LCG hat im März diesen Jahres Version 2 der Middleware herausgegeben, die auf dem GridKa-Cluster installiert wurde. Der Aufbau liegt damit gut in der Planung.

Mit regelmäßigen *Data Challenges* oder Probeläufen wird nicht nur die Funktionsfähigkeit der Software getestet, sondern auch

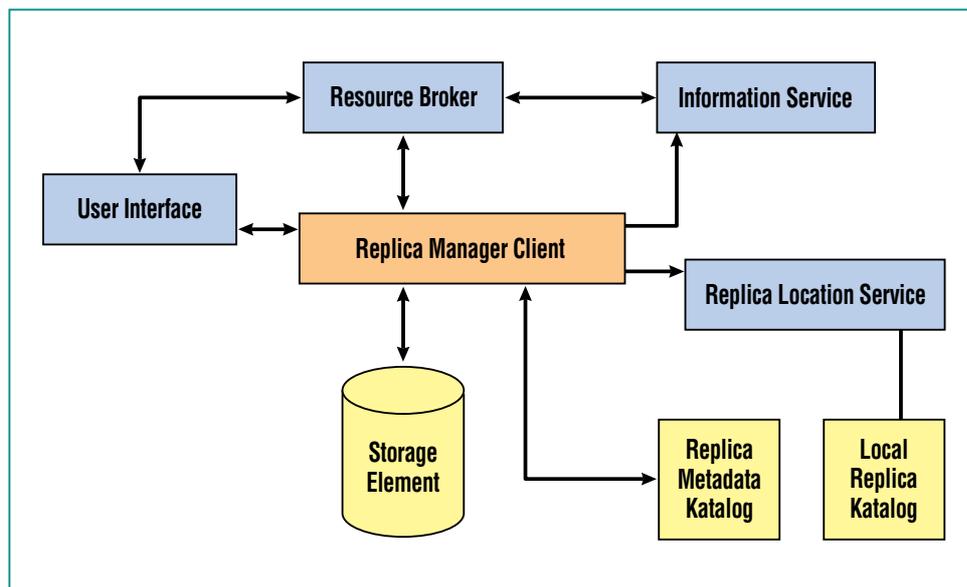


Abb. 4: Aufgabe der LCG „Replica Manager“ Middleware ist das Datenmanagement. Dateien und ihre Speicherorte werden in so genannte Katalogen gehalten.

die Stabilität und Leistungsfähigkeit der einzelnen Kategorie-1-Zentren. Die Experimentgruppen setzen hierzu eigene Applikationen ein und benutzen die LCG-Software zum Starten der Programme und zur Verwaltung der zu verarbeitenden Daten.

Die LCG-Softwareentwickler werden künftig immer häufiger neue Versionen bereitstellen, weil jetzt alle Komponenten erstmals zusammen eingesetzt werden können. Ein wichtiger Teil der Arbeit am GridKa ist die Installation und Verifikation neuer Versionen oder Teilen von Versionen sowie die Rückkopplung mit den Entwicklern und Anwendern. Es entsteht ein Zyklus von Anwendung über Fehlerbehebung bis zur Herausgabe einer neuen Softwareversion. Die Cluster an den Zentren sind nicht alle gleich, eine fehlerfreie Funktion muss aber in vielen unterschiedlichen Einrichtungen

gewährleistet sein. Daher ist die Testphase für das LCG sehr wichtig.

Zugleich stellt das Einbeziehen der laufenden Produktion in die Weiterentwicklung des Clusters eine interessante Herausforderung an den Betrieb dar. Der Aufbau des Clusters ist hierfür in Blöcken organisiert, und wichtige Software- und Hardwarekomponenten sind redundant ausgelegt. Auf diese Weise können Teile des Clusters während der Installation neuer Software oder Hardware zeitweise außer Betrieb genommen werden, während die Produktion auf den übrigen Teilen mit geringerer Kapazität weiterläuft.

GridKa Umfang und Erweiterung

Die Erweiterungen von GridKa verlaufen stufenweise und basieren auf Planungen der Teilchen-

physikexperimente und Fortschritten des LCG-Projekts. Nur weil jede Generation von Prozessoren und Speichermedien durch ständige Weiterentwicklung mehr Leistung bietet, sind die steigenden Anforderungen realisierbar. Die benötigte Anzahl der Prozessoren ist der zu erwartenden Vergrößerung der CPU-Leistung laut dem Mooreschen¹⁾ Gesetz laufend anzupassen. Die CPU-Anforderungen werden in den LHC-Modellen normiert auf die Leistung eines Intel Pentium-Prozessors von 1995 und mit kSI95 angegeben.

Der technologische Fortschritt in der verfügbaren Speicherdichte von Magnetbändern und Festplatten ist demjenigen der CPU-Leistung vergleichbar. Dennoch wird der größte Teil der Computerräume von GridKa im Endausbau mit Plattenspeichern und Magnetbandrobotern gefüllt sein (Abb. 5).

Im Laufe des Jahres 2004 wird das GridKa-Zentrum über das Deutsche Forschungsnetz (DFN) mit 10 Gigabit mit dem von DANTE²⁾ betriebenen europäischen Netzwerk Géant für Wissenschaft und Bildung verbunden werden. Diese Verbindung verspricht eine Datentransfer-Kapazität von 1000 MegaByte/Sekunde. Für eine effiziente Nutzung von Netzen mit derart hohen Bandbreiten ist aber noch eine Menge Entwicklungs- und Optimierungsarbeit zu leisten. Das IWR kooperiert dazu eng mit dem DFN und CERN.

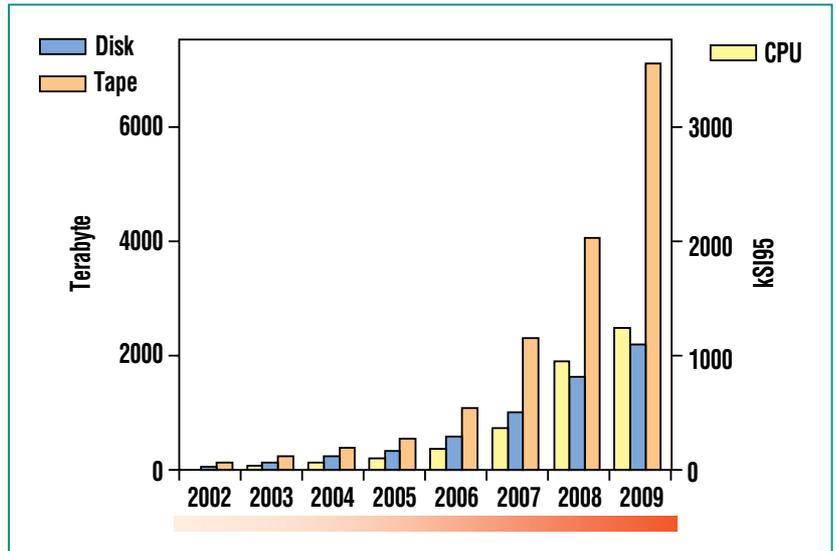


Abb. 5: Hardware-Erweiterung des GridKa-Clusters bis 2009.

Zusammenfassung

Grid Computing verspricht die Bereitstellung verteilter Speicher- und Rechenleistung je nach Anforderung unterschiedlichster Nutzer. Die dazu benötigten Techniken werden im Rahmen des LCG-Projekts entwickelt. Davon wird letztendlich jeder Computer-Besitzer profitieren – so, wie das auch bei der Entwicklung des World Wide Web der Fall war. Die Anwendung von Grid-Werkzeugen für die Datenanalyse der Experimente am Large Hadron Collider bietet der Teilchenphysik die Möglichkeit, Rechenkapazität weltweit und daher effizienter zu nutzen. Das Cluster des Grid Computing Centre Karlsruhe stellt Rechen- und Speicherleistung für die deutsche Hochenergiephysik bereit und ist über LCG2-Grid-Software zugänglich.

Literatur

- [1] *A Grid Computing Centre at Forschungszentrum Karlsruhe:*
http://lhcg.grid.web.cern.ch/LHCgrid/PEB/Documents/FZK_RDCCG-answer-v7.doc
- [2] *MONARC –*
<http://monarc.web.cern.ch/MONARC/docs/Phase2Report/Phase2Report.pdf>
- [3] *Web-Seite des BaBar experiment:*
<http://www.slac.stanford.edu/BFROOT>
- [4] *Web-Seite der Experimente D0:*
<http://www-d0.fnal.gov>
<http://www-cdf.fnal.gov>
- [5] *Web-Seite des Compass Experiment:*
<http://wwwcompass.cern.ch>
- [6] *Report of the LHC Computing Grid Project, Document identifier CERN-LCG-2002-16:*
<http://lcg.web.cern.ch/LCG/Documents/>

¹⁾ Gordon Moore ist Mitgründer des Halbleiterproduzenten Intel und hat 1965 erstmals einen exponentiellen Anstieg der Transistorzahl pro integrierte Schaltung festgestellt.

²⁾ Delivery of Advanced Network Technology to Europe, DANTE ist eine Non-profit-Organisation für Planung, Bau und Betrieb europaweiter Forschungsnetzwerke.

Aufbau des großen GridKa-Clusters am Forschungszentrum Karlsruhe

M. Alef, B. Hoefft, H. Marten, J. van Wezel, IWR

Einleitung

Teilchenphysiker aus aller Welt arbeiten daran, ihre Meßergebnisse in einem extrem leistungsfähigen Daten-Grid auszuwerten. Die riesigen zu verarbeitenden Datenmengen stellen enorme Anforderungen an die beteiligten Rechenzentren. Das „Grid Computing Centre Karlsruhe“ (GridKa), das seit Herbst 2001 im Institut für Wissenschaftliches Rechnen (IWR) des Forschungszentrums Karlsruhe als deutsches LHC-Tier1-Regionalrechenzentrum aufgebaut wird [1, 2], hat diese Herausforderung angenommen (Abb. 1, 2). Dort wird von den 4 LHC-Experimenten Alice, Atlas, CMS und LHCb in weltweiten so genannten „Data Challenges“ die Funktionsfähigkeit, Stabilität und Skalierbarkeit der Software zur Detektorentwicklung, zur Simulation und zur Da-



Abb. 1: Blick in das GridKa-Rechenzentrum.

tenauswertung getestet. Gleichzeitig steht das GridKa auch den 4 weiteren schon heute aktiven Hochenergiephysik-Experimenten BaBar (SLAC, Stanford), CDF

und D0 (Fermilab) und Compass (CERN) für deren Datenanalyse zur Verfügung. Mit den daraus gewonnenen Erfahrungen entwickeln GridKa-Mitarbeiter zusammen mit den Nutzern sowie den anderen Grid-Zentren die Grid-Infrastruktur weiter.

Die Ressourcen des GridKa werden, entsprechend den Anforderungen der beteiligten Experimente, zweimal im Jahr erweitert. Zum „Meilenstein Oktober 2004“ umfaßt das GridKa rund 1060 CPUs, 220 TB (netto) Online-Platz auf Magnetplattensystemen sowie mehrere 100 TB im Magnetbandarchiv. Diese Kapazitäten werden bis zum Jahr 2007 auf rund das 4-fache anwachsen.

In mehreren Disziplinen – Netzwerkanbindung, Datenhaltung und Clusterbetrieb/-Kühlung – zählt GridKa inzwischen schon zur Riege der weltweit führenden Rechenzentren.

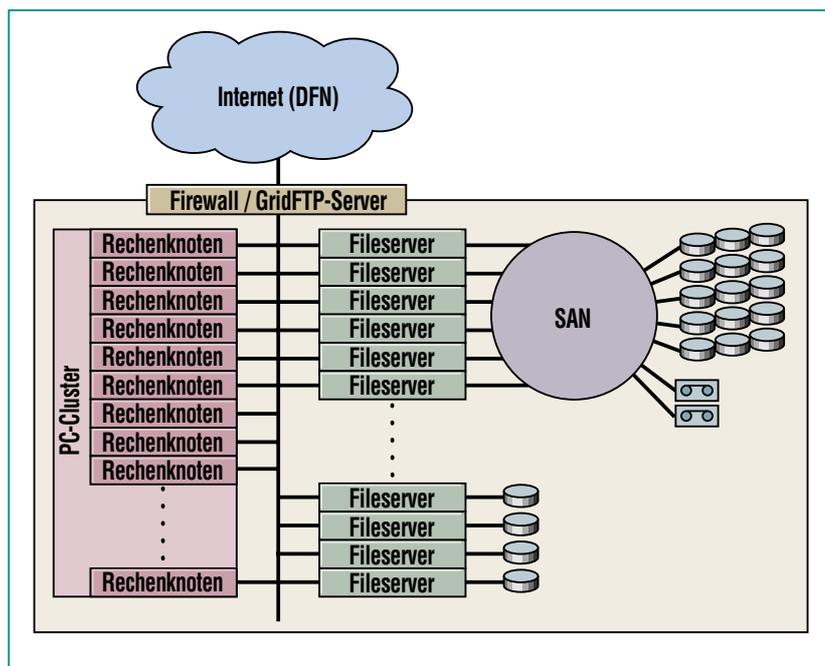


Abb. 2: Struktur des GridKa-Clusters (Rechenknoten, Fileserver, Netzwerkanbindung).

PC-Cluster

Für die Auswertung der Experimentdaten, sowie für Simulationsrechnungen, ist ein PC-Cluster installiert, der im Oktober 2004 etwa 1.000 Intel-Prozessoren (Pentium 3, 1,26 GHz, bis Xeon, 3,06 GHz) sowie rund 70 AMD-Opteron-CPU's enthält [3]. Jeder einzelne Rechenknoten dieses Clusters ist als Doppelprozessorsystem mit jeweils 1 bis 2 GB Hauptspeicher sowie einer lokalen Festplatte ausgelegt. Als Betriebssystem ist Linux installiert. Es ist bisher nicht geplant, parallele Anwendungen laufen zu lassen. Deshalb sind die meisten Rechner lediglich per Fast-Ethernet angebunden. Die neueren Knoten haben eine Gigabit-Ethernet-Anbindung.

Um eine solch große Anzahl an Rechnern effizient verwalten zu können, sind leistungsfähige, standardisierte und skalierbare Installations-, Administrations- und Überwachungsverfahren unbedingt notwendig. Für die Linux-Betriebssysteminstallation wurde das Rocks-Toolkit des San Diego Supercomputer-Zentrums [4] auf einem zentralen Installationsserver bereitgestellt und an die Er-

fordernisse des GridKa angepaßt. Bei der Anlieferung neuer Rechner kann ein kompletter Schrank mit 36 Rechnern innerhalb von 1 Stunde (!) installiert werden.

Das GridKa-Cluster wird im sogenannten Batchmodus betrieben. Dabei übergibt der Benutzer dem System vordefinierte Prozeduren und erhält einige Zeit später die Ergebnisse. Im Rocks-Toolkit ist das Batchsystem OpenPBS [5] als Standard enthalten; aus Stabilitäts- und Supportgründen wird im GridKa zur Zeit jedoch die kommerzielle Weiterentwicklung PBS-Pro verwendet [6] (Abb. 3a).

Schwachstellen und Engpässe im System können mit dem Auslastungsmonitor Ganglia erkannt werden [7], und zur „Gesundheitsüberwachung“ wird Nagios verwendet [8] (Abb. 3b).

Kühlung

Ein nicht zu unterschätzendes Problem beim Aufbau kompakter Clustersysteme ist die Kühlung. Der Stromverbrauch – und damit die Wärmeabgabe – der installierten Rechner ist in den vergangenen Jahren parallel zur Steigerung der Leistungsfähigkeit

enorm angewachsen (Abb. 4). Ein zuverlässiger Betrieb des Clusters ist nur möglich, wenn die Rechner ausreichend gekühlt werden.

Schon zu Beginn des GridKa-Aufbauprojekts war abzusehen, daß die im Rechenzentrum vorhandene Klimaanlage nicht ausreicht. Eine Erweiterung wäre aufwendig geworden und hätte sehr hohe Luftströmungsgeschwindigkeiten notwendig gemacht, die die Aufenthaltsqualität im Rechnerraum und damit die Konzentrationsfähigkeit der Mitarbeiter zum Beispiel bei einer Störungssuche empfindlich eingeschränkt hätte.

Bei den früheren Großrechnern, auch bei denen des IWR, war eine Wasserkühlung üblich. Im PC-Bereich sind zwar wasserdurchströmte CPU-Kühlkörper auf dem Markt und werden inzwischen auch schon von einem Hersteller von Rechnerschrank angeboten. Für den Aufbau des GridKa-Clusters kam eine solche Technik aber nicht infrage: Einerseits dürfen die durch die Nähe von Wasser und Elektrizität denkbaren Sicherheitsprobleme nicht vernachlässigt werden [9]. Andererseits erzeugen nicht nur die Prozessoren Wärme, sondern auch die anderen Rechnerkomponenten (Hauptspeicher, Mainboard, Festplatte, Netzteil, ...). Schätzungen sagen, daß deshalb auf diese Art nur rund 60% der anfallenden Wärme per Wasser abgeleitet werden können [10].

Im GridKa wurden als praxisgerechte Alternative – in Zusammenarbeit mit dem Elektronikschrankhersteller Knürr – ge-

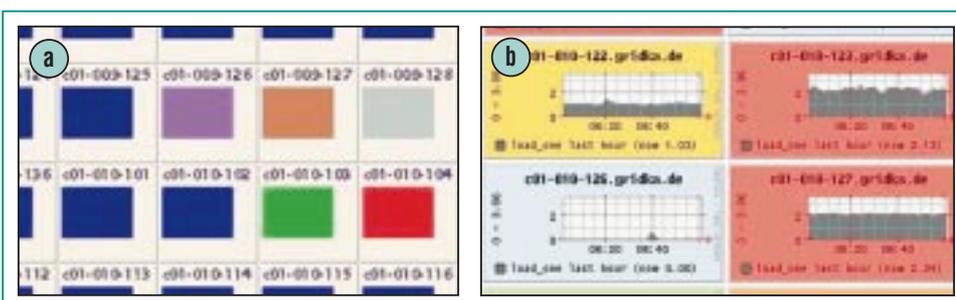


Abb. 3a+b: Grafische Benutzungsoberflächen helfen bei der Bedienung und Überwachung der GridKa-Rechner: Übersicht über die Auslastung des Clusters a) im Batchbetrieb (PBS-Monitor), b) detailliertere Ansicht (Ganglia).

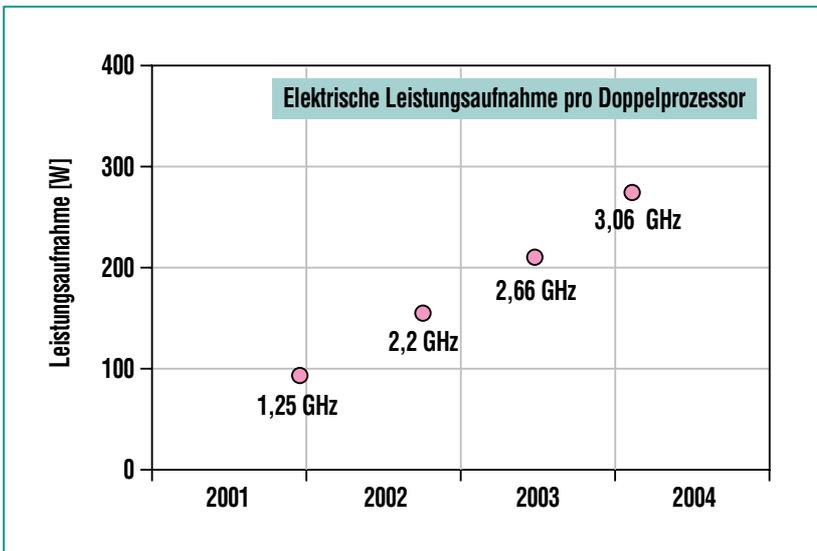


Abb. 4: In den letzten Jahren ist die elektrische Leistungsaufnahme – und damit auch die Wärmeabgabe – von PCs drastisch angestiegen. Die Grafik zeigt die Leistungsaufnahme der im GridKa eingesetzten Doppelprozessorsysteme in Abhängigkeit vom Zeitpunkt der Inbetriebnahme (bzw. dem Prozessortakt).

geschlossene Rechnerschränke konzipiert, in denen ein zentraler Luft-Wasser-Wärmetauscher die von Lüftern innerhalb des Schrankes umgewälzte Luft abkühlt (Abb. 5a+b). Im Oktober 2002 wurde ein Prototyp geliefert und in Betrieb genommen, ohne daß bisher grundsätzliche Probleme aufgetreten sind [3]. Mit den inzwischen knapp 30 weiteren solcher Schränke ist das GridKa des Forschungszentrums Karlsruhe weltweit das erste Rechenzentrum, das konsequent auf eine solche eigentlich simple, aber wirkungsvolle Technik zur vollständigen Ableitung der Rechnerabwärme setzt!

Lokale Datenhaltung

In einem Grid kann im Prinzip jeder Rechner auf die weltweit verteilten Daten zugreifen. Im Prinzip könnte ein Programm seine Eingabedaten

ten bei ausreichend performanter Internet-(WAN-)Anbindung aus einem entfernten Rechenzentrum abholen und die Ergebnisse dort wieder ablegen. In der Praxis ist das jedoch nicht realisierbar: Die Antwortzeiten wachsen mit der Entfernung, bei einem Datentransfer zum Beispiel vom Fermilab in Chicago liegen diese bei mindestens etwa 50 Millisekunden. (Ein 3-GHz-Prozessor verbraucht derzeit 150 Millionen Taktzyklen mit Warten!) Die Zugriffszeiten auf eine lokale Festplatte sind um ein Vielfaches kleiner. Ein weiteres Argument ist, daß zur direkten Versorgung der vielen Clusterknoten eine WAN-Anbindung zu langsam und auch zu teuer wäre – das heutige GridKa-Cluster mit seinen rund 530 Knoten könnte theoretisch immerhin über 5 GBytes/s verarbeiten. Deshalb werden auch vor Ort weiterhin Datenspeicher benötigt.

