

Die neue Netzmaschine

Grid Computing wurde zwar als basisdemokratisches, kollektives Netzwerkzeug erdacht, aber die aktuelle funktionierende Software elektrisiert vor allem Forscher, Informatiker und IT-Industrie. Die etwas anämische deutsche Grid-Gemeinde bekommt jetzt von Politik und Industrie flott Rückenwind. Rüdiger Berlich



1998 veröffentlichten Ian Foster und Carl Kesselman ihr Buch „The Grid – Blueprint for a New Computing Infrastructure“. Das Ereignis gilt als der Beginn einer neuen Ära des verteilten Rechnens (**Abbildung 1 und 2**). Fosters und Kesselmans Vision war es, auf der Basis standardisierter Schnittstellen transparent nutzbare, verteilte Ressourcen für jedermann verfügbar zu machen. Rechenpower, Informationen und andere Computer-Dienstleistungen sollten erhältlich sein wie der Strom aus der Steckdose. Von dieser 230-Volt-Vision leiteten die Gründer auch ihren Namen ab – von Power Grid (Stromnetz) an. Bereits 1991 hatte Tim Berners-Lee mit dem World Wide Web am Genfer Cern (Centre Europeenne pour la Recherche Nucleaire) Ähnliches erdacht. Und wie damals beim WWW gehen heute wieder wesentliche Impulse fürs Grid Compu-

ting vom Cern aus. Mancher verspottet die Grid-Technologie deshalb gerne schon als „Web on Steroids“ – als ein gedoptes Web.

Grid – was soll das sein?

Grid Computing gilt zurecht als wichtige Zukunftstechnologie, weshalb bei jeder Nennung des Begriffs zügig Forschungsgelder sprudeln. Forschungsgruppen, die auch nur entfernt etwas mit verteiltem Rechnen zu tun haben, nehmen das Zauberwort in ihre Projektanträge auf. In diesem Treibhausklima den Grid-Begriff scharf zu definieren fällt schwer, driftet er doch zusehens in Richtung „verteilt Rechnen“.

Die Definition orientiert sich notwendigerweise an der Praxis: Grob unterscheidet man zwischen der ursprüngliche Vision von Foster und Kesselman und den-

Inhalt

- 48 Globus-Toolkit, Version 2**
Die meisten Anwenderprojekte benutzen diese Globus-Version für ihr – meist weltweit – verteiltes Rechenzentrum.
- 51 Safer Grid**
Die Grid Security Infrastructure authentifiziert und autorisiert die Benutzer von Globus-Grids.
- 54 Globus-Toolkits 3 und 4**
Die mit Webservice-Technologie arbeitenden neuen Versionen bringen Vorteile, die Umstellung bestehender Installationen ist aber schwierig.
- 58 European Data Grid**
Die EDG-Middleware haben 21 Organisationen gemeinsam entwickelt.
- 62 Alien im Wunderland**
Das Alice Environment ist die Grid-Variante für den neuen Cern-Teilchenbeschleuniger.
- 66 Das 10. Global Grid Forum**
Berlin war im März Zentrum der internationalen Grid-Aktiven. Mit gutem Grund tauchte dort auch Politprominenz auf.

jenigen, die ein real existierendes Problem mit Hilfe weit verteilter Ressourcen lösen wollen. Zur zweiten Gruppe zählen die Teilchenphysiker. Tausende über die Welt verstreut arbeitende Wissenschaftler wollen ab 2007 mit vielen Petabyte dezentral gespeicherter Daten der Experimente des Large Hadron Colliders (LHC) des Cern hantieren. Die Rechenressourcen bei jedem einzelnen Wissenschaftler vorort reichen keinesfalls. Auch wäre es aufgrund der hohen Datenmengen sehr viel einfacher, die Programme zu den Daten zu bringen, als umgekehrt.

Der Gedanke an eine Infrastruktur liegt nahe, die die insgesamt riesigen Speicher und die zehntausende CPUs virtuell zueinander rückt. Herkömmliche Clustertechnologien taugen angesichts dieser Größenordnung und wegen des Zoos unterschiedlicher Hardware hierfür nicht



Abbildung 1: Carl Kesselman auf der Cern School of Computing 2002 in Vico Equense, in der Nähe von Neapel. Der Mitautor des ersten Grid-Buchs ist einer der beiden geistigen Väter des Grid Computings.



Abbildung 2: Ian Foster – er gilt wie Carl Kesselman als Grid Father – am Stand von Sun auf der Supercomputing 2001 in Denver. Der Slogan „Sun Powers The Grid“ zeigt das eindringliche Interesse der Industrie an der neuen Technik.