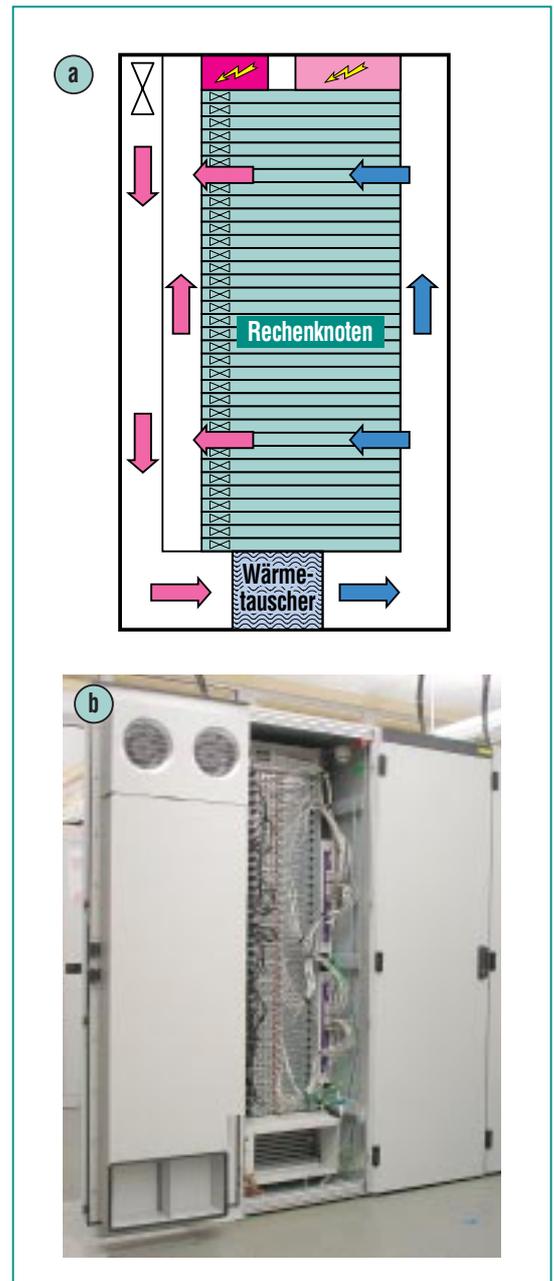


Abb. 5a+b: GridKa ist weltweit das erste Rechenzentrum, das die Abwärme der Rechner vollständig per Wasser ableitet. Die Luft wird von Ventilatoren, die oben in der rückseitigen Schranktür eingebaut sind, umgewälzt und durch einen Luft-Wasser-Wärmetauscher (unten) gedrückt. Die Rechner saugen vorne die gekühlte Luft an. Modifikationen an den Rechnern sind nicht notwendig.

Dennoch verschwimmen die Grenzen zwischen direkt oder per WAN angebundenem Speicher immer mehr. Die Gigabit-Strecke zwischen Hannover und Berlin wird zum Beispiel auch für den direkten Datenzugriff zwischen den dortigen Hochleistungsrechnern benutzt [11]. Es werden auch immer mehr Produkte entwickelt, mit denen Online-Daten verteilt und in sicherer Entfernung zu den Rechnern gehalten werden, um einen Datenverlust im Katastrophenfall zu minimieren. Und bei der jährlichen Supercomputer-Konferenz findet seit einigen Jahren der „High Performance Bandwidth Challenge“ statt, bei dem Firmen ihre Software für die direkte Speicheranbindung über WAN ins Rennen schicken [12]. Im folgenden sind aber mit „Online“ nur solche Daten gemeint, die lokal vor Ort gehalten werden.

Um allen Knoten des Clusters eine ausreichend hohe Bandbreite zu den Daten zu garantieren, wird für die Online-Datenhaltung im GridKa ein skalierbares paralleles Filesystem eingesetzt. Mehrere Rechner können damit gleichzeitig auf dieselbe Festplatte zugreifen. Zugleich können in einem derartigen System Daten schneller gelesen oder geschrieben werden, indem die Festplattenzugriffe parallelisiert werden. Dazu wird jede Datei über mehrere Platten verteilt („Striping“).

In Ethernet-Netzwerk sind mehrere gleichzeitige Zugriffe auf Dateien schon seit vielen Jahren möglich; NFS, SMB und AFS sind einige etablierte Protokolle zur gemeinsamen Nutzung von Dateisystemen in einem Netzwerk. Ein paralleles Filesystem bietet

im Prinzip die gleiche Möglichkeit, hier jedoch sind die Speicher z. B. mittels SCSI-, ATA- oder Fibre-Channel-Verbindungen direkt angebunden. Ein im Filesystem enthaltener, aufwendiger Locking-Mechanismus verhindert, daß mehrere angeschlossene Rechner gleichzeitig versuchen, dieselbe Datei zu verändern.

Für die Online-Datenhaltung wurde im GridKa das „General Parallel File System“ (GPFS) von IBM gewählt [13]. GPFS wird in Computer-Clustern unter AIX schon lange verwendet und ist seit rund 2 Jahren auch unter Linux verfügbar. Das parallele Filesystem läuft im GridKa auf einem eigenen, kleineren Fileserver-Cluster, das die Daten per NFS für die übrigen Rechner bereitstellt. Die GPFS-Fileserver sind weitgehend redundant, der Ausfall maximal knapp der Hälfte dieser Server oder deren Plattenanbindungen würde lediglich zu einer verringerten Datenübertragungsrates, aber nicht zu einem Totalausfall von Daten führen. Ein weiterer Vorteil ist, daß viele Wartungsarbeiten wie der Austausch von Platten oder Fileservern, oder das Vergrößern oder Verkleinern von Filesystemen, im laufenden Betrieb erfolgen können.

Datenaustausch und Sicherheit im Grid

Die Grid-Zugangrechner, die von außen direkt erreichbar sein müssen, sind durch eine Firewall vor potentiellen Angriffen aus dem Internet geschützt. Jedoch werden Firewall-Konzepte gerade im Grid-Zusammenhang kontrovers diskutiert, weil die er-

reichbare Bandbreite traditioneller Firewalls begrenzt ist und weil die Konfiguration der offenen Ports in einer weltweiten, dynamischen Kollaboration mit einem hohen Administrationsaufwand und vielen Fehlermöglichkeiten verbunden ist.

Eine sichere und skalierbare Lösung dazu können dedizierte Server sein, die direkt am Internet-Backbone angeschlossen sind und auf denen das „Grid File Transfer Protocol“ (GridFTP) läuft [14].

Um die Performance von GridFTP zwischen den beiden Grid-Partnern GridKa und CERN zu messen, wurde der WAN-Anschluß des Forschungszentrums Karlsruhe von ursprünglich 34 Mbit/s Ende 2001 auf derzeit 1 Gbit/s aufgerüstet (Abb. 6). Messungen haben ergeben, daß nicht das GridFTP-Protokoll selbst, sondern die darunter liegende TCP/IP-Schicht ein begrenzender Faktor ist. Während mit Standard-Implementierungen von TCP/IP über eine einzelne Gbit-Leitung eine mittlere Transferleistung von nur 470 Mbit/s erreicht wurde, konnten mit weiterentwickelten TCP/IP-Implementierungen bis zu 980 Mbit/s übertragen werden [15]. Allerdings können derartige Geschwindigkeiten nur beim Senden von Daten erzielt werden, die vom Hauptspeicher des Senders direkt in den Hauptspeicher des Empfängers geschrieben werden. Sobald der Transfer von/auf Festplatten oder gar Magnetbändern verläuft, ist deren Mechanik der begrenzende Faktor. Eine der Herausforderungen für die Zukunft wird darin bestehen, die Datenübertragungen

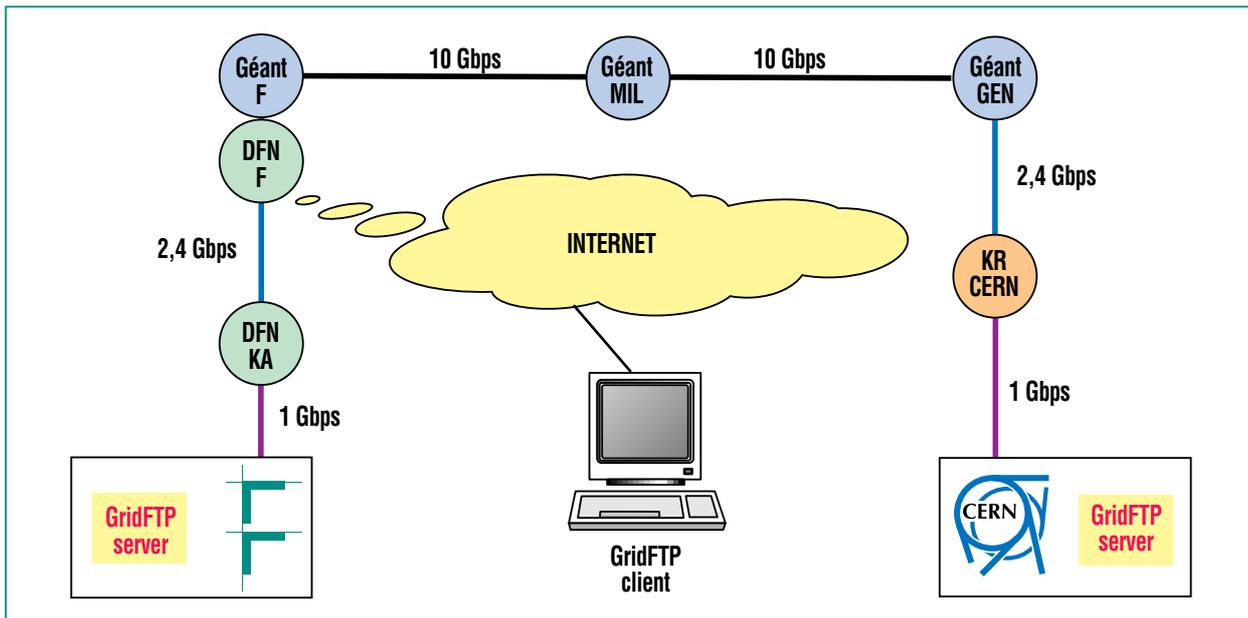


Abb. 6: GridFTP zwischen dem Forschungszentrum Karlsruhe (GridKa) und CERN. Externe Nutzer haben über das Internet weltweiten Zugriff auf ihre Daten.

bis auf die Speichermedien hinab zu parallelisieren. Die GridFTP-Spezifikation berücksichtigt bereits nicht nur die Parallelisierung eines Datenstromes über mehrere Prozesse einer Server-Client-Kommunikation, sondern auch die Aufspaltung eines Datentransfers über mehrere parallel geschaltete Server. Letzteres ist allerdings zur Zeit noch nicht implementiert.

Während die beschriebene Gigabit-Internet-Anbindung im Grid-

Ka schon genutzt wird, werden ab Herbst 2004 Tests mit der 10-Gigabit-Technologie beginnen.

Ausblick

Das Grid wird über enorme Rechner- und Speicherkapazitäten verfügen, die über die weltweiten Rechenzentren verteilt werden. Im Grid Computing Centre Karlsruhe (GridKa) entsteht ein PC-Cluster, das zur Zeit mehr als 1.000 CPUs, über 200 TB Plattenplatz und mehrere 100 TB

Magnetbandkapazitäten enthält. Diese Installation entspricht einem im Jahre 2001 von 8 deutschen Hoherenergiephysik-Gruppen aufgestellten Plan, der vom GridKa in jeweils 2 Erweiterungsstufen (Meilensteinen) pro Jahr bisher pünktlich erfüllt werden konnte. Der weitere Ausbau sieht bis zum Jahr 2007 über 4.000 Prozessoren und mehrere Petabyte Speicherplatz auf Magnetplatten und -bändern vor.

Literatur

- | | | |
|---|--|--|
| [1] http://www.gridka.de | [6] http://www.pbspro.com | [12] http://www.sc-conference.org/sc2003/infra_bwc2.html |
| [2] H. Marten, <i>Beitrag in dieser Ausgabe der Nachrichten</i> | [7] http://ganglia.sourceforge.net | [13] http://www.ibm.com/servers/eserver/clusters/software/gpfs.html |
| [3] http://www.gridka.de/hardware/hardware.html | [8] http://nagios.org | [14] http://www.globus.org/datagrid/gridftp.html |
| [4] http://rocksclusters.org/Rocks | [9] C. Windeck, <i>c't 7/2004, S. 150-161</i> | [15] R. Stoy, B. Hoeft, <i>DFN-Mitteilungen 64 (2004) 15-17</i> |
| [5] http://www.openpbs.org | [10] http://www.mpibpc.gwdg.de/inform/MpiNews/cientif/jahrg9/1.03/1a.03/scta.html | |
| | [11] http://www.hlrn.de | |

Unterstützung der Grid-Benutzer: Global Grid User Support (GGUS) – weltweit rund um die Uhr

H. Dres, G. Grein, Ra. Kupsch, R. Pietschmann, W. Thöne, IWR

Einleitung

Zu den Aufgaben eines klassischen Rechenzentrums als IT-Dienstleister gehört neben der Bereitstellung der IT-Ressourcen auch die Unterstützung seiner Benutzer durch Schulung, Beratung, Information, Problembearbeitung und -lösung.

Für das „World Wide Grid“ (WWG), bei dem die Ressourcen und Daten weltweit verteilt sind, das sich aber für den Benutzer als ein Grid-Rechenzentrum darstellt, gilt diese Aufgabenstellung gleichermaßen.

Das im Forschungszentrum angesiedelte Grid-Rechenzentrum „GridKa“ ist einerseits Produktionsumgebung für vier Experimente der Hochenergiephysik außerhalb des vom CERN organisierten Projekts „Large Hadron Collider“ (LHC), andererseits Testumgebung für die vier LHC-Experimente. In der Testphase liefern die LHC-Experimente Simulationsdaten mit zunehmendem Volumen, ab 2007 reale Messdaten in der Größenordnung von 12 bis 14 Petabyte pro Jahr. Da die Nutzer des WWG jedoch auch weltweit verteilt sind und somit Unterstützung rund um die Uhr geboten werden muss, stellen sich hierbei völlig neue Herausforderungen beim User Support, wenn man nicht den klassischen Ansatz eines Dreischichtbetriebes wählt. Hinzu kommt die spezielle Grid-Problematik, dass sowohl die Ressourcenverteilung als auch der Speicherort der Daten nicht bekannt sind und sich der Support auf völlig unterschiedliche Hardware, Software und Orte beziehen kann.

Der vorliegende Bericht soll Lösungen dazu aufzeigen, den Status quo von GGUS beschreiben und einen Ausblick in die Zukunft geben.

Historie

Die Arbeitsgruppe 5 (WG5: „Grid User Support“) des Grid Deployment Board (GDB), deren Sprecher Herr Mickel ist, stellte sich Ende 2002 die Aufgabe, ein Modell für die Unterstützung der Benutzer von LCG-1, des ersten Test-Grid-Rechenzentrums für die LHC-Experimente, zu entwerfen und Vorschläge zur technischen und organisatorischen Realisierung zu machen. Dabei waren folgende Vorgaben zu erfüllen:

- Empfehlung für ein LCG-User-Support-Modell mit Definition der Zuständigkeiten (LCG: Large Hadron Collider Computing Grid)
- Definition einer zentralen oder verteilten Call-Centre-Struktur
- Festlegung von Prozessen zum Problemmanagement und zur Kommunikation mit den Benutzern und beteiligten Supportgruppen
- Definition der Rahmenbedingungen für den Zugriff durch den Benutzer und Festlegung der Dienste, die Benutzer vom Call-Centre erwarten können (Service Level Agreements, SLA)
- Benutzerunterstützung „Rund um die Uhr“
- Entwurf eines Konzeptes zur Kommunikation des Call-Cen-

tre mit lokalen und anderen Support-Gruppen

- Empfehlungen für Prozesse zur Koordination und Verfolgung von Problemen auf der Basis eines Problemmanagement-Systems

Konzept

Mit der Übernahme der Leitung der WG5 war gleichzeitig die Verpflichtung des Forschungszentrums verbunden, im GGUS eine führende Rolle einzunehmen. Das GDB ist zuständig für das LHC-Projekt und die beteiligten Experimente. Es galt aber ein Konzept zu erarbeiten, das sich genauso auf die vier (internationalen) Nicht-LHC-Experimente anwenden ließ. Für deren nationale und internationale Nutzer war das Grid-Rechenzentrum des Forschungszentrums (GridKa) von Anfang an Produktionsumgebung.

Um die Forderung nach einer Benutzerunterstützung innerhalb des LCG an 7 Tagen in der Woche und 24 Stunden pro Tag zu erfüllen, wurde folgendes Modell (Abb. 1) vorgeschlagen:

Drei Zentren in drei verschiedenen Zeitzonen – idealerweise mit jeweils 8 Stunden Zeitverschiebung – bilden ein gemeinsames Support-Zentrum und leisten den „Global Grid User Support“ (GGUS).

Dabei ist der wichtigste Aspekt für den Benutzer, dass sich die drei Partner wie ein einziges Support-Team darstellen müssen. Der Grid-Benutzer, der ein Pro-

blem meldet, soll gar nicht wissen, welcher der drei Partner gerade aktiv ist. Das Konzept kann nur dann zum Erfolg geführt werden, wenn GGUS über ein Internet-Portal vom Benutzer kontaktiert werden kann. Für das Melden von Problemen darf es ebenso nur eine einzige Kontaktadresse geben, erreichbar über drei Zugänge – das Internet-Portal, die E-Mail-Adresse und eine einzige Telefonnummer.

Für andere im Umfeld des LCG tätige Support-Gruppen wie der experimentspezifische User Support (ESUS) oder der CERN Deployment Support (CDS) wurden Schnittstellen zu den dortigen Problemmanagementsystemen auf der Basis von XML oder E-Mail vorgeschlagen. Noch in der konzeptionellen Phase wurden Gespräche mit dem Grid Operations Centre (GOC) geführt und Übereinkunft darüber erzielt, das GGUS-Portal und seine darin enthaltenen Support Funktionen im Forschungszentrum durch das GOC direkt zu nutzen.

Das vom IWR vorgelegte Konzept wurde mit wenigen Änderungswünschen vom GDB akzeptiert und der Auftrag zur Entwicklung eines Prototypen für das Web-Portal und eines Problemverfolgungssystems auf der Basis von Remedy erteilt – zunächst für die GridKa-Benutzer.

Im Juni 2003 wurde der Prototyp im GDB erfolgreich vorgeführt. Der Auftrag für die Folgezeit war klar definiert: Ausbau des Prototypen zu einem funktionsfähigen Help-Desk-System und die Inbetriebnahme für die GridKa-Benutzer. Um die weltweite Öffnung

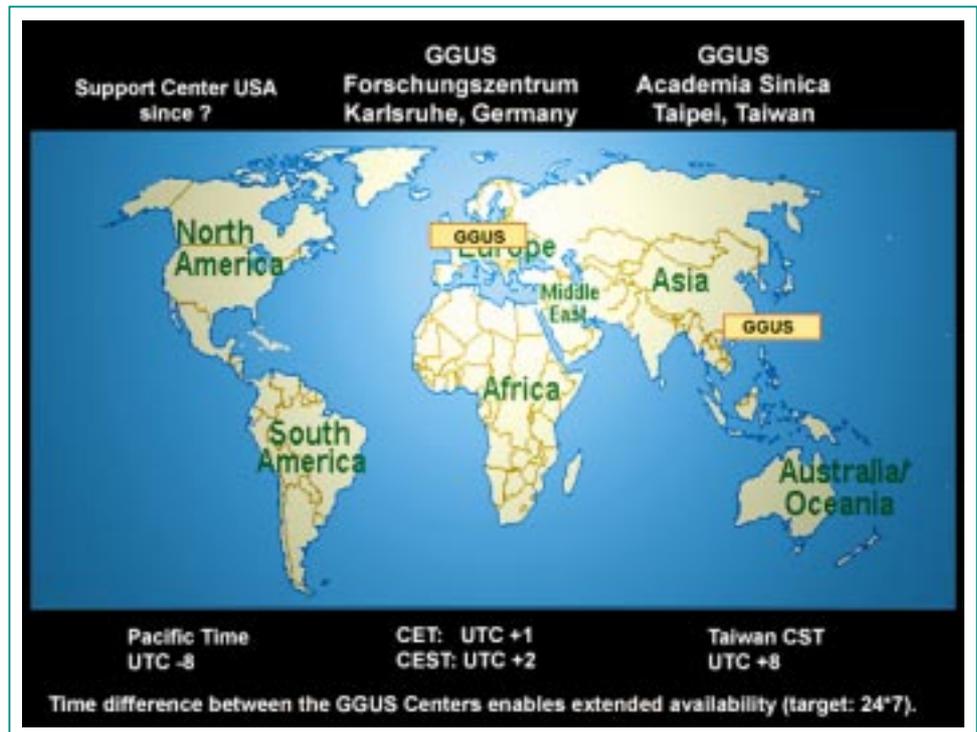


Abb. 1: GGUS-Zeitzone Modell.

des Portals von Anfang zu ermöglichen, kaufte das IWR die Internet-Domäne www.ggus.org.

Realisierung

Zur Realisierung des Konzeptes mit drei Zentren sind unterschiedliche Modelle möglich:

1. Ein Zentrum betreibt die gesamte Infrastruktur des Portals, die beiden anderen greifen über einen Web-Browser darauf zu.
2. Zwei Zentren betreiben eine möglichst gleichartige Infrastruktur und synchronisieren sich zeitnah, das dritte Zentrum nutzt die Infrastruktur eines Zentrums über einen Web-Browser.
3. Alle drei Zentren betreiben eine möglichst gleichartige In-

frastruktur und synchronisieren sich zeitnah.

Die geringste Komplexität hat das Modell 1, allerdings mit der Konsequenz, dass die Infrastruktur in einem Zentrum möglichst ausfallsicher aufgebaut sein muss. Aus Zeit- und Kostengründen wurde beschlossen, zunächst auf der Basis von Modell 1 zu beginnen und die Anwendungen für das GGUS-Portal im Forschungszentrum zu erstellen. Um eine gute Verfügbarkeit zu gewährleisten, wurde an das Problem-Managementsystem Remedy ein ausfallsicheres Oracle-Datenbank-Cluster angebunden. Das FZK-GGUS-Team hat sich nach sorgfältiger Überlegungen gegen ein Open-Source-Produkt und für Remedy ausgesprochen, da Know-how und Lizenzen bereits vorhanden waren,

Remedy ein sehr offenes System mit hoher Funktionalität ist und somit die Entwicklung des Portals rasch vorangetrieben werden konnte.

Der Prototyp für die Unterstützung der Benutzer aller acht HEP-Experimente bot folgende Funktionen:

Statusinformationen von Jobs und Jobqueues in GridKa, geplante und ungeplante Ausfallzeiten, Erstellen eines Tickets, Ticketverfolgung, News, Downloads und FAQs.

Er wurde am 2. Oktober 2003 den Vertretern der acht Experimente im Rahmen eines TAB-Meetings (Technical Advisory Board des GridKa) vorgestellt und stieß auf Zustimmung.

Im Anschluss daran wurde die erste Version des GGUS-Portals für alle GridKa-Benutzer sowie für ausgewählte Interessenten aus anderen europäischen Grid-Zentren freigegeben (Abb. 2 u. 3).