– ein neuer Ansatz muss her: Grid Computing. Grids eignen sich für die Teilchenphysik auch deshalb so gut, weil die Wissenschaftler oft gleiche Programminstanzen separate Teildatensätze parallel berechnen lassen. Die Teilchenphysik ist damit, wenn schon nicht der Auslöser, so doch die Triebfeder des Grid Computings.

Beim Grid Computing geht es also um die weltweite Verteilung identischer Programminstanzen, ähnlich den üblichen Batchsystemen in lokalen Clustern. Mit etwas Wohlwollen könnte man diese Grid-Anwendung als Cluster von Clustern verstehen. Den Cluster-Aspekt bringen auch der **Kasten „Prinzipien des verteilten Rechnens“** und der Beitrag über das European Data Grid ins Spiel.

Echte parallele Applikationen, die viele Daten zwischen einzelnen Rechenknoten austauschen, werden dagegen kaum je eine tragende Rolle im Grid Computing einnehmen.

Anderes ist von verteilten Datenbanken zu erwarten: So wären Grids im Gesundheitswesen (Zugriffe auf Patientendaten), beim Koppeln von Informationen einer weltweit agierenden Firma oder als Suchmaschinen-Technologie sehr effektiv einsetzbar.

Netze, Filesysteme und Middleware

Um sich dem Ziel des weltweit verteilten Rechnens zu nähern, sind vielerorts neue Entwicklungen und Verbesserun-

gen notwendig. Ein schnell offensichtliches Muss sind Hochgeschwindigkeitsnetze. Die in den Industrieländern recht gut bestückten nationalen Netze müssen nach und nach zu einem „World Wide Grid“ zusammenwachsen.

Die Netztechnologie ist vorhanden und die nominellen Bandbreiten wachsen auch hinreichend schnell. Die Geschwindigkeit des Ausbaus ist dabei mehr eine Frage von nationalen Budgets und Interessen, als die der Technik. Die dominierende Rolle spielt in Europa GÉANT [1], eine Kooperation von 26 nationalen Forschungsnetzen.

Viel Arbeit wird momentan in die Entwicklung verteilter Filesysteme gesteckt, die zum Betrieb der lokalen Rechenzentren – den Rechenknoten des Grid – un-

erlässlich sind. In manchen Fällen sind sie sogar global nutzbar, obwohl dem natürlich die in Weitverkehrsnetzen sehr hohe Latenzzeit entgegen steht (siehe **Kasten „Prinzipien des verteilten Rechnens“**).

Das Ausführen eines Programms in einem Grid setzt in der Praxis eine Middleware voraus. Ihre Rolle ist ähnlich der des Netzwerklayers eines Betriebssystems: Eine Applikation, die Daten über ein Netzwerk transferieren will, braucht nicht über die verwendete Netzwerkhardware informiert sein. Analog soll sich eine Grid-Applikation nicht um Fragen der Authentifizierung und Autorisierung oder um Fragen der Abrechnung kümmern müssen – das ist Aufgabe der Middleware.

Die Verhüll-Möglichkeiten der Middleware hat aber Grenzen. Solange die Grid-Benutzer Programme verschicken, die die Zielsysteme interpretativ abarbeiten, ist alles in Butter. (Für Interpreter mit Just-in-Time-Compilern gilt das Glei-

che.) Sobald aber fertig kompilierte Binaries auftauchen, muss diesem Umstand Rechnung getragen werden – unvorhersagbar in einem großen Grid ist, ob am anderen Ende eine 64-bittige Risc-Architektur, ein 32-Bit-Intel-System oder etwas ganz anderes lauert.

Hieraus ließe sich ein Bedeutungszuwachs für Java und C# konstruieren, die solche Unterschiede verstecken. Andererseits sehen moderne Grid-Middlewares vor, die Ausführung einer Grid-Applikation auf vorher definierte Architekturtypen zu beschränken.

Kreativität kontra Standards

Es mag auf die Eigenheiten freier und unbeschränkter Forschung zurückzuführen sein, dass im Bereich der Middleware ein Chaos fast babylonischen Ausmaßes herrscht. Einerseits ist es gut und normal, dass in einem kreativen Umfeld viele kreative Ansätze unterschiedlich-

ter Art miteinander konkurrieren – man denke an Mailclients oder die Desktopumgebungen unter Linux. Andererseits erwarten wirtschaftlich interessierte Anwender Standards, um ihre Programme schreiben zu können.

Die Evolution wird die Lösung bringen: Wie bei Open Source sonst meist auch, wird sich nach und nach der beste Kandidat durchsetzen. Der Schwerpunkt dieses Linux-Magazins beschreibt die heißesten Kandidaten in diesem Rennen. Aktuell und bei der Anzahl der laufenden Installationen liegt die Middleware Globus-Toolkit vorn – der zugehörige Artikel hier im Schwerpunkt folgt darum auch sofort. Ian Foster und Carl Kesselman, die beiden Autoren der Grid-Computing-Bibel, verleihen Globus durch ihre Mitgliedschaften in Globus-Forschungsgruppen hohe Weihen.