Aktueller Stand

Im Januar 2004 wurde mit dem „Academia Sinica Computing

Centre“ (ASCC) der erste Partner gefunden. Damit mussten Verantwortlichkeiten und Zuständigkeiten zwischen beiden Partner definiert werden. Die Bereiche News und FAQs sowie die Bereitstellung des einheitlichen Telefonzugs wurden in die Verantwortlichkeit von ASCC gelegt.

Die Wissensdatenbank soll auf der Basis von Remedy gemeinsam konzipiert und aufgebaut werden. Mit ASCC finden regelmäßige Videokonferenzen statt. Breiten Raum nehmen dabei Diskussionen ein, wie Probleme über Schichtgrenzen hinweg behandelt werden sollen. Seit dem 24. Mai 2004 wird GGUS im Zweischicht-Betrieb in der Zeit von 2 Uhr bis 16 Uhr MESZ betrieben.

Die in der Konzeptphase mit GOC vereinbarte Zusammenarbeit wurde intensiviert. Das GOC liefert Statusinformationen über das Grid und betreibt Monitoring über die Ressourcen. GOC kommuniziert nicht direkt mit den Benutzern, sondern mit den Administratoren der Grid-Rechenzentren. Trotzdem fließen Informationen von GOC zu GGUS und umgekehrt. Zur Meldung von Problemen nutzt GOC wie vereinbart das GGUS-Portal.

Überwiegend für lokale Administratoren wird Beratung durch den CERN Deployment Support (CDS) erbracht. Der CDS hat sich zwischenzeitlich ebenfalls bereit erklärt, die GGUS-Remedy-Lösung zur Problembehandlung mit zu benutzen.

Bei den Experimenten gibt es lokalen User-Support mit experimentenspezifischem Wissen (Experiment Specific User Support –



Abb. 2: Einstiegseite des GGUS-Portals.

ESUS). An diese lokalen Support-Teams leitet GGUS bei Bedarf Probleme zur Lösung vor Ort weiter. Die optimale Anbindung von ESUS wäre durch die Nutzung des GGUS-Portals gegeben. Das wird nicht bei allen Experimenten zu erreichen sein. Deswegen müssen Schnittstellen zu den lokalen Ticket-Systemen definiert werden, um auch hier zu gewährleisten, dass alle experimentsspezifischen Probleme ebenfalls in der GGUS-Datenbank dokumentiert werden.

Probleme bereitet zurzeit noch die Benutzerverwaltung von GGUS. Da es noch kein vollständiges (logisches) Verzeichnis aller Grid-Benutzer gibt und noch nicht alle mit Zertifikaten arbeiten, müssen die Nutzer ohne Zertifikate noch lokal bei GGUS eingetragen und verwaltet werden. Der Zugang zu den internen Bereichen des Portals wird über die lokale Benutzerverwaltung bzw. über im Browser integrierte Zertifikate ermöglicht. Es soll dadurch sichergestellt werden, dass nur Grid-User ihre Probleme an GGUS melden. Zurzeit sind mehr als 270 Benutzer eingetragen. GGUS leistet inzwischen Unterstützung für LCG-2, den Nachfolger von LCG-1 (Abb. 4).

Ausblick

Eine wichtige Zielsetzung von GGUS ist der Aufbau einer „Datenbank aller für das Grid relevanten Probleme“.

Erst nachdem eine größere Anzahl von gelösten Problemen dokumentiert ist, lässt sich daraus eine Wissensdatenbank konzi-

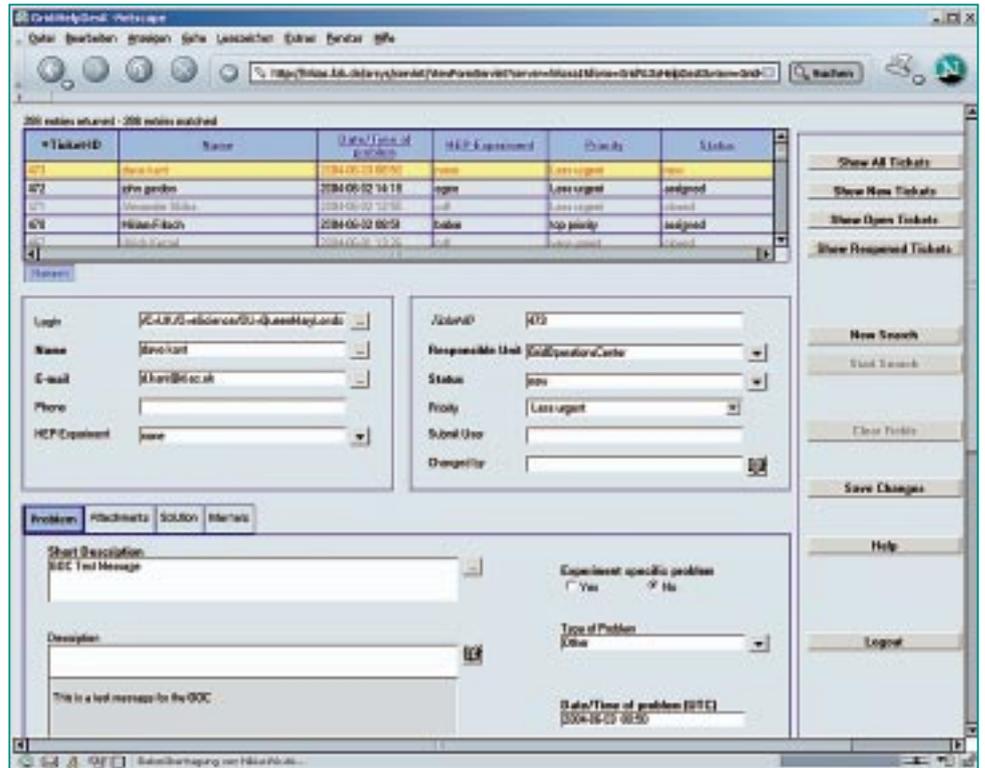


Abb. 3: Ticket-Bearbeitung.

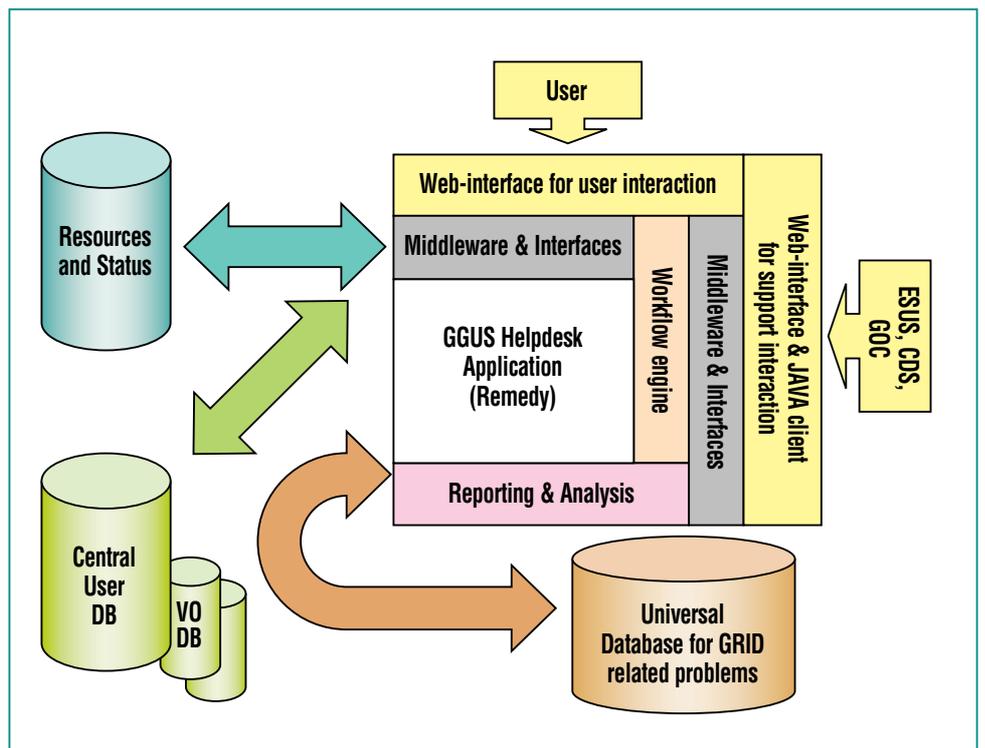


Abb. 4: GGUS-Architektur.

pieren. Relativ wenig Arbeit wurde bisher in den Aufbau der FAQs investiert. Hier sollen sowohl aus gelösten Problemen abgeleitete, selbst erstellte Beschreibungen eingestellt werden, aber auch das Grid betreffende, weltweit verteilte Dokumentationen in Form einer Link-Sammlung zusammengeführt werden. Für die News muss noch ein Konzept entwickelt werden, damit die in einem beliebigen Grid-Tier-x-Rechenzentrum oder bei GOC veröffentlichten News automatisch im GGUS-Portal publiziert werden.

Einen großen Stellenwert wird die Suche nach dem dritten Partner (möglichst an der Westküste der USA) einnehmen. Erst wenn dieser Partner gefunden und integriert ist, kann Benutzerunterstützung rund um die Uhr erbracht werden.

Eine Herausforderung stellt die Integration eines Support-Konzeptes innerhalb des EGEE-Projektes (Enabling Grid E-science in Europe) dar. Mit der Unterstützung für andere Wissenschaften und für vielfältige Anwendungen müssen bei der Klassifizierung von Problemen völlig neue Wege

beschritten werden. Darüber hinaus muss der dezentrale Ansatz der Support-Struktur im Konzeptpapier zum EGEE mit dem zentralen Ansatz von GGUS zur Deckung gebracht werden.

Verschlüsselung und Authentifizierung als Basis internationalen Vertrauens

U. Epting, IWR

Einleitung

Grid Computing gibt seinen Nutzern die Möglichkeit zum Zugang zu weltweit verteilten Rechenressourcen mit potentiell zehntausenden von CPUs. Ausgefeilte Sicherheitsmechanismen müssen deshalb für die Einhaltung vorgegebener Sicherheitskriterien sorgen. Wird über das Grid etwa ein Programm auf einen Cluster in Italien oder Taiwan geschickt, so wird der Benutzer dort in aller Regel nicht persönlich bekannt sein. Umgekehrt möchten Wissenschaftler sicher sein, ihre neuesten Forschungsergebnisse auf dem richtigen Datenserver abzulegen – und nicht etwa bei einem Hacker. Es wird also eine Methode zur sicheren, automatischen Authentifizierung von Benutzern und Servern benötigt. Das Globus-Toolkit [1], das in der GridKa-Umgebung des IWR eingesetzt wird, stellt ein Sicherheitsmodell bereit – die Grid Security Infrastructure (GSI) [2].

Die Grid Security Infrastruktur (GSI)

Public Key Kryptographie und RSA-Algorithmus

Die Sicherheitsinfrastruktur GSI des Globus Toolkits dient hauptsächlich der sicheren Authentifizierung von Benutzern oder Endgeräten. Sie bedient sich dazu der Public-Key-Kryptographie. Dieses 1976 durch Whitfield Diffie und Martin Hellman entwickelte Konzept, auch als asymmetrische Kryptographie bekannt, beruht auf mathematischen Algorithmen, mit deren Hil-

fe man sehr einfach Daten mit einem öffentlich bekannten Schlüssel verschlüsseln kann, wohingegen es fast unmöglich ist, sie ohne zusätzliche Informationen wieder zu entschlüsseln. Diese zusätzlichen Informationen bezeichnet man als privaten Schlüssel, der nur dem jeweiligen Besitzer bekannt ist [3] [4]. Öffentlicher und privater Schlüssel bilden zusammen ein Paar.

Der bekannteste in der Public-Key-Kryptographie verwendete Algorithmus ist der 1977 nach seinen Erfindern – Ron Rivest, Adi Shamir und Leonard Adleman – benannte RSA-Algorithmus. Die Sicherheit dieses Verfahrens beruht auf der Schwierigkeit, große Zahlen zu faktorisieren, d.h. in ihre Primfaktoren zu zerlegen. Zur Herstellung des Schlüsselpaars werden zwei sehr große Primzahlen p und q – mit 150-200 Dezimalstellen oder mehr – verwendet. Bei jeweils etwa 150 Dezimalstellen ergibt sich für das Produkt $N = p \times q$ ein Wert mit etwa 300 Dezimalstellen, was einer Schlüssellänge von 1024 Bits entspricht. Zusätzlich wird eine zufällige Zahl e so gewählt, dass e und $(p - 1) \times (q - 1)$ keinen gemeinsamen Teiler haben. Die beiden Zahlen N und e sind der öffentliche Schlüssel. Der private Schlüssel d wird mit Hilfe des sogenannten erweiterten Euklidischen Algorithmus berechnet und geheim gehalten. Der Klartext M wird über die ASCII-Tabelle zunächst in Binär- und dann in Dezimalzahlen verwandelt und blockweise mit der in Abb.1 dargestellten Formel berechnet. Es ergibt sich der verschlüsselte Text C – der so genannte Cipher-

text. Der Empfänger kann ihn nur dann mit der Umkehrfunktion, die in der zweiten Formel von Abb. 1 dargestellt ist, wieder entschlüsseln, wenn er d kennt oder berechnen kann.

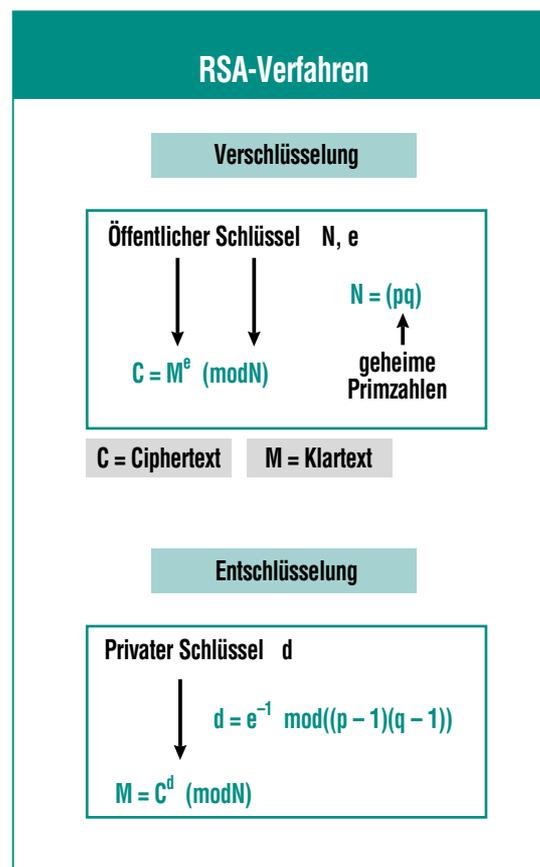


Abb.1: Der RSA-Algorithmus zur Kryptographie mit öffentlichem Schlüssel beruht auf der Schwierigkeit, sehr große Zahlen N zu faktorisieren, d.h. in ihre (geheimgehaltenen) Primfaktoren p und q zu zerlegen. Die beiden Zahlen N und e sind der „öffentliche Schlüssel“. Der Klartext M wird über die ASCII-Tabelle zunächst in Binär- und dann in Dezimalzahlen verwandelt und mit der dargestellten Formel in den verschlüsselten Text C umgerechnet. Ein Empfänger kann C nur dann mit der Umkehrfunktion entschlüsseln, wenn er den geheimgehaltenen „privaten Schlüssel“ d kennt (oder bei gelungener Faktorisierung von N berechnen kann!)

Die verwendete Modulararithmetik untersucht endliche Gruppen von Zahlen. Sie ist als „Uhrenarithmetik“ jedem aus dem Alltag bekannt: Wenn Ihnen abends um 23.00 Uhr gesagt wird, dass in 5 Stunden ein Treffen stattfindet, wissen Sie, dass das Treffen nicht um 28.00 Uhr, sondern um 4.00 Uhr nachts stattfindet. Sie berechnen dies folgendermaßen: $23 + 5 = 28$, 28 geteilt durch 24 = 1 Rest 4. Mathematisch dargestellt: $23 + 5 = 4 \pmod{24}$, also (Rest) modulus (Teiler). Dieser Ausdruck ist aber offenbar nicht eindeutig, sondern könnte auch $23 + 29$ bedeuten. Modulfunktionen werden wegen ihrer Eigenschaften als Einwegfunktionen geschätzt. Es ist sehr schwer, zu einer gegebenen Funktion die Umkehrung zu finden. [5]

Im Fall von RSA müsste man die Zahl N faktorisieren. Dies ist zwar im Prinzip möglich, aber bei entsprechend großem N sehr zeitaufwändig. Man schätzt, dass man für die Faktorisierung einer 1024-Bit langen Zahl $3 \cdot 10^{11}$ MIPS-Jahre braucht. (Ein MIPS-Jahr erfordert einen Rechner, der z. B. 1 Million Instruktionen pro Sekunde durchführen kann und ein Jahr läuft.) [6]

Das Verfahren eignet sich auch dafür, die Herkunft von Daten oder eines Textes durch eine Unterschrift (Signatur) zu bestätigen. Hierbei wird über die Daten oder den Text mit einer Einwegfunktion eine eindeutige Prüfsumme (Hashwert) ermittelt. Dieser wird sodann mit dem privaten Schlüssel des Absenders verschlüsselt und ebenfalls übermittelt. Der Empfänger entschlüsselt nun mit dem öffentlichen Schlüssel

des Absenders den gesendeten Wert und vergleicht ihn mit dem von ihm selbst berechneten Wert. Stimmen beide überein, so ist die Herkunft der Daten oder des Textes bestätigt, da keine andere Person den zum verwendeten öffentlichen Schlüssel gehörigen privaten Schlüssel besitzt.

Zertifizierung

Um sicherzustellen, dass ein öffentlicher Schlüssel wirklich zu einer bestimmten Person gehört, wurde das Konzept der Zertifizierung entwickelt. Eine dritte, vertrauenswürdige Instanz – eine Zertifizierungsstelle (Certification Authority, CA) – beglaubigt mit ihrer unabhängigen Unterschrift die Verbindung einer Identität mit einem öffentlichen Schlüssel. Ein Zertifikat enthält einen weltweit eindeutigen Namen, den so genannten Distinguished Name (DN), den öffentlichen Schlüssel der Person und zusätzliche Informationen wie Gültigkeitsdauer, Ausstellungsdatum, verwendete Algorithmen und Schlüssellängen. Ferner finden sich darin der Name der Zertifizierungsstelle sowie ihre Unterschrift, die bestätigt, dass die im Zertifikat enthaltenen Informationen korrekt sind.

Die Art und Weise, wie eine Zertifizierungsstelle die Identität der Benutzer oder Endgeräte überprüft und welche Sicherheitsvorkehrungen sie rund um die Zertifizierungsstelle trifft sind in der Certification Policy (CP)/Certification Practice Statement (CPS) festgelegt und veröffentlicht. Jede Institution, die Ressourcen für das Grid zur Verfügung stellt,

kann entscheiden, ob sie der Zertifizierungsstelle vertraut. Über veröffentlichte Widerruflisten kann ein Zertifikat schnell für ungültig erklärt werden, wenn zum Beispiel der private Schlüssel gestohlen wurde.

Die Zertifikate, die von der GSI verwendet werden, folgen in ihrem Datenformat dem X.509-Standard, der von der Internet Engineering Task Force (IETF) für das Internet standardisiert wurde und auch von anderer Software, etwa den üblichen Webbrowsern (Konqueror, Opera, Netscape, Internet Explorer), eingesetzt wird.

Zertifizierungsstelle GridKa-CA

Das Institut für wissenschaftliches Rechnen (IWR) betreibt seit Dezember 2001 die Zertifizierungsstelle GridKa-CA [7]. Hier erhalten Wissenschaftler aus 22 verschiedenen deutschen Institutionen Zertifikate, die von mehr als 65 Institutionen in insgesamt 24 Ländern akzeptiert werden. Die stetige Zunahme an Benutzern zeigt deutlich die Akzeptanz der GridKa-CA Zertifizierungsstelle des Forschungszentrums Karlsruhe. Der Betrieb einer Zertifizierungsstelle stellt besonders hohe Ansprüche an den Betrieb des zur Zertifizierung verwendeten Computers sowie den Schutz der darauf gespeicherten Daten (Abb. 2).

Die praktische Arbeit in großen, nationale Grenzen überschreitenden Projekten führte zu einem Zusammenschluss von CA-Betreibern aus verschiedenen Ländern. Diese in vielen internationalen Grid-Projekten tätige „Euro-

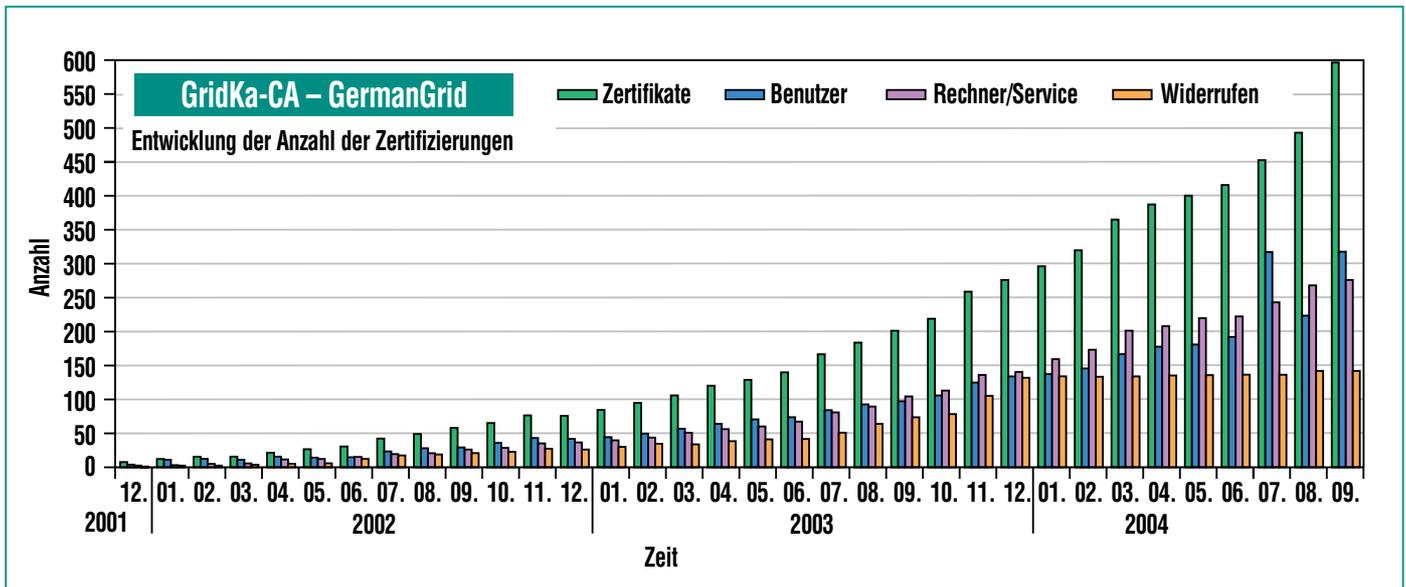


Abb. 2: Das IWR betreibt seit Dezember 2001 eine Zertifizierungsstelle. Das kontinuierliche Ansteigen der ausgegebenen Zertifikate zeigt das wachsende Interesse an Grid-Projekten in Deutschland.

pean Policy Management Authority Group for Grid-Authentication in e-science“ (EuGridPMA)[8] entwickelt für Zertifizierungsstellen Mindestanforderungen und Verfahren zur gegenseitigen Überprüfung, um eine gemeinsame Vertrauens-Domäne herzustellen. Auf diese Weise wird (projektabhängig und demokratisch) eine „Liste“ [9] [10] von vertrauenswürdigen Partner-CA’s erstellt. So ist ein Netz des Vertrauens entstanden, das die meisten europäischen Staaten, Amerika [11] und Teile Asiens [12] umfasst. Das IWR ist in dieser Gruppe seit ihrer Gründung vertreten.

Authentifizierung versus Autorisierung

Ein Zertifikat ist mit einem Personalausweis oder einem Reisepass vergleichbar. Ein Reisepass bescheinigt zwar die Identität des Besitzers, berechtigt aber nicht automatisch zur Einreise in jedes

Land, sondern oft wird zusätzlich ein Visum benötigt. Ähnlich verhält es sich mit dem Besitz eines Zertifikats: um Zugang zu Grid-Ressourcen zu erhalten muss nach der Authentifizierung die Autorisierung erfolgen. Es muss festgelegt sein, welche Rechte der Benutzer hat – z.B. wie viel Rechenzeit er in Anspruch nehmen darf, welche Dateien er schreiben oder lesen darf. Die Rechte eines Benutzers werden über die Mitgliedschaft in einer virtuellen Organisation (VO) festgelegt. (Siehe Artikel „Grid Computing – Basis of multi-institutional Virtual Organizations“ im gleichen Heft.) Durch eine Abbildung (mapping) des eindeutigen DN („Distinguished Name“) auf ein lokales Userkonto bewahrt die GSI den lokalen Administratoren die Lufthoheit über ihre Ressourcen und macht sie von einem zentral verwalteten Sicherheitssystem unabhängig.

Das Secure Socket Layer (SSL) Protokoll

Das im Internet zur Sicherung der Kommunikation häufig verwendete Secure Socket Layer Protokoll – auch unter dem neuem Namen Transport Layer (TLS) Protokoll bekannt – ist auch in der GSI implementiert und ermöglicht eine Zwei-Wege-Authentifizierung. Zwei Parteien können mittels asymmetrischer Kryptographie beweisen, dass sie diejenigen sind, die sie vorgeben zu sein. Voraussetzung ist, dass beide Parteien ein Zertifikat besitzen und dass den jeweiligen Zertifizierungsstellen vertraut wird (Abb. 3).

Dies geschieht folgendermaßen: Partei A sendet B das Zertifikat (ZertA) zu. B entnimmt ZertA, wer A ist, und überprüft zunächst die Gültigkeit des Zertifikats und die Richtigkeit der Signatur der Zertifizierungsstelle. Dies geschieht mittels des öffentlichen Schlüs-

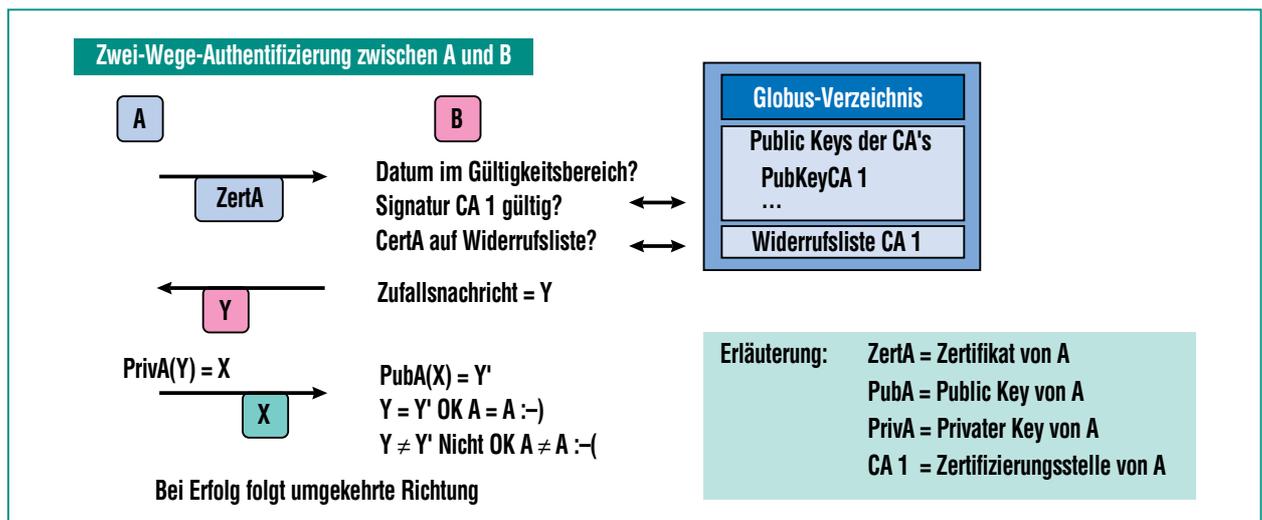


Abb. 3: Die 2-Wege-Authentifizierung zwischen den Parteien A und B erfolgt ohne direkte Absprache („stumm“): A sendet B das Zertifikat (ZertA) zu, dem B entnimmt, wer A ist. B überprüft die Gültigkeit des Zertifikats und die Richtigkeit der Signatur der Zertifizierungsstelle. Dann sendet B eine zufällige Nachricht an A. Diese wird von A mit seinem privaten Schlüssel verschlüsselt und an B zurückgesendet. B entschlüsselt die Nachricht mit dem öffentlichen Schlüssel von A und vergleicht das Ergebnis mit seiner ursprünglichen Nachricht. Stimmen beide überein, ist A der Besitzer des privaten Schlüssels, der zum beglaubigten öffentlichen Schlüssel gehört. Seine Identität ist somit bewiesen und es folgt in umgekehrter Richtung die Identifizierung von B.

sels der CA, der in einem Globus-Verzeichnis hinterlegt ist. Einem Check der Widerrufsliste wird entnommen, ob das verwendete Zertifikat zwischenzeitlich widerrufen wurde. Ist alles in Ordnung, so sendet B eine zufällige Nachricht zu A. Diese wird nun von A mit dem privaten Schlüssel verschlüsselt und zu B zurückgesendet. B entschlüsselt die Nachricht mit dem öffentlichen Schlüssel von A und vergleicht das Ergebnis mit der ursprünglichen Nachricht. Stimmen beide überein, so besitzt A den passenden privaten Schlüssel – seine Identität ist somit bewiesen.

Der Vorgang wird wiederholt, wobei nun B sein Zertifikat an A sendet. Sollte gewünscht sein, dass die gesamte Kommunikation verschlüsselt wird, werden anschlie-

ßend die Algorithmen festgelegt und ein symmetrischer Sitzungsschlüssel erzeugt, mit dem die Daten verschlüsselt werden. Aus Performancegründen (Übertragung etlicher GBits pro Sekunde) wird im Grid-Umfeld darauf jedoch verzichtet.

Schlussbemerkung

Dieses Jahr konnte ein Team der Universität Bonn mit Unterstützung des Instituts für Experimentelle Mathematik in Essen und des Bundesamts für Sicherheit (BSI) erstmals eine 576 Bit lange Zahl – das sind 174 Dezimalstellen – faktorisieren [13]. Dennoch schätzen Experten, dass Schlüssel mit einer Länge von 1024 Bit – die im Grid-Umfeld als Mindestlänge betrachtet wird – noch bis zum Jahre 2020 sicher sein werden.

Jedoch hängt die Sicherheit von RSA zum einen von der Entwicklung der Mathematik ab: Bisher ist es nicht gelungen zu beweisen, dass es keinen Algorithmus für die Primfaktorzerlegung gibt, wenn auch Mathematiker seit 2000 Jahren verbissen danach suchen. Fände aber morgen ein Mathematiker solch einen Algorithmus, so wäre RSA für alle Zeit unbrauchbar geworden. Andererseits ermöglicht auch der Fortschritt in der Computertechnik immer kürzere Berechnungszeiten. So könnte man auch das Grid mit seinen riesigen Ressourcen zur Faktorisierung großer Zahlen verwenden und so die sichere Basis, auf der es jetzt ruht, zerstören, frei nach dem Motto:

„Die Revolution ist wie Saturn, sie frisst ihre eigenen Kinder.“ [14]

Literaturverzeichnis und weiterführende Informationen

- [1] *Globus Software*,
<http://www.globus.org>
- [2] *GSI*,
<http://www.globus.org/security/overview.html>
- [3] Bruce Schneier,
Angewandte Kryptographie, Protokolle, Algorithmen und Sourcecode in C, Addison-Wesley, 1996, S. 525 - 540
- [4] *Simon Singh*,
Geheime Botschaften, Die Kunst der Verschlüsselung von der Antike bis in die Zeiten des Internet, Carl Hanser Verlag, München Wien, 2000
- [5] Bruce Schneier,
ibid. S. 288 – 290
- [6] Bruce Schneier,
ibid. S. 187
- [7] <http://grid.fzk.de/ca>
- [8] *EuGridPMA*,
<http://www.eugridpma.org>
- [9] *European Data Grid Project*,
<http://marianne.in2p3.fr/datagrid/ca/ca-table-ca.html>
- [10] *CrossGrid Projekt*,
<http://www.crossgrid.org>
- [11] www.gridpma.org bzw.
<http://www.doe grids.org/pages/doegr idspma.html>
- [12] <http://www.apgridpma.org/>
- [13] <http://www.proka.de/start.htm?ipool/inform/krypto/rsa576.htm>
- [14] Georg Büchner,
Dantons Tod, Reclam Verlag, Ditzingen, 1989, S. 22

Verteilte Simulationen auf einem Grid

K.-U. Stucky, S. Halstenberg, W. Jakob, A. Quinte, W. Süß, IAI

Einleitung

In Technik und Wissenschaft sind computergestützte Simulationen und Optimierungen heute zu einer wesentlichen, nicht mehr wegzudenkenden Grundlage geworden. Ihr Einsatz erlaubt in sehr frühen Entwicklungsphasen eines Produkts die Überprüfung von technischen Konzepten und Designs. Eine große Anzahl von Entwurfsvarianten kann ohne kosten- und zeitaufwändigen Bau von Prototypen bewertet werden. Die exponentiell wachsende Rechnerleistung und die ständige Fortentwicklung der Software-Werkzeuge bestimmen die Dynamik auf diesem Gebiet.

Bei Simulationen handelt es sich um die Beschreibung komplexer Systeme auf Basis von parametrierbaren Modellen und die Berechnung des Systemverhaltens mithilfe mathematischer Formulierungen. Optimierungen verfolgen das Ziel, aus einer Menge von Alternativen die unter Berücksichtigung bestimmter Kriterien beste Lösung zu finden, mathematisch entspricht das der Minimierung bzw. Maximierung einer gegebenen Funktion unter Randbedingungen.

Im IAI konnte in den letzten Jahren viel Know-how auf diesem Gebiet erarbeitet werden. Dabei standen das Design von Mikrosystemen und die Medizintechnik im Vordergrund. Die Finite-Elemente-Methode (FEM) zur Simulation der Strukturmechanik, Verhaltenssimulationen auf Basis von Netzwerkmodellen (Makromodellierung) und die Simulation optischer Abbildungen sind hier zu nennen. Evolutionäre Algorithmen

[1] bilden die Grundlage für die Lösung verschiedener Optimierungsaufgaben.

Die Einsatzmöglichkeiten von Simulation und Optimierung sind stark von der Verfügbarkeit von Rechenleistung und anderer Computerressourcen abhängig. Die Genauigkeit von Vorhersagen wird bestimmt von der Genauigkeit der Modelle. Verbesserungen bedeuten hier meist ein Anwachsen der Datenmengen und eine Zunahme der Berechnungen während der Simulation. Ebenso wachsen die Modellgrößen mit der Komplexität der zu simulierenden Systeme, was ebenfalls größere Datenmengen und höhere Anforderungen an die Software bedingt. Parameterstudien und der Wunsch nach mehr und komplizierteren Optimierungen stellen ebenfalls steigende Anforderungen an Ressourcen, sowohl quantitativ als auch qualitativ.

In dieser Situation bietet die Technologie des Grid Computing neue Möglichkeiten. Bei einem Grid handelt es sich um ein System von standardisierten, allgemein einsetzbaren Protokollen und Schnittstellen zu Servicediensten (Services) in einem weiträumigen heterogenen Computernetzwerk. Hauptaufgabe besteht darin, über die Dienste einen koordinierten Zugriff von Anwendern auf Ressourcen zu ermöglichen. Sogenannte virtuelle Organisationen sind Gruppen von Anwendern oder Institutionen, die über die Grid Services verteilten Zugriff auf bestimmte Ressourcen haben [2]. Ressourcenmanagement, Informationsdienste und ausreichende Sicherheitsmechanismen sind Be-

reiche, in denen Grid Services bereitgestellt werden.

Open Grid Services Architecture (OGSA) ist ein Framework für Grid Standards [3]. Wichtige Gridsoftware sind das Globus-Toolkit [4], Condor [5] und UNICORE [6]. Gridsysteme, die speziell für den Bereich Simulation und Optimierung entwickelt wurden sind zum Beispiel Geodise [7] und Nimrod/G [8].

Global optimierender Resource Broker

Die Servicedienste eines Grids werden auch Middleware genannt, da sie eine anwendungsunabhängige Ebene zwischen den Applikationen und den Ressourcen bilden. Hinsichtlich der Bereitstellung von Informationen kann ein Grid mit dem Internet verglichen werden. Es bietet aber darüberhinaus weitere grundlegende Ressourcen; die wichtigsten Typen sind Rechenleistung, Speicherkapazität, Daten und Software.

Die Entwicklungen im IAI gehen von dem in Abb. 1 dargestellten Gridaufbau aus:

- Der Zugang zu Anwendungen erfolgt über Gridportale. Die Anwendungen selbst werden als Workflows beschrieben, die aus einzelnen, verteilt im Grid zu bearbeitenden Jobs bestehen. Den Jobs sind Daten und Beschreibungen der benötigten Ressourcen zugeordnet.
- Die Ressourcen sind sogenannten Worknodes (Hardware) zugeordnet. Derzeit wer-

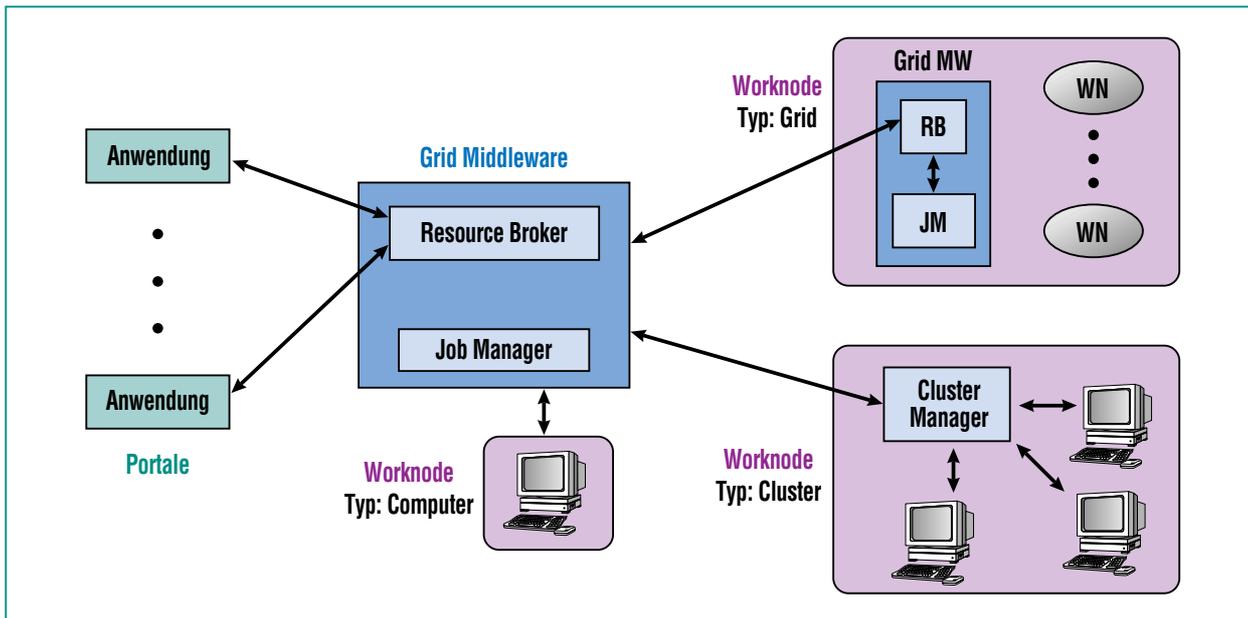


Abb. 1: Gridaufbau mit Middleware-Schwerpunkt Ressourcenmanagement. Gridanwendungen haben Zugriff auf unterschiedliche Ressourcentypen, die im Grid auf mehrere Arten von Worknodes verteilt sind.

den drei Typen unterschieden: Einzelrechner, Cluster mit einem individuellen Clustermanagement und – rekursiv definiert – komplette Grids.

- Im Zentrum der Middleware befindet sich das Ressourcenmanagement. Andere Services – etwa für die Sicherheit – sind nicht dargestellt. Der Resource Broker ist für die Planung und Festlegung der Ressourcenverteilung zuständig, während der Job Manager die ausführende Komponente ist. Grundlage für die Funktion beider Komponenten ist der Workflow.

Die verteilte Ressourcennutzung ist das bestimmende Kennzeichen eines Grids. Deshalb widmet sich ein wesentlicher Teil der Grid-Arbeiten im IAI dem Thema Resource Broker. Aus den früheren Arbeiten zur Optimierung mit

Evolutionären Algorithmen ist das allgemein einsetzbare Werkzeug HyGLEAM [1] hervorgegangen. Es kombiniert den Evolutionären Algorithmus GLEAM [9] mit lokalen Suchverfahren und erzielt in verschiedenen Anwendungen gute Ergebnisse. Aufbauend auf HyGLEAM wurde das Konzept eines global optimierenden Resource Brokers entwickelt, um den Gridanwendern im konkurrierenden Zugriff auf verfügbare Ressourcen eine optimale Zuteilung zu ermöglichen.

Die ersten zu realisierenden Optimierungsziele sind in Bezug auf die Ressource Rechenleistung die Zeiten und Kosten für Berechnungen im Grid. Dabei wird die Sicht des Anwenders verfolgt. So hat etwa bei den Kosten der Anbieter einer Ressource eher gegensätzliche Optimierungsziele. Diese wären dann aber lokal beim

Anbieter und nicht global zu verfolgen. Ähnliche Zielsetzungen hat auch der Resource Broker, der für Nimrod/G [8] entwickelt wurde. Der Resource Broker des IAI soll jedoch alle Optimierungsziele und alle noch nicht begonnenen Jobs berücksichtigen und dabei den gesamte Suchraum und nicht nur eine lokale Umgebung der aktuellen Lösung betrachten (Globale Optimierung). Außerdem soll er auf weitere Optimierungsziele und andere Ressourcentypen ausgedehnt werden.

Abb. 2 zeigt die Architektur für das Ressourcenmanagement:

- Der Resource Broker muss zunächst den Workflow der Anwendung geeignet zerlegen.
- Über eine Planungssteuerung erfolgt zuerst eine konventio-

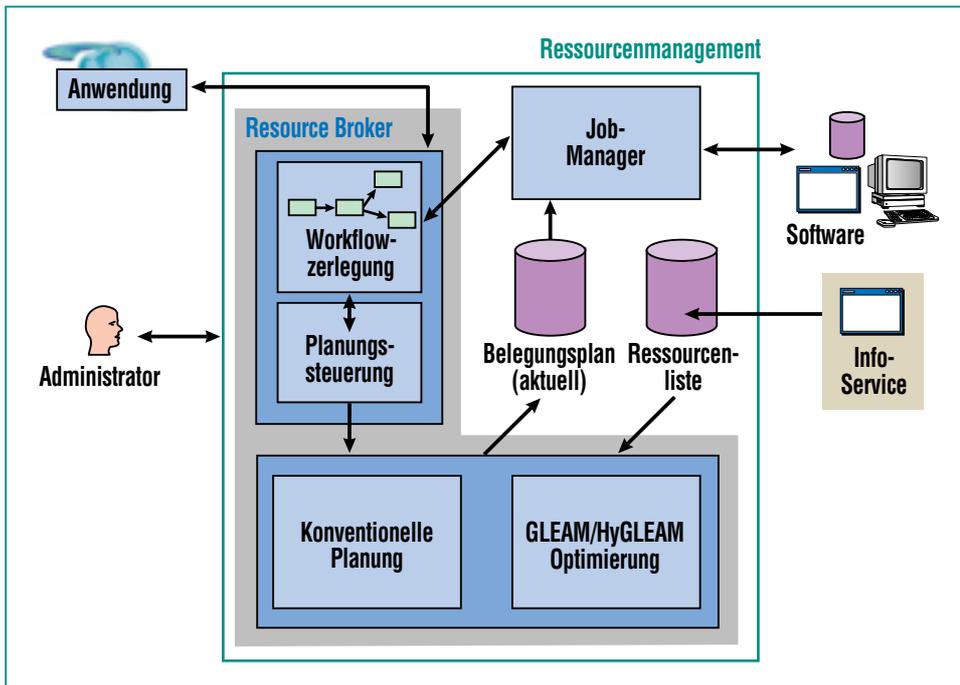


Abb. 2: Aufbau des Ressourcenmanagements. Zentrale Bedeutung hat ein global optimierender Resource Broker, der ausgehend von der Workflowdarstellung der Anwendungen eine optimale Ressourcenverteilung ermöglicht. Die Optimierungen basieren auf dem hybriden Evolutionären Algorithmen HyGLEAM. Für die Umsetzung ist ein Job Manager zuständig.

- Die unter dem ständig wechselnden Zustand des Grids im laufenden Betrieb aufwändige globale Optimierung erfolgt nur bei Bedarf. Sie wird umso lohnender sein, je mehr disponible Jobs im Verhältnis zu den verfügbaren Ressourcen vorhanden sind.
- Informationsdienste liefern notwendige Daten zu den verfügbaren Ressourcen.
- Planungsergebnisse werden im Belegungsplan eingetragen.

- Der Job Manager ist für die Ausführung auf der Grundlage des Workflows und des Belegungsplans zuständig. Er überwacht die Ausführung und sorgt auch für den Datentransfer.
- Wichtig ist zudem, dass der gesamte Ablauf in der Regel automatisch erfolgen kann. Eine umfangreiche Steuerung durch den Anwender, wie bei anderen Lösungen vorgesehen, soll es nicht geben. Daher ist in Abb. 2 auch nur für den Systemadministrator ein direkter Zugriff auf den Resource Broker dargestellt. Der Nutzer arbeitet nur über das Gridportal mit seiner Applikation. Gridspezifische Aufgaben sind für ihn transparent.

Referenzapplikation „Virtuelles Auge“

Beim „Virtuellen Auge“ handelt es sich um ein im IAI entwickeltes Softwaresystem zur Planung chirurgischer Eingriffe an der Hornhaut des Auges [10]. Grundlage des Systems sind biomechanische FEM-Simulationen und Optik-Simulationen. Die verwendeten Werkzeuge sind ANSYS [11] für die Biomechanik und SOLSTIS [12] für die Optik. Ursprünglich wurde das „Virtuelle Auge“ als Web Service realisiert. Die Anwenderschnittstelle basiert auf Java Servlets und JSP, der Zugang für den Anwender erfolgt über das Internet. Die Kommunikation zwischen einzelnen Systemteilen, die auf verschiedenen Rechnern installiert sind, ist über XML und SOAP realisiert.

Eine Referenzapplikation für den Aufbau eines Simulationsgrids dient sowohl zum Test der Ideen der Gridentwickler als auch zur Demonstration des Erreichten. Das „Virtuelle Auge“ ist aus folgenden Gründen bestens geeignet:

- Es benötigt alle wichtigen Ressourcentypen: Rechenleistung für Simulationen je nach erforderlicher Genauigkeit, Speicherkapazitäten für Modelle und Ergebnisarchivierung sowie Anwendungssoftware wie etwa Simulatoren.
- Die Voraussetzungen für eine weiträumige Verteilung sind gegeben mit Anwendern in Kliniken und Firmen, die häufig auch zusammenarbeiten und oft nicht jeder für sich teure Softwaresysteme oder ausrei-

chende Speicherkapazitäten vorhalten können.

- Einzelne Systemteile laufen unter verschiedenen Betriebssystemen (UNIX/Linux, Windows), es wird verschiedene Datenbanksysteme geben und sowohl Einzelrechner als auch Cluster werden einbezogen. Damit ist das System unter verschiedenen Blickwinkeln heterogen.
- Bei einem weiteren Ausbau wird das „Virtuelle Auge“ zu einem Operations- und Therapieplanungssystem für den gesamten Sehapparat. Ferner ist ein Einsatz in der klinischen und industriellen Forschung denkbar sowie eine Integration in weitere Simulationssysteme der Humanmedizin. Das System und seine Anwender erfüllen damit die Voraussetzungen für eine virtuelle Organisation.
- Das Arbeitsfeld Gridportale ist abgedeckt. Bereits jetzt wird intensiv an der Thematik Visualisierung gearbeitet. Komfortable Möglichkeiten sind hier entscheidend für die Akzeptanz der Endanwender.
- Nicht zuletzt ist der direkte Kontakt zu den Entwicklern der Anwendung im IAI und über diese zu einer großen Zahl potentieller Anwender ein wichtiger Vorteil bei der Realisierung anspruchsvoller Gridbausteine wie etwa dem global optimierenden Resource Broker.

Eine erste Gridversion des „Virtuellen Auges“ wurde realisiert. Sie erlaubt die Durchführung be-

stimmter Parameterstudien in der Biomechanik. Basis ist das Globus-Toolkit in der Version 2.0 und als Programmierschnittstelle Java CoG; passend zur Programmiersprache Java, in der ein erster Prototyp eines Job Managers realisiert wurde. Die Wahl einer objektorientierten Programmiersprache wird auch dem serviceorientierten Ansatz im Grid Computing gerecht.

Aktuell wird an einer umfassenden Gridanwendung „Virtuelles

Auge“ gearbeitet, die Anwender mit einbezieht und damit zum ersten Mal das lokale Testgrid verlässt. Abb. 3 vermittelt einen Eindruck des Gridsystems mit verteilten Ressourcen und mehreren Anwendern, die zusammen eine virtuelle Organisation bilden.

Ausblick

Die Entwicklungen im Grid Computing des IAI werden zunächst von der Referenzapplikation „Virtuelles Auge“ bestimmt sein.

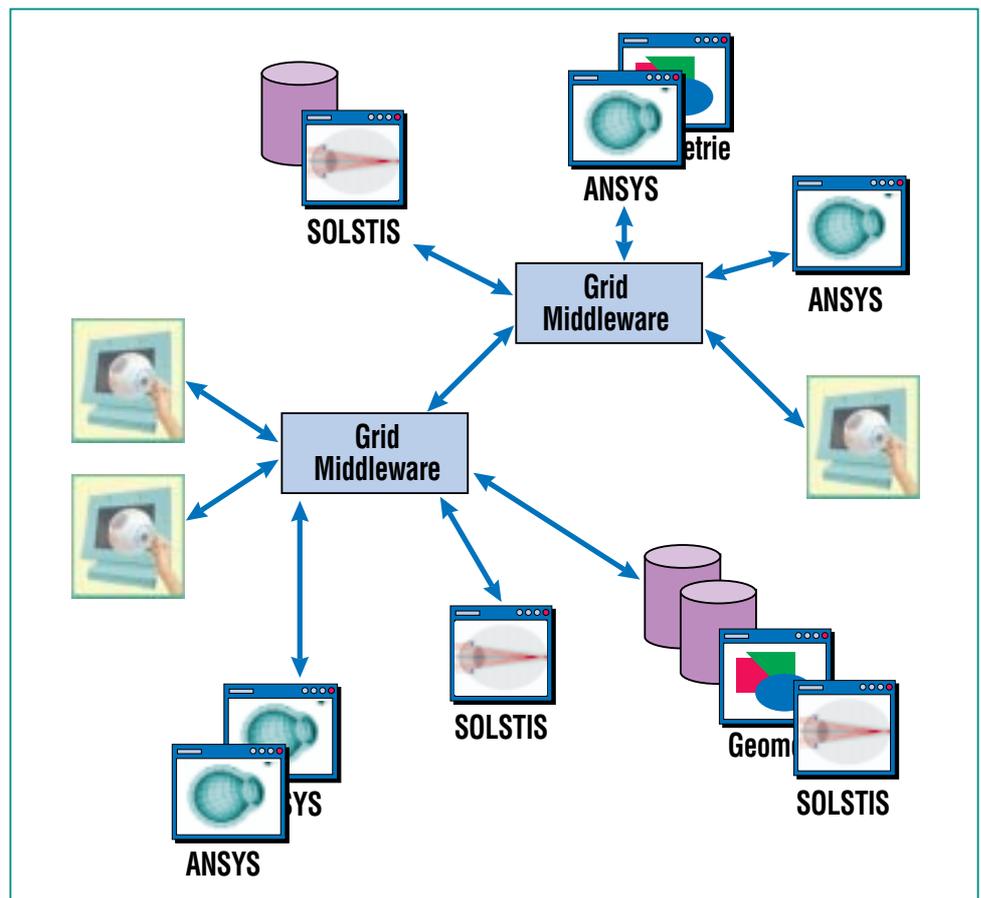


Abb. 3: Das „Virtuelle Auge“ im Grid. Dargestellt sind Softwareressourcen, ANSYS für die Biomechanik und SOLSTIS für die Optik, sowie Datenbanken. Die Ressource Rechenleistung steht überall zur Verfügung. Die Gridsoftware selbst ist kein einzelnes Programm sondern hat ebenfalls einen verteilten Charakter in Form von Services. Mehrere Anwender teilen sich die vorhandenen Ressourcen.

Kontakte zur klinischen Forschung und zu Firmen zeigen die Notwendigkeit für eine verstärkte Beschäftigung mit Möglichkeiten zur Parallelisierung. Für Parameterstudien und Optimierungen müssen einfach zu handhabende Simulatoren vorliegen, die vielfach installiert auf Clustern und Einzelrechnern paralleles Simulieren ermöglichen. Je nach Genauigkeitsbedarf kann auch die Parallelisierung einzelner FEM-Simulationsläufe im Grid zu ei-

nem wichtigen Schwerpunkt werden.

Nach der Implementierung des global optimierenden Resource Brokers mit den vorerst anvisierten Optimierungszielen Zeit und Kosten ist als nächstes Ziel die Optimierung von Datenbankzugriffen, also eine Erweiterung um den Ressourcentyp Speicherkapazität, abzusehen.

Neue Referenzapplikationen auf dem Gebiet der Designoptimie-

rung oder der Strukturbestimmung von Biomolekülen werden diskutiert. Paralleles Rechnen für Optimierungen basierend auf Evolutionären Algorithmen wird hier angestrebt.

Literatur

- [1] W. Jakob, *Dissertation, Universität Karlsruhe, Bericht FZKA-6965, 2004.*
- [2] I. Foster, C. Kesselman, S. Tuecke, *Intern. Journal of Supercomputer Applications, Vol 15(3), 2001, 200-222.*
- [3] I. Foster, C. Kesselman, J. Nick, S. Tuecke, *Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002.*
- [4] I. Foster, C. Kesselman, *in I. Foster, C. Kesselman, eds., The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 1999, 259-278.*
- [5] D. Thain, T. Tannenbaum, M. Livny, (www.cs.wisc.edu/condor/doc/condor-practice.pdf), *erscheint 2004.*
- [6] D.W. Erwin, *Concurrency Computat.: Pract. Exper., 2002, 14, 1395-1410.*
- [7] G. Xue, M.J. Fairman, G.E. Pound, S.J. Cox, *In: H. Kosch, L. Böszörményi, H. Hellwanger (eds.), Proc. of Euro-Par 2003. Lecture Notes in Computer Science, Springer-Verlag, 2003, 357-365.*
- [8] D. Abramson, J. Giddy, L. Kotler, *International Parallel and Distributed Processing Symposium (IPDPS), pp 520-528, Cancun, Mexico, May 2000.*
- [9] C. Blume, *Proc. of PPSN I, LNCS 496, Springer-Verlag, S. 48-54, 1991.*
- [10] H. Eggert, H. Guth, A. Hellmann, K.P. Scherer, P. Stiller, W. Süß, *in Proceedings of the IASTED International Conference on Internet and Multimedia Systems and Application, Honolulu, Aug. 2001, Acta Press, 148-153.*
- [11] ANSYS, Inc. (www.ansys.com).
- [12] OPTIS (www.optis-world.com).

Grid-Computing in Echtzeit

W. Eppler, T. Müller, V. Hartmann, S. Tchilingarian, A. Augenstein, M. Naji, H. Gemmeke, IPE

Einleitung

Unter Grid-Computing versteht man das Zusammentreffen von hoher Rechenleistung, großen Speicherkapazitäten, breiter Vernetzung und Integration von Diensten. Die Dienste (engl: services) spielen im folgenden Beitrag (wie meist auch in der Industrie) die Hauptrolle. Die Integration von Diensten bedeutet eine Vereinheitlichung von Softwarekomponenten bei teilweise ganz unterschiedlicher Hardware und unterschiedlichen Betriebssystemen, so dass praktisch alles, was über das Grid angefordert wird, nach demselben Schema abläuft. Erreicht wird dies durch Schnittstellen, die dafür sorgen, dass ein Benutzer auf eine Anfrage nicht, wie derzeit im Internet mit der Suchmaschine Google, unsinnige Daten gleichrangig neben sinnvollen Daten zurückerhält, sondern über einen rechnergeführten Dialog gezielt zum gewünschten Ergebnis geführt wird. Die Integration von Datenstrukturen ist daher ein wichtiges Ziel, das von zwei Seiten her angegangen wird: einerseits wird bei neuen Entwicklungen darauf geachtet, dass auf bestehenden Datenstrukturen aufgesetzt wird und anwendungsspezifische Datentypen daraus abgeleitet werden. Andererseits werden Mechanismen geschaffen, die Daten mit unterschiedlichen Formaten ineinander überführen können.

Das Global Grid Forum (GGF) [1] ist eines der wichtigsten Konsortien zur Standardisierung des Grids. Der bisher noch nicht verabschiedete Standard OGSA (Open Grid Services Architecture)

[2] definiert die grundlegenden Dienste im Grid (zu Benutzermanagement, Finden von Diensten, Datenzugriff, ...). Für den einheitlichen Datenzugriff gibt es die Arbeitsgruppe DAIS (Data Access Integration Services), die einen Standard unter gleichem Namen erarbeitet. Wichtige Vorarbeiten werden von einem schottischen Konsortium aus Universitäten und den Firmen IBM und Oracle im Rahmen des Projektes OGSA-DAI [3] geleistet. Resultat ist bisher eine Programmierumgebung für Datenzugriffe im Grid, welche auf der bekanntesten Grid-Middleware Globus aufbaut [4]. Als Middleware wird im Grid die Software-Schicht bezeichnet, die zwischen Betriebssystem und Anwendungsprogramm für Plattformunabhängigkeit und die Bereitstellung von Grid-Diensten sorgt.

Diese Arbeiten reichen noch nicht aus, um Forschungsprojekte wie das Neutrinoexperiment KATRIN [5] auf dem Grid zu realisieren, weil bisher wenig Arbeit aufgewendet wurde, das Grid echtzeitfähig zu machen. Mit Echtzeit ist die strikte Einhaltung von Antwortzeiten gemeint, die ein Computer benötigt, um auf eingehende Daten zu reagieren. Wenn z.B. ein Computer das Signal eines Temperatursensors zu spät an ein Kontrollsystem weiterleitet, weil er durch andere Aufgaben überlastet ist, kann es zu gefährlichen Systemzuständen kommen. Da im Grid die verschiedenen Rechenressourcen üblicherweise über das nicht echtzeitfähige Internetprotokoll TCP/IP vernetzt sind, gibt es hier Handlungsbedarf.

Anwendungen als Motivation

Grid-Computing als momentaner Schrittmacher moderner Informationstechnologien ist eindeutig anwendungsgetrieben. Es erhält seinen Reiz durch die Generalisierung von Komponenten, die einzelnen Anwendungssystemen entspringen und auf einmal auf vielfältige Weise in anderen Systemen verwendet werden oder zumindest mit ihnen interagieren können. Ohne Blick auf die einzelnen Anwendungen könnten die Anforderungen an das Grid nie formuliert werden.

Das IPE ist für die Messwerterfassung, die Datenakquisition, die Systemkontrolle, die Datensichtung und die Datenanalyse einer Reihe von Experimenten im Forschungszentrum zuständig. Dazu zählen Projekte aus der Teilchenphysik, Fusion, Krebsforschung und Proteomik. Bei diesen Projekten sind meist die gleichen Aufgaben zu lösen, allerdings unter verschiedenen Umgebungen und unterschiedlichen Anforderungen. Bisher ist es nicht gelungen, ein einheitliches System zu finden, das alle diese Aufgaben löst. Es ist aber äußerst unökonomisch, bei jeder neuen Systementwicklung wieder von vorne beginnen zu müssen. Unabhängig von den übrigen Vorteilen des Grid mit seinen fast unbegrenzten Rechen-, Speicher- und Netzwerkressourcen ist sein Bestreben zur Vereinheitlichung von Software-Diensten genau das Richtige, um modulare und wiederverwendbare Software-Komponenten für die zuvor erwähnten Anwendungen zu entwerfen. Zum ersten Mal ist es mit

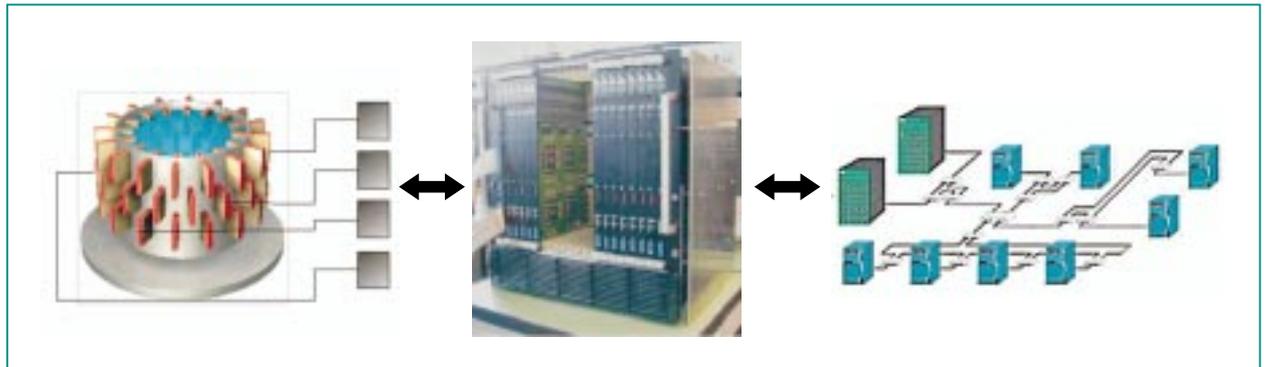


Abb. 1: Ultraschallcomputertomograph mit den Komponenten Sensortopf, Auswerteelektronik und Grid. Die Auswerteelektronik wurde ursprünglich für das Teilchenphysikprojekt AUGER entwickelt.

Hilfe von Grid-Technologien möglich, Systemteile, die beispielsweise für das Projekt TOSKA [6] entworfen wurden, für die Projekte KATRIN [5], AUGER North [7] und für einen Ultraschallcomputertomographen (USCT) [8] zur Brustkrebsuntersuchung zu benutzen (Abb. 1). In allen diesen Fällen verwenden wir eine Datenbank, einen Web-Applikationsserver und

ein Kontrollsystem, die über einheitliche Schnittstellen und einen Standard-Datentransfer miteinander kommunizieren (Abb. 2). Der Übergang zu den neuen Grid-Technologien ist für diese Aufgaben mit vernünftigem Aufwand machbar, da diese sich stark auf die von uns bereits bisher verwendeten Internet-Technologien beziehen. Weitere Anwendungen

des IPE, die mit denselben Technologien behandelt werden, kommen aus der Proteomik (Vorhersage der dreidimensionalen Proteinstruktur durch Optimierungsverfahren – zusammen mit INT und IAI [9]; Protein-Protein-Interaktionen, insbesondere bei der Hefe – zusammen mit ITG [10]; und Proteinanalytik im Chip-Format – zusammen mit IFIA [11]).

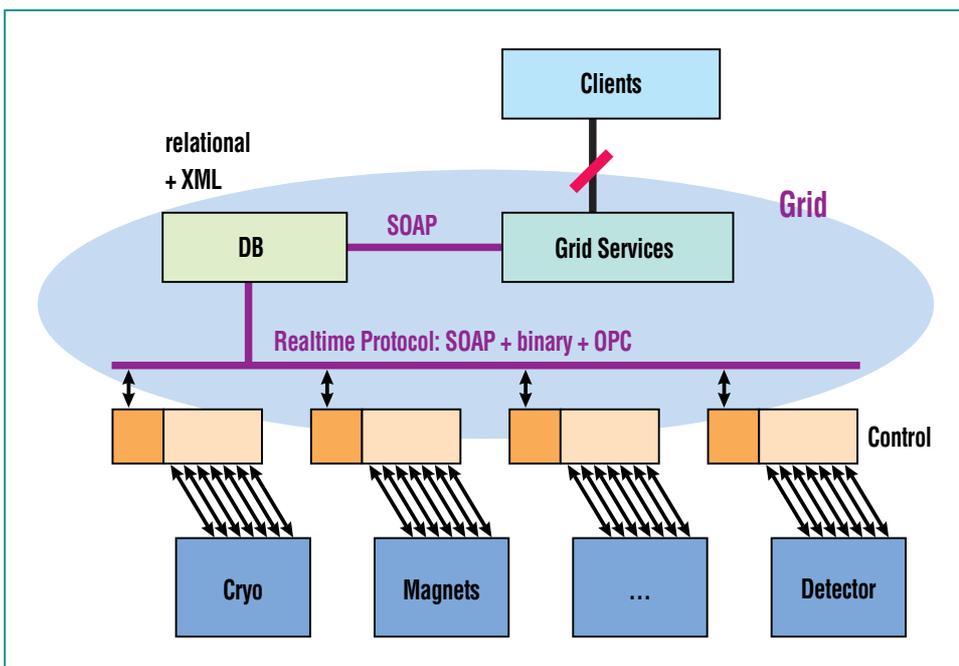


Abb. 2: Grid-Komponenten mit Subsystemen (hier für KATRIN), Kontrollsystem, Datenbank und Web-Applikationsserver.

Herausforderungen für die Forschung

Die Forschungsaktivitäten des IPE konzentrieren sich auf Erweiterungen der derzeitigen Grid-Middleware um Echtzeit-Komponenten. Ziel ist dabei, auf entstehende Spezifikationen einzuwirken bzw. entsprechende Ergänzungsspezifikationen für Echtzeit-Anwendungen in die entsprechenden Gremien einzubringen. Eine gute Basis dafür ist die Spezifikation OGSA-DAIS, welche sich zur Zeit im Diskussionsstadium befindet (siehe Abb. 3). IPE schlägt dafür eine Erweiterung vor (Spezifikation für Echtzeit-Datenintegration: OGSA-R(ealtime)DAIS), die sich bereits in Arbeit befindet und folgende Punkte umfasst:

1. Schnelles Standard-Protokoll für den Datentransfer

– basierend auf den Grid-Standards OGSA [2] und WSRF [12]

– basierend auf dem OPC-XML-Standard [13]

– basierend auf den binären Transportprotokollen XDR [14] und PBIO [15]

2. Datenschema für Experimentdaten (Primärdaten, relational) und Verwaltungsdaten (Metadaten, XML [16]) getrennt, aber durch einheitlichen Zugriff

3. Einbinden von Legacy-Systemen (Altsysteme) in das Grid durch standardisierte Datentransformationen (sog. Stylesheet-Transformationen XSLT [17] und offene Frameworks wie Struts [18] und Chiba [19])

4. Einbinden von Arbeitsplatzrechnern in Virtuelle Organisationen (VO) [2] durch den Benutzer

Im Folgenden werden diese Aktivitäten erläutert.

1. Datentransferprotokoll

Das derzeit einzige Datenprotokoll GridFTP [20] muss ergänzt werden, da es nicht echtzeitfähig und nicht ereignisgesteuert ist und somit für asynchron gesendete kleine bis mittelgroße Datenpakete nicht in Frage kommt. Interessant ist die Kombination von Standards, die aus der Automatisierungsindustrie kommen (OPC XML) [13], mit dem Web Service Standard SOAP [21] und der Erweiterung auf binäre Daten, um so ein standardisiertes, flexibles

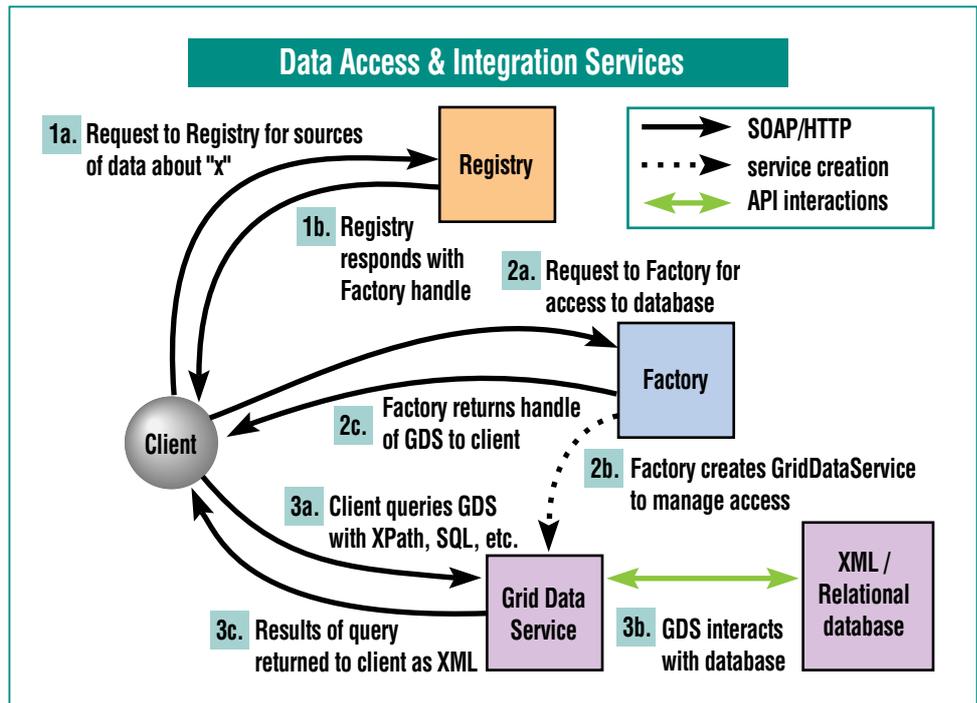


Abb. 3: DAIS-Ablauf für eine Datenbank-Abfrage.

und dennoch äußerst effektives Datentransferprotokoll zu schaffen. Weitere Details hierzu finden sich in [22].

2. Einheitlicher Datenzugriff

In praktisch allen relevanten Anwendungen, die mit Datenbeständen zu tun haben, gibt es die Unterscheidung in Primär- und Metadaten (bzw. Nutz- und Verwaltungsdaten). Die Metadaten beschreiben dabei die Struktur der Primärdaten und geben ihnen einen gewissen Sinn. Bei technischen Experimenten entsprechen die Primärdaten den Messdaten und die Metadaten sowohl den Parameterdatensätzen von Gerätekonfigurationen als auch Verwaltungsinformationen zu Benutzern oder Sicherheitseinstellungen. Die Primärdaten sind in der Regel homogen strukturiert, die Metadaten sehr heterogen. Man

spricht von strukturierten bzw. unstrukturierten Daten [23], was aber irreführend ist, da z.B. Texte sehr wohl eine gewisse Struktur aufweisen. Mit (homogen) strukturierten Daten sind in Tabellen organisierte Datenreihen gemeint, die in relationalen Datenbanksystemen mit SQL (Structured Query Language) als Zugriffssprache sehr effizient organisiert werden können. Heterogen strukturierte Daten sind z.B. Beschreibungen, denen auch Strukturelemente fehlen können. Sie werden am besten mit baumartigen Strukturen dargestellt, was direkt auf die moderne Datenbeschreibungssprache XML (eXtensible Markup Language) und die Zugriffssprache XQuery [24] deutet. Vorteil ist die flexible Struktur, die gegenüber der Organisation in Tabellen sehr leicht Änderungen zulässt. Nachteil ist der große Overhead, da die Struktur

beschreibenden Elemente direkt in die Daten eingebettet sind. Für die Darstellung von Metadaten ist diese Art also wie geschaffen.

Für den Entwurf modularer Systeme wäre es sehr hinderlich, für die verschiedenen Datensorten jeweils unterschiedliche Zugriffsmechanismen zu verwenden. Für die Darstellung von Ergebnissen auf einer Web-Seite sind meist beiderlei Datentypen vonnöten. Eine unterschiedliche Behandlung verkompliziert das System drastisch. OGSA-DAI legte für einen integrierten Zugriff den ersten exzellenten Vorschlag vor. Er definiert allerdings nur die *Möglichkeit* des Zugriffs auf unterschiedliche Datenbanktypen. Wir wollen einen Schritt weitergehen und – ausgehend von der Datenakquisition – die grundsätzliche Datenorganisation durch Basisfor-

mate festlegen. Grundlegend ist das XML Schema [25], dessen Basisdatentypen für diese Aufgaben erweitert werden. Für spezifische Anwendungen werden weitere Datentypen benötigt, die aus den Basisdatentypen abgeleitet werden. Dies wird für die Datenakquisition in einer zweiten Stufe weiter betrieben, so dass für ein spezielles Experiment wie KATRIN in einer dritten, applikationsspezifischen Schicht auf beiden vorangegangenen Stufen aufgebaut werden kann. In Abb. 4 ist der Basisdatentyp für die Administrationsdaten von KATRIN zu sehen. Ziel der Erforschung einheitlicher Datentypen ist die Erzeugung von Standardschnittstellen bis weit in die Anwendungen hinein, um ein dynamisches Plug-and-Play von Diensten zu erreichen.

3. Einbinden von Legacy-Systemen

Obwohl die Integration von Diensten das Ziel von Grid-Computing ist, kann bei den wenigsten Entwürfen neuer Systeme davon ausgegangen werden, dass alle Systemkomponenten tatsächlich einer einheitlichen Schnittstellenbeschreibung gehorchen. Die Zahl der Aufgaben und Hersteller ist einfach zu groß und die Komplexität mancher Systeme zu hoch, um sie an die neuen Standards anzupassen. Mit den neuen Web- und Grid-Technologien können solche Alt- oder Legacy-Systeme so verpackt werden, dass sie sich nach außen wie eine Standard-Komponente verhalten. Üblicherweise werden auch Legacy-Systeme laufend verbessert und verändert, was ohne intelligente Um-

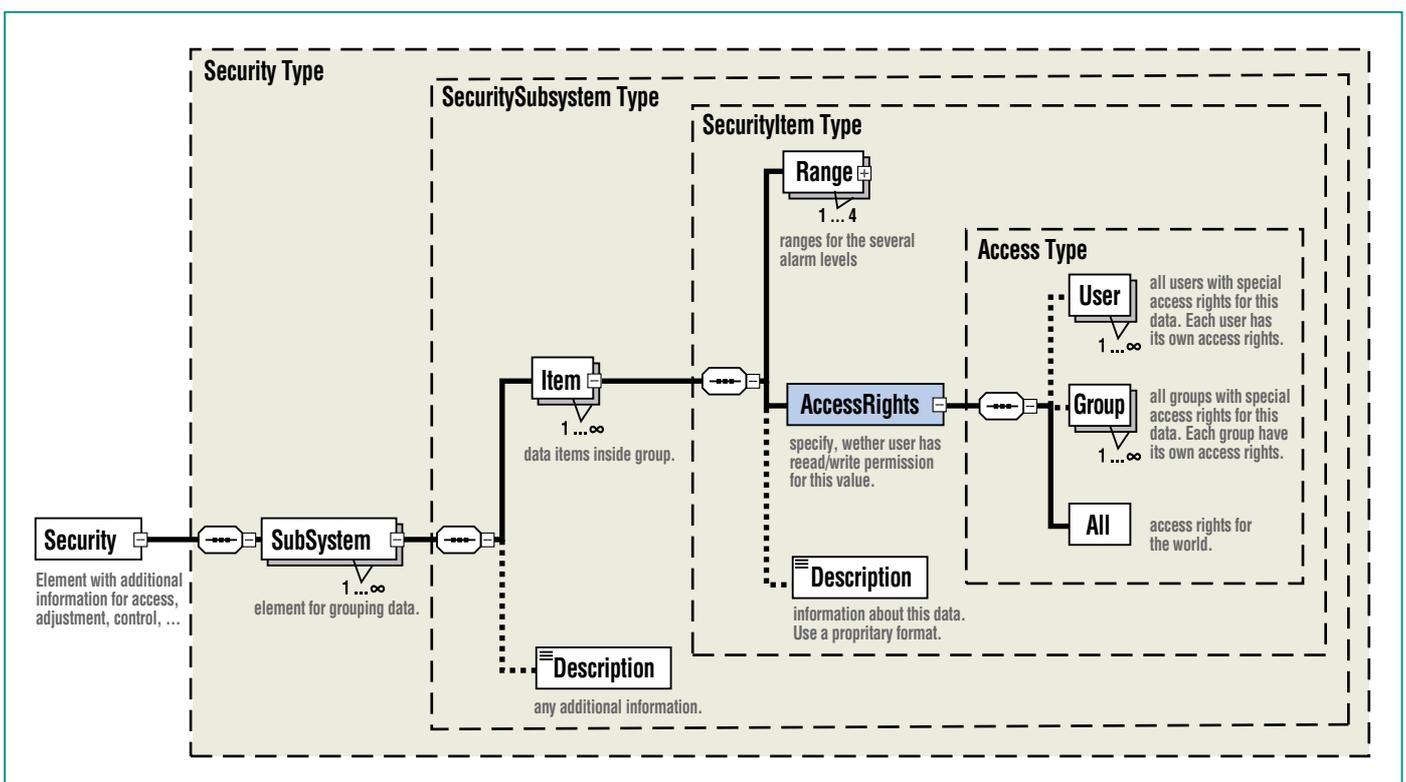


Abb. 4: XML-Schema für die Verwaltungsdaten von KATRIN.

setzungstechniken fehleranfällige Änderungen der Verpackung nach sich ziehen kann. Dieses Problem wird umgangen, wenn eine grundsätzliche Transformation des einen abstrakten Datenformats in das andere beschrieben wird. Für das Projekt KATRIN wurde diese Methode bereits umgesetzt und implementiert. Sie lässt sich verallgemeinern immer dann, wenn eine XML-Beschreibung des Legacy-Datenformats vorliegt. Bei KATRIN ist das Kontrollsystem LabView [26] die Legacy-Komponente, welche nach insgesamt vier Transformationen dem übrigen System dynamisch angepasst wird [27].

4. Einbinden von Arbeitsplatzrechnern in Virtuelle Organisationen (VO)

Es gibt zur Zeit zwei divergierende Richtungen des Grid-Computings. Die eine, hier vorwiegend dargestellte Herangehensweise, konzentriert sich auf die Standardisierung von Grid-Komponenten und deren Zusammenschluss zu einem großen Grid. Dies muss mehrere Aufgaben und Benutzer gleichzeitig bedienen können, die sich nicht gegenseitig stören dürfen. Dafür wurden so genannte Virtuelle Organisationen (VO) [2] geschaffen, die eine logische Auf-

teilung der gemeinsamen physikalischen Ressourcen vornehmen. Die andere Herangehensweise wird durch Projekte wie seti@home [28] oder zetaGrid [29] repräsentiert, die – vorwiegend als Bildschirmschoner – versuchen, möglichst viele brachliegende Arbeitsplatz-PCs zu nutzen. Wir wollen eine Synthese beider Richtungen vorantreiben, um die begrenzten Grid-Ressourcen bei Einhaltung von Standards und Sicherheitsvorkehrungen zu vergrößern. Dafür muss das Konzept der VO erweitert werden, so dass neue Ressourcen dynamisch eingebunden werden können. Damit ist möglich, für bestimmte Zeiten ungeheure Rechen- und Speicherkapazitäten preisgünstig zu nutzen. Dies wird von uns (in Zusammenarbeit mit INT und IAI) exemplarisch für die Vorhersage dreidimensionaler Proteinstrukturen angegangen [9].

Ausblick

Grid-Computing stellt eine gewaltige Herausforderung dar. Für die reibungslose Integration unterschiedlichster Daten und Dienste ist die Durchformalisierung (bzw. „Ontologisierung“) nahezu aller Wissensbereiche nötig, die stark untereinander vernetzt sind. Dies wird zunächst nur für spezielle,

von einander entkoppelte Anwendungsbereiche funktionieren. Die Standardisierung schreitet zügig voran, kann aber nicht verhindern, dass sich manche Standards nicht oder nur zögerlich durchsetzen und immer wieder korrigiert werden. Wir müssen uns daher immer wieder mit veränderten oder halbfertigen Softwarewerkzeugen beschäftigen. Diese zählen zur Infrastruktur, welche immer komplexer wird und immer mehr Zeit und Personal zu ihrer Verwaltung erforderlich macht.

Die Chancen zur Weiterentwicklung des Grid-Computing sind allerdings groß. Die sich entwickelnden Standards werden im Internet offen diskutiert, jeder kann Vorschläge unterbreiten und mit kompetenten Experten diskutieren. Die ungeahnten Möglichkeiten durch vereinheitlichten Zugriff auf Daten und Dienste verursachen eine Aufbruchstimmung, die sehr produktiv ist. Wir nützen sie für wichtige Experimente, die im Forschungszentrum durchgeführt werden.

Literatur

- | | | |
|--|--|--|
| <p>[1] GGF: http://www.ggf.org</p> <p>[2] I. Foster, C. Kesselman, J. Nick, S. Tuecke, <i>Global Grid Forum, June 22, 2002</i>, Available: http://www.globus.org/research/papers/ogsa.pdf</p> <p>[3] M. Antonioletti, et al., Available: http://www.ogsadai.org.uk/</p> | <p>[4] <i>The Globus Alliance</i>, Available: http://www-unix.globus.org/toolkit/</p> <p>[5] A. Osipowicz et al., <i>Letter of Intent</i>, <i>Los Alamos e-print archive</i> http://xxx.lanl.gov/abs/hep-ex/0109033</p> | <p>[6] TOSKA: http://hikwww4.fzk.de/itp/toska/d_index.html</p> <p>[7] AUGER North: http://www.hpe.fzk.de/projekt/auger/auger-north/e_index.html</p> <p>[8] USCT: http://www.ipe.fzk.de/projekt/med/usct/d_index.html</p> |
|--|--|--|

-
-
- [9] W. Eppler, V. Hartmann, *2nd Workshop on Information Technology and its Disciplines, WITID 2004, Kish, Iran*
- [10] J. Silva, R. Stotzka, N. Ruitter, P. Uetz, *In Bildverarbeitung für die Medizin 2004, Informatik Aktuell, 2004*
- [11] M. Kautt, J. Reichert, W. Hoffmann, M. Strasser, M. Rapp, A. Voigt, K. Länge, K.-F. Weibezahn, E. Gottwald, *Nachrichten, Jahrgang 34 · 2-3/2002*
- [12] K. Czajkowsky et al., *Version 1.0, Available: <http://www.globus.org/wsrp/>, 12 Feb 2004*
- [13] *OPC Foundation, June 2002, Available: <http://www.opcfoundation.org>*
- [14] B. Meyers, G. Chastek, *Carnegie Mellon Univ, Software Engineering Institute, Technical Report CMU/SEI-93-TR-10 October 1993*
- [15] F. Bustamante, G. Eisenhauer, K. Schwan, P. Widener, *Proc. ACM/IEEE conference on Supercomputing 2000*
- [16] T. Bray, J. Paoli, C. Sperberg-McQueen, *February 1998. Available: <http://www.w3.org/TR/REC-xml>*
- [17] J. Clark, *November 1999. Available: <http://www.w3.org/TR/xslt>*
- [18] *Struts: <http://jakarta.apache.org/struts>*
- [19] J. Turner, U. N. Lissé, *2003, <http://chiba.sourceforge.net/>*
- [20] W. Allcock, J. Bester, J. Bresnahan, A. Chervenak, L. Liming, S. Tuecke, *<http://www-fp.mcs.anl.gov/dsl/GridFTP-Protocol-RFC-Draft.pdf>, 2001*
- [21] *W3C, September 2002. Available: <http://www.w3.org/TR/soap12-af/>*
- [22] W. Eppler, A. Beglarian, S. Chilingarian, S. Kelly, V. Hartmann, H. Gemmeke, *IEEE TNS, Vol. 51, No. 3, June 2004*
- [23] S Banerjee, *The Twelfth International World Wide Web Conference, 2003, Budapest, HUNGARY, Available: www2003.org*
- [24] *XQuery: <http://www.w3.org/XML/Query>*
- [25] D. Fallside, *May 2001. Available: <http://www.w3.org/TR/xmlschema-0/>*
- [26] *LabView: <http://www.ni.com/labview/>*
- [27] W. Eppler, V. Hartmann, *Multilateral Workshop on Intelligent Information Technology in Control and Data Processing, IITinCDP 2004, Tehran, Iran*
- [28] *seti@home: <http://setiathome.ssl.berkeley.edu/>*
- [29] *zetaGrid: <http://www.zetagrid.net/>*

Das CrossGrid-Projekt

M. Kunze, IWR

Grid-Anwendungen

Die moderne Forschung und Entwicklung bedient sich der Expertise einer großen Zahl von Wissenschaftlern, die mit unterschiedlichem Know-how zusammenarbeiten müssen. Diese interdisziplinäre Zusammenarbeit stützt sich dabei in der Regel auf den Austausch von Informationen in weltweit verteilten Organisationen. Grid-Computing ist ein Mittel der Wahl um diesen Prozess zu unterstützen: Grids bilden eine Infrastrukturbasis zur generischen Unterstützung verteilter Anwendungen z.B. in der Hochenergiephysik, der Medizin und der Umweltforschung [1].

Das von der EU geförderte CrossGrid Konsortium¹⁾ umfasst 21 Partner aus 11 Ländern und beschäftigt sich mit interaktiven Grid-basierten Anwendungen aus den genannten Bereichen [2]. Die Anwendungen sind dadurch gekennzeichnet, dass eine Person in einer Verarbeitungsschleife permanent mit dem verteilten System wechselwirkt: Charakteristisch dabei ist, dass eine schnelle Antwort vom Computersystem als Reaktion auf eine datenintensive Eingabe garantiert sein muss. Beispiele für diesen Typ von Anwendungen sind:

- Virtuelle chirurgische Modelle zur Operationsplanung und Simulation,
- Entscheidungsunterstützungssysteme für Flutkatastrophen,
- Verteilte Datenanalyse in der Hochenergiephysik,

- Kombination von Luftverschmutzungsmodellen mit Wettervorhersagen.

Exemplarisch soll im Folgenden ein Beispiel aus dem medizinischen Bereich angeführt werden: Gefäßerkrankungen stellen ein großes Gesundheitsproblem dar und die Behandlung erfordert oftmals einen chirurgischen Eingriff. Dieser beinhaltet das Legen eines Bypass um die betroffenen Arterien zur Verbesserung des Blutflusses. Die Operationsplanung stützt sich dabei auf 3D-Aufnahmen, die üblicherweise mit Computertomographen gewonnen werden. Die hier zum Einsatz kommenden Informationssysteme profitieren erheblich durch Kombination mit Werkzeugen zur Simulation und Visualisierung. Diese Herangehensweise erfordert aber im Augenblick der Behandlung schnellen Zugriff auf große Rechnerleistung, die aus

Kostengründen in der Regel vor Ort nicht installiert ist. Grid-Computing bietet hier einen eleganten Ansatz zur Problemlösung, indem verteilte Rechenressourcen dynamisch zugeschaltet werden können. Im CrossGrid-Projekt wurde ein solches prototypisches Grid-System zur Unterstützung virtueller Modelle in der Operationsplanung und Simulation entwickelt. Das System unterstützt den Chirurgen beim Entscheidungsprozess über verschiedene Operationsmethoden, indem der Blutfluss simuliert und vorhergesagt wird auf der Basis von vermessener Arteriengeometrie sowie von gespeichertem Expertenwissen über die verschiedenen Behandlungsformen. Interaktive 3D-Visualisierung wird eingesetzt, um den Chirurgen in klarer und intuitiver Weise zu leiten (siehe Abb. 1). Verschiedene Operationspläne können so vor der eigentlichen Operation im

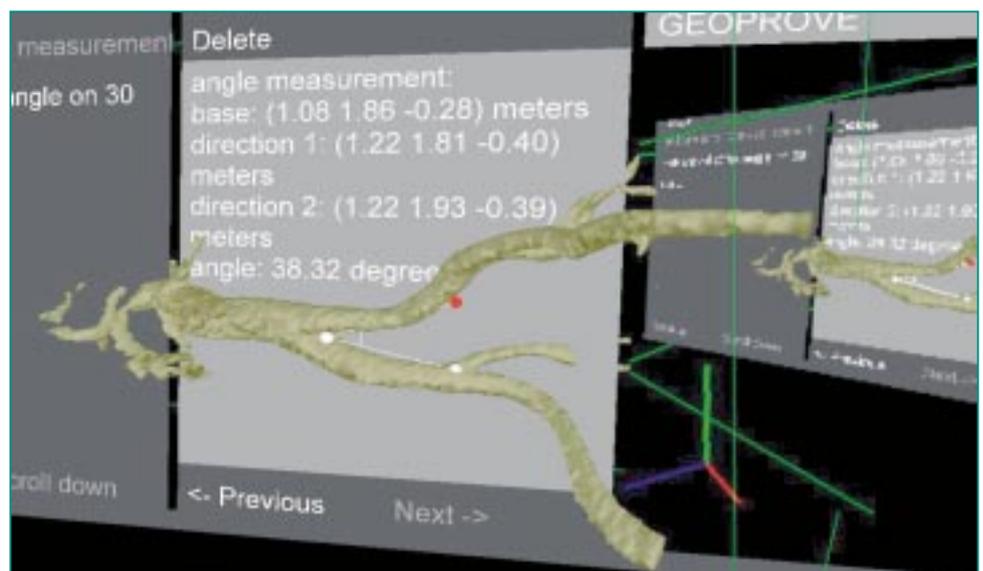


Abb. 1: Aus computertomografischen Daten gewonnenes 3D-Modell einer Arterie. In einer Virtual Reality-Umgebung kann der Chirurg die Operation planen (Messung der Länge eines Bypass) und durch Simulation den Blutfluss vorhersagen.

¹⁾ Vertrag IST-2001-32243, Laufzeit 2002-2005, Fördersumme ca. 2.3 Millionen Euro.

Hinblick auf ihre Effizienz und Machbarkeit interaktiv überprüft werden.

Der transparente und sichere Zugriff auf verteilte Ressourcen wird dabei durch einen Resourcebroker organisiert. Die CrossGrid-Infrastruktur erlaubt in diesem Zusammenhang die nahtlose Einbindung von medizinischen Scannern und Daten, der Visualisierungsumgebung sowie die Vermittlung und die Bereitstellung der zur Simulation des Blutflusses nötigen Ressourcen in örtlich verschiedenen administrativen Rechenzentrumsdomänen.

Grid-Entwicklungs- umgebung

Nicht nur für die Anwender, sondern auch für Programmentwickler ist einer der kritischen Aspekte des Grid die inhärente Komplexität. Die Entwicklung, die Fehlersuche und die Optimierung von verteilten Anwendungen auf dem Grid ist schwierig und zeitaufwändig aufgrund der sich dynamisch ändernden heterogenen Infrastruktur aus Rechnern, Speichern und Netzen. Um die genannten Anwendungsfelder effizient auf einem Grid implementieren zu können, werden daher nicht nur neue Werkzeuge zur syntaktischen Überprüfung paralleler Programme, sondern auch Methoden zur Leistungsvorhersage und Leistungsbewertung benötigt. Ferner werden Hilfsmittel gebraucht zur Ermittlung der Anwendungsleistung sowie zum interaktiven Management der zur

Durchführung der Aufgaben benötigten Ressourcen im Grid.

Es ist ein weiteres Ziel des CrossGrid-Projekts, solche Monitoring-Werkzeuge zu entwickeln und in die Grid-Umgebung zu integrieren, um damit die Erstellung und die Optimierung von verteilten, parallel auf mehreren Rechnern ausgeführten Programmen zu erleichtern. Um solche Parallelapplikationen bei reproduzierbaren Ergebnissen portabel zu gestalten, wurde ein spezielles Verifikationstool für MPI-Programme erstellt (MARMOT). Die automatische Überprüfung, ob Benutzeranwendungen mit dem MPI²⁾-Standard übereinstimmen, verringert die Fehleranfälligkeit schon im Vorfeld der Implementierung. Es werden dabei Programmiermodelle unterstützt, die sowohl rechenzeitintensiv sind als auch datenintensiv. Um das Grid auch für im Hinblick auf Ressourcenanforderungen anspruchsvolle interaktive Anwendungen tauglich zu machen, benötigen die Entwickler darüber hinaus Werkzeuge, um die Leistung ihrer Programme überwachen und analysieren zu können. In diesem Bereich wurden entsprechende Hilfsmittel entwickelt zum Aufspüren von Engpässen, sowie entsprechende Benchmark-Programme entworfen, die sensitiv im Hinblick auf Datentransfer, Synchronisation und Latenzzeit sind. Zusammen mit den Monitoring-Tools kann ein verteiltes System so mit nahezu beliebiger Granularität (Ort und Zeit) bis auf den Prozess-Level untersucht und optimiert werden.

Grid-Portale

Es ist eine der Grid-Visionen, dass Benutzer ihre Anwendungen auf dem Grid in einer einfachen und transparenten Weise ausführen können, ohne die genauen Details der Grid-Infrastruktur zu kennen. Internetportale mit dynamisch zusammengestellten Informationen und Verweisen helfen, den Zugang zur Grid-Welt zu vereinfachen: Der Zugriff über Internetportale sowie die Plattform- und Ortsunabhängigkeit spielen eine entscheidende Rolle für die Akzeptanz der neuen Technologie seitens der Endanwender. Entsprechend einer Umfrage über die Anforderungen von Gridnutzern im Rahmen des ENACTS-Projekts [3] würde eine größere Zahl von Anwendern Grid-Services in Anspruch nehmen, wenn der Zugriff einfacher wäre. Portaltechnologien, die etwa im Grid Portal Development Toolkit (GPDK) propagiert werden, zielen in diese Richtung [4].

CrossGrid entwickelt aus diesem Grund benutzerfreundliche Portale und mobile Arbeitsumgebungen. Eine der Portaltechniken, die in CrossGrid implementiert wird, ist „Roaming“, d.h. der mobile persönliche Zugriff auf Ressourcen. Ein Grid-Service (Portal-Server) führt Buch über alle Interaktionen des Benutzers, der sich auf seinem Endgerät (Portal-Client) zunächst über eine Eingabemaske authentifizieren muss. Die Benutzerschnittstelle ist dabei gängigen Desktop-Oberflächen nachempfunden.

²⁾ MPI = Message Passing Interface

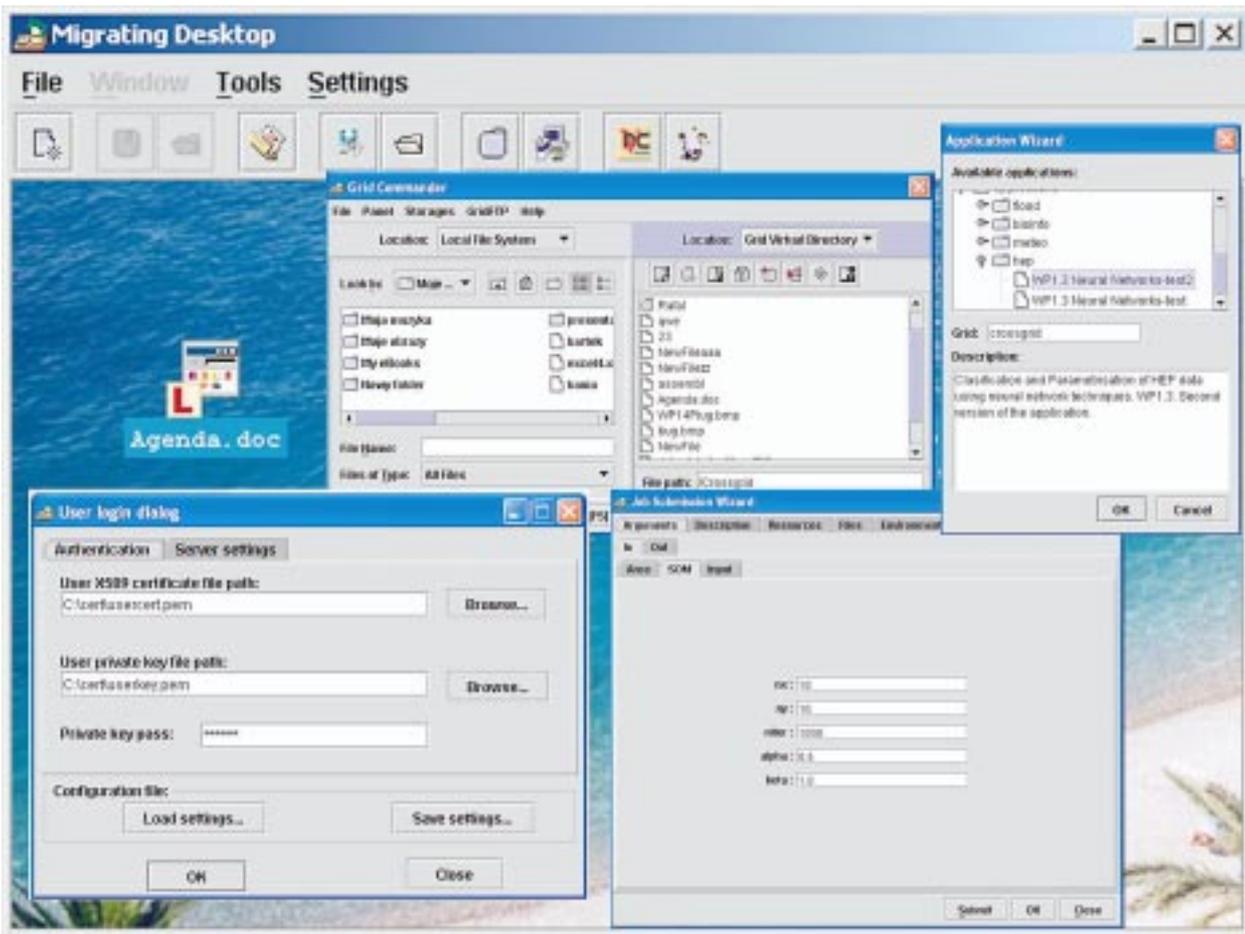


Abb. 2: Mobile Benutzerschnittstelle für CrossGrid-Applikationen. Durch enge Anlehnung an das Erscheinungsbild gängiger Desktop-Oberflächen finden sich auch Neueinsteiger schnell zurecht. Zu sehen sind neben Verknüpfungen zu Dateien im Grid (Agenda.doc, oben links) auch Fenster zur Anmeldung (User Login Dialog, unten links) und Job-Abgabe (Job Submission Wizard, unten rechts) sowie zur Navigation (Grid Commander, oben Mitte) und Erstellung von Benutzerschnittstellen (Application Wizard, oben rechts).

den (siehe Abb. 2), um die Einarbeitungszeit zu verkürzen. Es ist nicht nur möglich, Grid-Ressourcen interaktiv zu verwalten (Grid Commander), sondern es können Dateien im Grid auch so verarbeitet werden, als ob sie lokal gespeichert wären (Virtual Directory Browser). Für Grid-Applikationen können durch einen sog. „Application Wizard“ auf einfache Art und Weise grafische Schnittstellen zur Eingabe und

Ausgabe von Informationen gebaut werden, die Applikation kann durch einen „Job Submission Wizard“ gestartet werden. Der Zustand einer Sitzung ist persistent, d.h. ein Benutzer kann sich jederzeit abmelden und seine Arbeit an einem anderen Ort zu einer späteren Zeit auf einer anderen Plattform wieder aufnehmen, wobei die Oberfläche identisch restauriert wird.

Internationales Testbett

Organisatorische Probleme sowie Fragen der Leistung und Sicherheit im Netzwerk können nur in einem realen Umfeld gelöst werden. Im CrossGrid Projekt wird daher eine internationale Testumgebung aufgebaut und gepflegt, um die Integration und Optimierung der verteilten Anwendungen und Werkzeuge zu ermöglichen. Die Zielsetzungen

der Testbettorganisation im Projekt sind:

- Bereitstellung verteilter Ressourcen,
- Integration nationaler Grids,
- Integration der Softwarepakete,
- Koordination der Software-Release Versionen,
- Sicherstellung der Verträglichkeit mit Grid-Standards.

Das Testbett wird an 10 unterschiedlichen europäischen Standorten verteilt betrieben. Ein wesentlicher Beitrag wird in diesem Bereich vom Forschungszentrum Karlsruhe geleistet: Neben der Bereitstellung von Ressourcen im Grid (Rechnerelemente, Speicherelemente) wird hier auch der

zentrale Server zur Verwaltung der Software-Releases betrieben. In regelmäßigen Abständen wird mittels einer speziell entwickelten Infrastruktur aus den Eingaben der Partner in automatisierter Weise ein Softwarepaket gebaut und in Testprozeduren auf Konsistenz und Fehlerfreiheit überprüft. Auf diese Weise ist sichergestellt, dass ein Software-Release eine gewisse Robustheit und Reife besitzt, bevor er den Server verlässt und in der Kollaboration sowie auf dem Testbett verteilt wird.

Zusammenfassung

Das CrossGrid-Projekt beschäftigt sich mit interaktiven Anwendungen im Grid. Interaktivität im Grid ist ein durchaus neues,

schwieriges und spannendes Thema, das hier erstmals erfolgreich aufgegriffen wird. Neben prototypischen Anwendungen aus dem Bereich der Hochenergiephysik, der Medizin und der Umweltforschung werden auch Spezialwerkzeuge bereitgestellt zur Programmentwicklung, sowie Portaltechniken, die den Umgang mit dem Grid erleichtern und damit die Akzeptanz der neuen Technik verbessern. Das Forschungszentrum Karlsruhe erbringt wesentliche Beiträge im Bereich der Grid-Infrastruktur und der Grid-Services.

Literatur

[1] M. Kunze, H.Marten, *Grid-Computing und e-Science.*, Nachrichten 4/2002

[2] *The CrossGrid Collaboration*, IST-2001-32243, <http://www.crossgrid.org/>

[3] <http://www.enacts.org/>

[4] <http://doesciencegrid.org/projects/GPDK/>

Das Projekt CampusGrid

F. Schmitz, IWR

Einleitung

Die tragende Idee des Grid Computings ist die Zurverfügungstellung von heterogenen Ressourcen wie CPU-Leistung, Platten- und Archivspeicher aber auch die Unterstützung der Benutzer (GGUS: Global Grid User Support) und die Einbindung von Experimental-Großgeräten. Um alle diese Herausforderungen angehen zu können, hat das Institut für Wissenschaftliches Rechnen (IWR) das Projekt CampusGrid mit Beginn des Jahres 2004 gestartet. Anwendungen aus der Physik, der Bioinformatik, Nanotechnologie und Meteorologie werden als Testanwendungen ins CampusGrid integriert und sollen dann die aus Vektorcomputern, SMP-Clustern und preisgünstigeren Standardkomponenten (Linux-Cluster, Windows-Cluster) bestehende CPU-Infrastruktur mit einer Hochgeschwindigkeitslösung im Speicherbereich (globales Filesystem) nutzen.

Es werden die Grundzüge des Projektes, der momentane Status sowie ein Ausblick in die Zukunft gegeben. Diese Aktivitäten stehen im Zusammenhang mit den anderen Grid und Hochleistungsrechner (HPC) Aktivitäten im IWR.

Motivation

Hochleistungsrechnen im Forschungszentrum Karlsruhe startete 1983 mit den gemeinsamen Aktivitäten der Universität Karlsruhe und des Forschungszentrums. Mit Standort an der Universität Karlsruhe wurde eine Cyber 205 beschafft und gemeinsam genutzt. Erschwerend für die Ak-

zeptanz war die völlig eigenständige Nutzung der Cyber an der Universität und der IBM Rechenlandschaft im Forschungszentrum, d.h. Dateiverwaltung, Benutzerverwaltung, Jobsprache und vieles mehr waren voneinander unabhängig und in höchstem Maße verschieden. Der Nutzer musste sich jeweils von neuem um seine maschinenspezifische Umgebung kümmern. 1997 änderte sich mit der Beschaffung eines Vektorrechners VP50 für das Forschungszentrum dieser Zustand. Das Rechenzentrum ermöglichte den Benutzern von dem zentralen Rechner des Forschungszentrums aus (IBM 3090 unter MVS) die einfache Nutzung der VP50 und später der VP400EX durch einen so genannten Präprozessor.

Jedoch wurde diese Lösung von den Benutzern nur bedingt angenommen, so dass die Nachfolgeinstallationen (Cray J916, VPP300, VPP5000) völlig unabhängig von den IBM (MVS, AIX) und später auch Linux Rechnern betrieben wurden. Die Benutzer mussten ihre Daten mehrfach halten, wenn sie auf mehr als einer Rechnerplattform arbeiten wollten.

Ziel des Projekts CampusGrid

Die Rechner im Projekt CampusGrid sollen für den Anwender im Forschungszentrum transparenter miteinander verknüpft werden um eine globale Sicht der Dinge zu ermöglichen. Daraus ergibt sich eine Zieldefinition für das Projekt, wobei einige Zielsetzungen in der aktuellen Produktions-

umgebung des Instituts für Wissenschaftliches Rechnen (IWR) bereits realisiert wurden, jedoch im CampusGrid Projekt neu zu überdenken sind:

- Globale Benutzerverwaltung und Verwendung eines Passwortes für die Benutzerauthentifizierung (Arbeitsplatzrechner und Computerserver) bis hin zum Single-Sign-On.
- Aufbau einer Virtuellen Organisation (VO) zur Akzeptanz auch von Zertifikaten von etwa GridKa.
- Mindestens ein schnelles gemeinsames Filesystem zur globalen Datensicht der Benutzer auf den Computerservern.
- Zugriff der Arbeitsplatzrechner (meist unter dem Betriebssystem Windows oder Linux) auf das schnelle Filesystem der Computerserver.
- Globale Jobverwaltung unter Hinzuziehung eines Brokers zur effizienten und bedarfsgerechten Nutzung der Ressourcen. Damit ist gemeint, dass je nach angefordertem Hauptspeicher- und Plattenspeicherbedarf sowie Anwendungsprofil (Vektor-, SMP-, Windows- oder Linuxrechner) und Antwortzeit bzw. Preisvorstellung der geeignete Computerserver automatisch gewählt wird. Auch ist damit die Definition eines Work-Flows unter Einbeziehung unterschiedlicher Rechner möglich.
- Integration weiterer Ressourcen des Forschungszentrums und anderer Zentren der HGF,

hier bietet sich etwa eine Schnittstelle zum Höchstleistungsrechner JUMP im Forschungszentrum Jülich an. Die Nutzung von UNICORE (www.fz-juelich.de/unicoreplus) als integrierendes Hilfsmittel zwischen unterschiedlichen Softwareansätzen ist denkbar.

- Berücksichtigung bestehender und in Entwicklung befindlicher Grid-Standards und „Best Practices“.
- Ein API (Application Programming Interface) zur Nutzung der CampusGrid-Ressourcen ist zu definieren.
- Nutzung des CampusGrids durch Experimente (Echtzeitfähigkeit).
- Einrichtung von Problemlösungsumgebungen (PSEs)
- Die Entwicklung netzwerkzentrischer Anwendungen muss verstärkt angegangen werden.

- Portierung von Anwendungen aus Industrie und Forschung ins CampusGrid (Meteorologie, Nanotechnologie, Festkörperphysik, Fluidmechanik, Bioinformatik, ...).

Das Projekt CampusGrid wurde im Rahmen der PoF (Programmorientierte Förderung) im Programm „Wissenschaftliches Rechnen“ im Mai 2004 positiv beurteilt. Dies lässt die Projektpartner hoffen, dass einer Förderung bis 2009 nichts im Wege steht. Bereits für 2004 wurden deshalb die folgenden Arbeiten in Angriff genommen:

- Auswahl zweier globalen Filesystems (SAN-FS von IBM, StorNextFS von Adic) zum Test unter den relevanten Serverplattformen, wobei zusätzlich das experimentelle Filesystem Oracle 10G hinzugenommen wird. Diese Filesysteme werden in einem Storage Area Network (SAN) Testumfeld auf ihre Tauglichkeit, speziell die

Skalierbarkeit und Ausfallsicherheit, untersucht.

- Aufbau unterschiedlicher Hardware zur Portierung von Anwendungen und zum Design der Middleware (Schnittstelle zwischen Benutzer/Anwendung und Betriebssystem):
 - SX-5 (siehe Abb. 1) und SX-6i Vektorcomputer von NEC
 - p630 AIX-Rechner von IBM
 - Infiniband-Cluster
 - Blade-Center mit PowerPC und Intel Prozessoren (Infiniband und SAN)
 - Xeon Testsystem mit Infiniband Netzwerk (siehe Abb. 2)
- Aufbau einer SAN Testumgebung mit Cisco MDS Fabric zum Test von FC (Fibre Channel, Protokoll), iSCSI (Protokoll SCSI über IP) und später Infiniband.



Abb. 1: Aufbau eines Vektorrechners.



Abb. 2: Kupferleitungen ermöglichen schnelle Kommunikation.

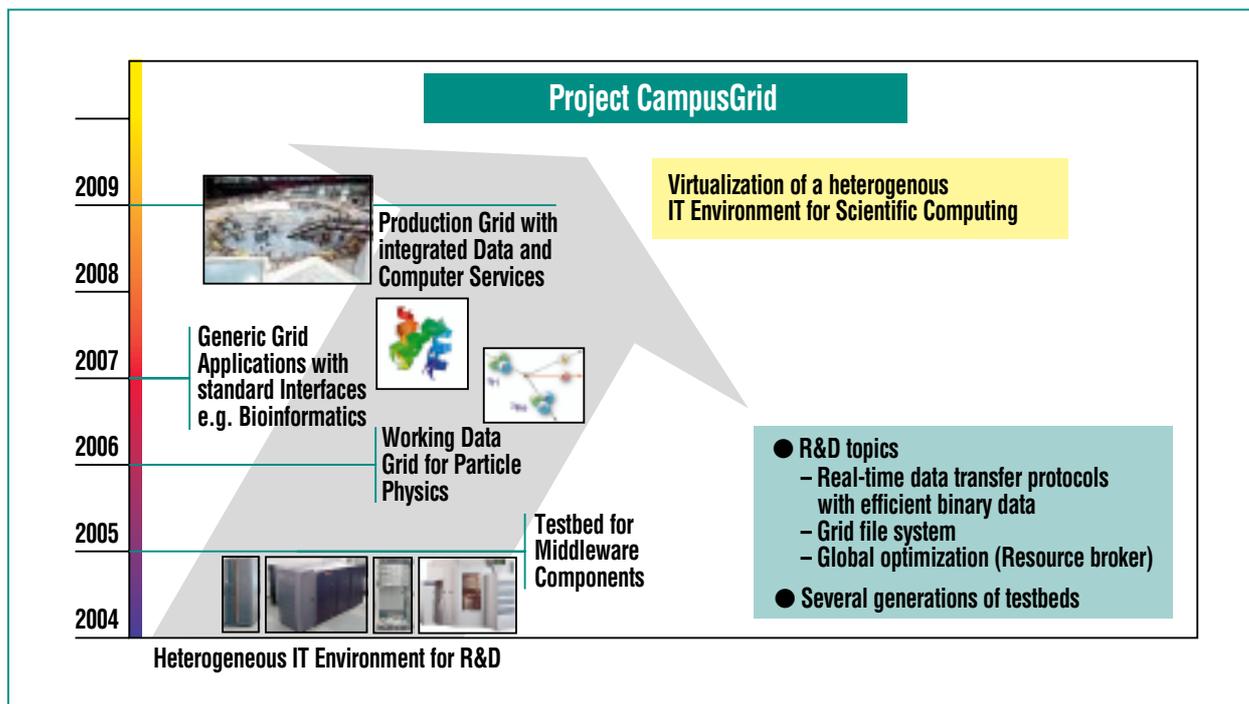


Abb. 4: Wo wird die Reise hingehen?

Schnittstelle (Middleware) ermöglicht werden. Zur besseren Verteilung wird ein Broker zur Verfügung gestellt. Die Entwicklung vom Beginn des Projektes bis 2009 ist in Abb. 4 dargestellt. Ziel für das Jahr 2009 wird die voll-

ständige Integration heterogener Hardwareressourcen (Rechner, Storage) sowie eine einheitliche Sicht auf die Daten und deren globale Nutzung auf dem Gelände des Forschungszentrums sein. Parallele Anwendungen über das

CampusGrid sollten dann selbstverständlich sein und der Broker die optimale (Preis, Durchsatz, ...) Ressourcenanforderung übernehmen.

Literatur

M. Kunze, H. Marten,
Nachrichten Jahrgang 34
4/2002, S. 280-283

F. Schmitz,
Nachrichten Jahrgang 34
4/2002, S. 269-274

Von der Vision zur Realität: Grid-Computing für kommerzielle Anwendungen

K.-P. Mickel, IWR

Einleitung

Das Bundesministerium für Bildung und Forschung hat unlängst seine Absicht erklärt, die „Deutsche Grid- und e-Science-Initiative“ von 2005 bis 2009 mit insgesamt 100 Millionen Euro zu unterstützen, die Europäische Union fördert im gegenwärtigen 6. Rahmenprogramm die Themen Grid-Computing, e-Science und Vernetzung mit 150 Millionen Euro (im kommenden 7. Rahmenprogramm voraussichtlich mit bis zu 300 Millionen Euro), Großbritannien finanziert sein nationales e-Science-Programm mit 40 Millionen Pfund pro Jahr und der US-amerikanischen National Science Foundation ist dasselbe Thema eine Milliarde US-Dollar wert. Man kann daraus unschwer ableiten, dass Grid-Computing wie auch der oft synonym verwendete Begriff „e-Science“ (digitally enhanced Science) sich anschicken, in einer Art technologischer Revolution die Informationstechnik (IT) der Welt zu erobern und vollkommen neu zu definieren. Den Begriff „Grid“ haben seine Schöpfer Foster und Kesselmann [1] in Analogie zum „Power Grid“, zum Stromversorgungsnetz, gewählt. Sie wollten damit plakativ darlegen, dass es mit „Grid-Computing“ in gar nicht ferner Zukunft möglich werde, Computerressourcen unterschiedlichster Art quasi aus der Steckdose („from a plug in the wall“) zu beziehen, ohne sich noch um die Herkunft und die Verteilung dieser Ressourcen Gedanken machen zu müssen.

Bei der Nutzung des WWW, des World Wide Web, das als Vorgän-

ger des Grid gelten kann, geht es im Wesentlichen darum, einem individuellen Anwender den Zugriff auf – in aller Regel einfach strukturierte – Informationen zu ermöglichen, die sich irgendwo im weltweiten Netz befinden (siehe Abb. 1). In einem Grid werden dagegen viele Ressourcen ganz unterschiedlicher Art logisch so vereinigt, dass die Anwender diese Ressourcen anschließend „aus der Steckdose“ beziehen können (siehe Abb. 2). Grid-Computing wurde, wie auch das WWW, zunächst von Wissenschaftlern entwickelt, die sich da-

mit ein Werkzeug schufen, um komplizierte wissenschaftliche Fragestellungen besser lösen zu können. Am Beispiel des WWW konnte man jedoch in den vergangenen etwa sechs Jahren miterleben, wie schnell Wirtschaft und Industrie dieses neue Werkzeug weltweit auf ihre Bedürfnisse angepasst haben, was schließlich dazu führte, dass die heutige kommerzielle Welt ohne das WWW nicht mehr vorstellbar ist. Wir erleben derzeit, wie die Industrie sich auch des neuen Werkzeugs „Grid“ bemächtigt, was für die kommerzielle Welt al-

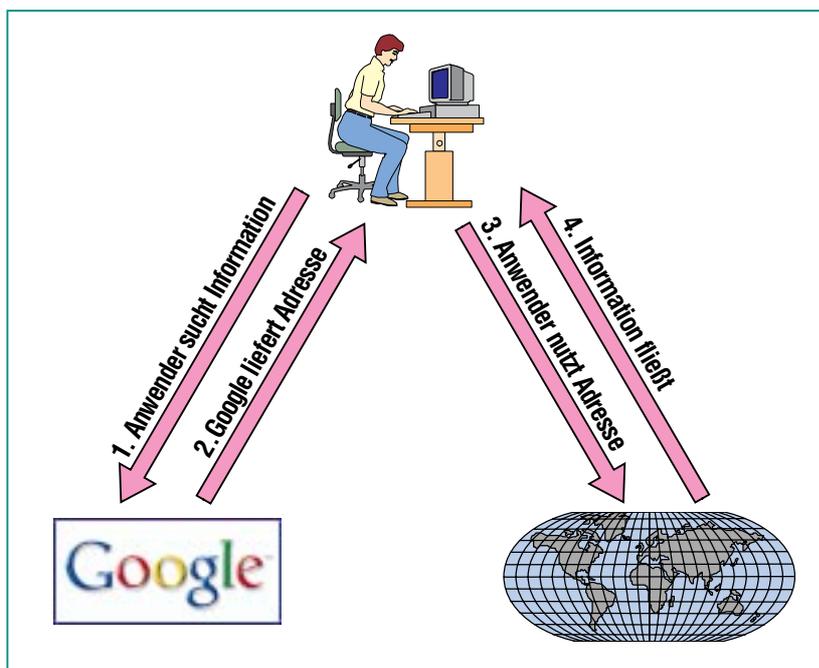


Abb. 1: Vereinfachte Darstellung der Informationsgewinnung im World Wide Web (WWW):

- 1. Ein Anwender gibt Suchbegriffe in eine Internetsuchmaschine ein.**
- 2. Die Internetsuchmaschine liefert dem Anwender die Adressen von Computern im Internet, in denen die angegebenen Suchbegriffe gespeichert sind.**
- 3. Der Anwender wählt eine dieser Internetadressen, eine Verbindung zu dem betreffenden Computer wird hergestellt.**
- 4. Die gesuchte Information fließt zum Anwender und kann von diesem genutzt werden.**

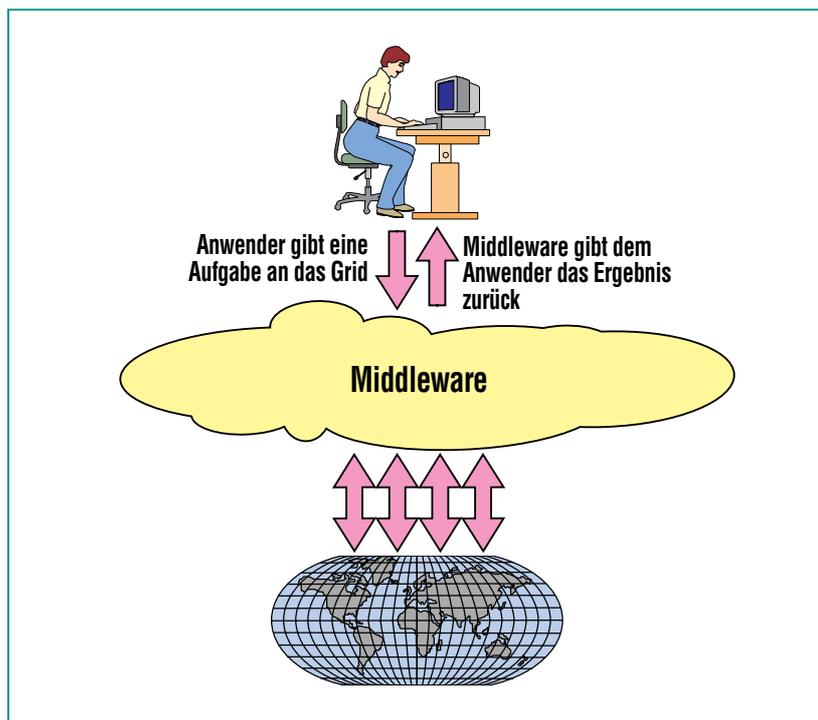


Abb. 2: Vereinfachte Darstellung der Arbeitsweise in einem World Wide Grid (WWG):

- 1. Ein Anwender übergibt dem WWG eine Aufgabe, für deren Lösung unterschiedliche Ressourcen (Computer, Datensammlungen, Software, Datennetze) benötigt werden.**
- 2. Die vermittelnde Software („Middleware“) koordiniert die im WWG vorhandenen, global verteilten Ressourcen und teilt sie dieser Aufgabe im benötigten Umfang zu, so dass die Aufgabe gelöst werden kann.**
- 3. Nach Abschluss der Bearbeitung sendet die Middleware dem Anwender das Ergebnis zurück.**

ler Voraussicht nach einen noch tiefgreifenderen Umbruch zur Folge haben wird, als wir das beim WWW erlebt haben.

Grid-Computing – von wissenschaftlichen zu kommerziellen Anwendungen

Der Beitrag „Grid-Computing – Basis für institutsübergreifende Virtuelle Organisationen“ im vorliegenden Heft schildert die Ideen und Konzepte des Grid-Compu-

ting sowie Anwendungen aus der Wissenschaft, die zu treibenden Kräften der Entwicklung des Grid-Computing geworden sind; weitere Beiträge beschreiben Grid-Anwendungen aus anderen wissenschaftlichen Disziplinen. Einige dieser Anwendungen sind schon heute weitgehend produktionsreif – dies gilt insbesondere für die Grid-Projekte der Hochenergie- und Elementarteilchenphysik –, andere befinden sich noch im Entwicklungszustand. Gleichzeitig ist festzustellen,

dass sich der Entwicklungsschwerpunkt der Grid-Technologie von den rein wissenschaftlichen Anwendungen hin zu Grids in der öffentlichen Verwaltung und in Industrieunternehmen verlagert, die nur vordergründig eine Verschlinkung des IT-Einsatzes zum Ziel haben. Die wesentlichen wirtschaftlichen Vorteile der Grid-Technologie liegen vielmehr in der damit erreichbaren deutlich höheren Flexibilität der Unternehmen gegenüber kurzfristigen Marktanforderungen und bei wesentlich kürzeren Entwicklungszeiten für neue Produkte, so dass solche Unternehmen dem globalen Konkurrenz- und Kostendruck zukünftig deutlich besser standhalten können.

Mittlere und große Industrieunternehmen haben in der Vergangenheit ihre ständig wachsenden IT-Anforderungen in aller Regel mit der bedarfsgerechten Erweiterung ihrer zentralen Rechenzentren befriedigt. Dort wurden große Rechen- und Speicherkapazitäten konzentriert, dort wurden die umfangreichen und oft existenziell wichtigen Datenbanken der Unternehmen gebündelt. Diese Rechenzentren mussten dabei meist so dimensioniert werden, dass sie auch die größten planbaren Benutzeranforderungen erfüllen konnten. Selbst für wirtschaftlich gesunde Unternehmen ist dieser Weg nahezu unbegrenzten IT-Wachstums zu Zeiten eines starken globalen Konkurrenzdrucks nicht mehr gangbar; für kleine und mittelständische Unternehmen (KMU), die sich erst nach und nach die Vorteile der modernen IT zunutze machen können, war er seit jeher

undenkbar. Für beide Szenarios verspricht das Grid-Computing faszinierende Möglichkeiten: Es wird Unternehmen möglich sein, IT-Ressourcen nur bei Bedarf zu nutzen, ohne sie ständig an einem Ort vorhalten zu müssen, sie werden bei Bedarf auf große und sehr große Datenbanken aus ganz unterschiedlichen Themengebieten zugreifen können, ohne diese Datensammlungen selbst pflegen zu müssen, und sie werden in der Lage sein, im Grid auch hochkomplexe Software-Pakete zu nutzen, ohne dafür eigene Spezialisten finanzieren zu müssen. Aber es gibt noch weitere Beweggründe, zentrale Rechenzentren durch Grid-Modelle zu ersetzen: Global operierende, multinationale Unternehmen sind i. A. nicht zentral organisiert, sondern haben verteilte Hierarchien und eben auch verteilte IT-Ressourcen. Gerade bei weltweit verteilten Unternehmen werden sehr oft virtuelle Teams gebildet, die zwar gemeinsam dieselben Projekte bearbeiten, wegen der Zeitverschiebung jedoch zu sehr unterschiedlichen Tageszeiten. Und es wird kaum noch der Fall sein, dass man es als Mitarbeiter eines großen Unternehmens nur mit Menschen zu tun hat, die am selben Ort arbeiten wie man selbst. Das alles spricht für die Ablösung herkömmlicher zentraler IT-Strukturen durch weiträumig verteilte Grid-Strukturen. Dennoch ist es sicherlich verfrüht, wenn die großen IT-Hersteller im Vorgriff auf zukünftige Grid-Entwicklungen schon heute als unmittelbare Folge aus der Einführung des Grid-Computing vollmundig eine Halbierung der industriellen Entwicklungszeiten, eine Verdoppe-

lung der Produktivität der Mitarbeiter, eine naht- und problemlose Zusammenarbeit aller Angestellten eines weltweit verteilten Konzerns, eine weitgehend problemlose Integration heterogener IT-Umgebungen sowie ganz allgemein eine deutliche Steigerung aller IT-Ressourcen eines Unternehmens in Aussicht stellen [2].

Besondere Herausforderungen bei kommerziellen Grid-Anwendungen

Die Idee des Grid-Computing entstammt unterschiedlichen wissenschaftlichen Fragestellungen. Hier, im Umfeld wissenschaftlicher Disziplinen mit teilweise extrem großem Bedarf an IT-Ressourcen wurden und werden in weltweiten gemeinsamen Anstrengungen die ersten funktionsfähigen Grid-Umgebungen entwickelt, und man arbeitet hier auch intensiv an der Weiterentwicklung der dazu nötigen Standardisierungen und an der Entwicklung geeigneter „Middleware“, also der Systemsoftware des Grid. Das Global Grid Forum (GGF) hat sich seit seiner Gründung im Jahre 2001 zu der bei diesen Bemühungen führenden Kooperation von wissenschaftlichen und kommerziellen Einrichtungen entwickelt [3]. Bei den meisten Entwicklungsarbeiten und Standardisierungsbemühungen im GGF lässt sich beobachten, dass nahezu immer die wissenschaftlichen Anwender die Vorreiter der Grid-Technologie sind. Als Gegenbewegung wurde allerdings jüngst innerhalb des GGF eine „Enterprise Grid Requirements Research Group“ gebil-

det, welche die Erforschung der künftigen Anwendbarkeit von Grid-Technologien im kommerziellen Umfeld zum Ziel hat [4]. Die kommende 12. Tagung des Global Grid Forum (GGF12) wird sich sogar gänzlich dem Thema „Grids Deployed in the Enterprise“ widmen [3] und Experten aus zahlreichen Branchen über bisherige Ansätze zum Aufbau und zur Nutzung von Daten- und Compute-Grids berichten lassen; vorgesehen sind bisher Erfahrungsberichte aus der pharmazeutischen Industrie, aus der Luft- und Raumfahrt, von Fahrzeugherstellern sowie von Banken und Versicherungen. Ein wesentlicher Grund für das bislang noch deutliche Übergewicht wissenschaftlicher Anwendungen ist sicherlich, dass manche der heute in der Grid-Technologie noch ungelösten technischen Herausforderungen bei kommerziellen Anwendungen notwendigerweise einen weit höheren Stellenwert haben müssen als bei wissenschaftlichen Anwendungen, so dass wissenschaftliche Anwendungen sozusagen „unbeschwerter“ entwickelt und in eine erste Produktionsphase übernommen werden können. Zu nennen sind hier insbesondere eine zuverlässig vorhersagbare Service-Qualität, die Stabilität des Produktionsbetriebs, die Vertraulichkeit der übermittelten und verarbeiteten Daten sowie die Sicherheit des Datenzugriffs. Während sich Service-Qualität und Betriebsstabilität durch eine Überdimensionierung der Ressourcen um 10 bis 20 Prozent recht zuverlässig sicherstellen lassen, bedarf es für Vertraulichkeit und Sicherheit komplizierter

und aufwändiger Verschlüsselungs- und Authentifizierungsverfahren (siehe Artikel „Verschlüsselung und Authentifizierung“ in diesem Heft). Manche unternehmensinternen und damit leider nicht zitierfähigen Untersuchungen sprechen deshalb hier von einem bis zu zehnfach höheren Entwicklungsaufwand für kommerzielle Grid-Anwendungen gegenüber rein wissenschaftlich orientierten Grid-Applikationen. Und genau hier liegt ein sehr wesentlicher Grund, warum bis heute ganz überwiegend wissenschaftliche Anwendungen das Grid-Feld beherrschen, während kommerziell eingesetzte Grid-Anwendungen bisher nur vergleichsweise selten zu finden sind. Einige dieser kommerziellen Grid-Anwendungen, die jedoch durchweg noch den Charakter von Pilotanwendungen haben, werden im folgenden Abschnitt vorgestellt.

Beispiele für erste industrielle Anwendungen des Grid-Computing

Petrochemie (Shell)

Eine wesentliche Herausforderung der Erdölindustrie besteht darin, anhand realer seismischer Messdaten und darauf basierender komplexer Rechenmodelle möglichst effizient präzise Vorhersagen über die noch abbaubare Kapazität von Erdgas- und Erdöllagerstätten zu treffen. Dafür müssen umfangreiche Datenbestände mit sehr großem Computeraufwand bearbeitet werden. Die Royal Dutch/Shell bemüht sich deshalb, diese hier benötigten Computerressourcen in ei-

nem Grid bereitzustellen, um so die Genauigkeit und Zuverlässigkeit der Ergebnisse erheblich zu verbessern, die Auswertungsgeschwindigkeit deutlich zu erhöhen und den zukünftigen Anforderungen des Marktes noch besser begegnen zu können [5].

Automobilhersteller (DaimlerChrysler)

„DaimlerChrysler spart sich mit Grids an die Spitze“: Mit diesen Worten gab der Konzern unlängst bekannt, dass man für umfangreiche Simulationsrechnungen bei der Fahrzeugentwicklung künftig die Grid-Technologie einsetzen werde. Erstes Anwendungsgebiet wird die Simulation von Crash-Tests sein, weil dort heute bis zu 70 Prozent aller im Rahmen der Fahrzeugentwicklung aufgewendeten Rechnerressourcen verbraucht werden, was auf den bisher dafür verwendeten Supercomputern erhebliche Kosten verursacht. Die von DaimlerChrysler gesetzten Ziele sind hoch: „Die Automobilindustrie muss sich weltweit selbst neu erfinden, um den Konkurrenzkampf zu gewinnen und den Eintritt ins 21. Jahrhundert zu schaffen. Für uns ist dabei das sicherste Zeichen für Produktqualität ein kontinuierliches Herunterfahren der Ausgaben für Garantieleistungen. Dazu gehört auch eine gut berechnete Schadensbegrenzung, wie sie im Crash-Test-Center geliefert wird.“ sagte dazu der Chrysler-Vorstandsvorsitzende Dieter Zetsche [6].

Luft- und Raumfahrt (EADS)

Ein großer Anteil der Computerressourcen der „European Aeronautic Defence and Space Com-

pany“ (EADS) wird im Umfeld der Entwicklung komplexer Raumfahrttechnik für Simulationsrechnungen eingesetzt. Im Rahmen eines Pilotprojektes prüft die EADS nun die Ablösung herkömmlicher zentraler Computer durch Grid-Architekturen. Das Unternehmen erhofft sich davon eine deutliche Verkürzung der Analyse-, Simulations- und Entwicklungszeiten sowie eine nachhaltige Senkung der für die IT aufzuwendenden Kosten bei gleichzeitiger Verbesserung der Produktqualität, was die EADS in eine dauerhafte und nachhaltige Festigung ihrer Marktposition umsetzen möchte [5].

Pharmazeutische Unternehmen (Novartis)

„Schlafende PC mutieren zum Supercomputer“ – so sieht das Schweizerische Pharmazieunternehmen Novartis seine Grid-Philosophie. Hier werden nach dem Vorbild des SETI-Projektes [7] tausende von PCs im Unternehmen immer dann zu einem Grid zusammengeschaltet, wenn sie nicht für andere Zwecke benötigt werden. So entsteht von Fall zu Fall und insbesondere nachts ein leistungsfähiger Rechnerverbund, der in der Forschungs- und Entwicklungsstrategie von Novartis inzwischen eine zentrale Rolle spielt. Nach Angaben von Novartis erreichen die bisher 2700 miteinander verbundenen Pentium-4-Prozessoren, welche an mehreren Standorten von Basel über Wien bis Cambridge (USA) stehen, eine kumulierte Rechenleistung, welche an die Leistung der stärksten Supercomputer heranreicht. Die vergleichsweise geringe Bandbreite von

100 Megabit (Ethernet) sei dabei ausreichend, weil dieses Grid zur Zeit nur für rechen-, aber nicht datenintensive Aufgaben genutzt werde [8].

Gießereitechnik (Fraunhofer-Gesellschaft)

Ein Beispiel für den erfolgreichen Einsatz der Grid-Technologie bei KMU's ist „Magmasoft“, eine vom Fraunhofer-Institut für Techno- und Wirtschaftsmathematik in Kaiserslautern entwickelte Software zur Verbesserung der Qualität von Gießereiprodukten. Dabei werden aus der Geometrie eines Bauteils mögliche Schwachstellen abgeleitet und sichtbar gemacht. Aufwändige Simulationsrechnungen, die anschließend in einem Grid-Umfeld durchgeführt werden, erlauben dann anhand von Veränderungen von Form, Material und anderen

Parametern eine kostengünstige Optimierung der Gießereiprodukte. Da gerade KMUs nicht über hohe Computerkapazitäten verfügen, bietet sich Grid-Computing in diesem Umfeld als viel versprechende Lösung für die Zukunft an [9].

Zusammenfassung und Ausblick

Innerhalb weniger Jahre hat das WWW die Welt schneller und grundlegender verändert, als jemals zuvor eine andere technische Erfindung binnen so kurzer Zeit. Mit Grid-Computing erleben wir soeben die ersten Anzeichen einer weiteren Revolution der Informationstechnologie. Es kann heute als sicher gelten, dass Grid-Computing die wissenschaftliche, aber auch die kommerzielle und die private Welt

noch weit stärker und nachhaltiger verändern wird, als wir das mit dem WWW erlebt haben. Noch sind „IT-Ressourcen aus der Steckdose“ eine Vision; aber es wird nur wenige Jahre dauern, bis unsere Welt ohne Grid-Computing nicht mehr vorstellbar sein wird. Die milliardenschweren Entwicklungsarbeiten der größten Unternehmen der Welt lassen keinen anderen Schluss zu.

Abkürzungen:

| | |
|-------------|---|
| GGF: | Global Grid Forum |
| IT: | Informationstechnik |
| KMU: | Kleine und mittelständische Unternehmen |
| WWW: | World Wide Web |

Literatur

- [1] I. Foster, C. Kesselmann, „The Grid – Blueprint for a New Computing Infrastructure“, Morgan Kaufmann Publ. 1999
- [2] IBM Corporation, *Integrated Marketing Communications*, Broschüre GM13-0474-00
- [3] Global Grid Forum: <http://www.ggf.org>
- [4] GGF Enterprise Grid Requirements Research Group (EGR-RG): <http://forge.ggf.org/projects/egr-rg>
- [5] IBM Corporation, Systems Group, Broschüre G522-2580-00
- [6] silicon.de Newsletter, 26.05.04: <http://www.silicon.de/cpo/news-csh/detail.php?nr=14796>
- [7] Seti@home: <http://setiathome.ssl.berkeley.edu>
- [8] Neue Züricher Zeitung Online, 07.11.03
- [9] Fraunhofer Magazin 01/2003, Seite 14 ff: <http://www.fraunhofer.de/fhg/archiv/magazin/pflege.zv.fhg.de/german/publications/df/df2003/mag1-2003-14.pdf>