Allerdings sorgt das Globus-Projekt mit seiner Programm- und Versionsvielfalt selber für Verwirrung im Grid-Universum, wenn auch aus guten Gründen.

Prinzipien des verteilten Rechnens

Grid-Anwendungen unterliegen den selben Gesetzmäßigkeiten wie jede verteilte Anwendung: Bandbreite und Latenzzeit des Netzwerks dominieren die Leistungsfähigkeit der meisten Applikationen. Die Bandbreite eines Netzwerks gibt an, welche Datenmenge es pro Zeiteinheit übertragen kann. Die Latenzzeit misst, wie lange ein Signal vom Sender zum Empfänger benötigt. Unter Linux ermittelt das »ping«-Kommando diesen Wert, es misst jedoch die doppelte Latenzzeit (den so ge-

nannten Roundtrip vom Sender zum Empfänger und wieder zurück). Vom Forschungszentrum Karlsruhe zur Ruhr-Universität Bochum liegt die Latenz bei etwa 20 ms (**Abbildung 3**), in einem lokalen Netzwerk wenigstens eine Größenordnung darunter. Latenzzeiten in Multiprozessorsystemen sind im Vergleich dazu fast vernachlässigbar klein.

Manche Anwendungen lassen sich von hoher Latenz und niedriger Bandbreite nicht beeindrucken. Dasselbe Programm bearbeitet gleichzeitig je einen Teildatensatz auf verschiedenen – eventuell weltweit verteilten – CPUs, die einzelnen Instanzen tauschen jedoch keine Informationen aus.

Diese Applikationen bezeichnet man oft als „embarrassingly parallel“ (beschämend parallel), oder optimistischer als „nicely parallel“ (angenehm parallel). Sie stellen keine besonderen Ansprüche an die Programmierung. Unter Umständen braucht der Entwickler nicht einmal zu wissen,

dass seine Software später in mehreren Instanzen parallel läuft. Am Ende des Programms müssen Anwender oder Software nur die Ergebnisse der Einzelberechnungen zusammenführen.

Datenaustausch bremst die Applikation

Herkömmliche parallele Cluster-Applikationen tauschen bereits während der Ausführung häufig Daten aus. Das wird zum Problem, wenn eine Programminstanz auf das Ergebnis einer anderen warten muss, um weiterrechnen zu können. Dann stören die hohe Latenzzeit und niedrige Bandbreite der WANs, besser geeignet sind lokale Netzwerke (LAN oder Cluster) oder Multiprozessorsysteme.

Aus dieser Unterscheidung folgt, welche Applikationen sich für ein Grid eignen. Da sie meist im WLAN operieren, entscheidet das Ausmaß der Kommunikation zwischen den Rechenknoten, ob eine Applikation sinnvoll im Grid arbeitet. Zwar stellt das Grid potenziell sehr hohe Rechenressourcen zur Verfügung. Jedoch verpufft die Leistung, wenn die Knoten ihre Rechenzeit damit verschwenden, auf die Antwort eines anderen Knotens zu warten.

Vom Grid Computing profitieren heute vor allem die Teilchenphysik und andere Forschungszweige, die sehr große Datenmengen in möglichst kurzer Zeit analysieren müssen. Sie erstellen fast ausschließlich Applikationen, die „nicely parallel“ arbeiten.

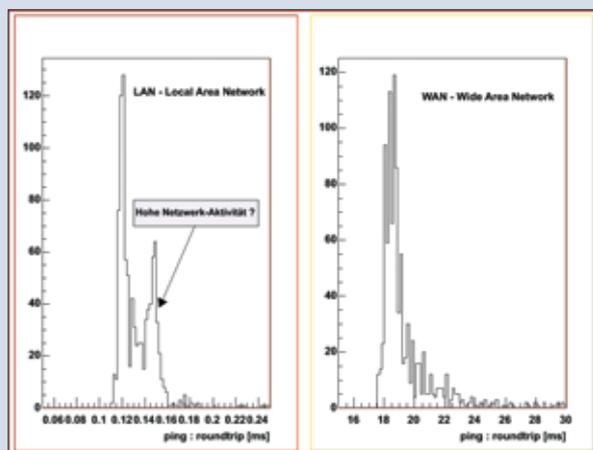


Abbildung 3: Die Häufigkeitsverteilung der Roundtrip-Zeiten zweier Netzverbindungen (doppelte Latenz, gemessen mit »ping«). Der linke Graph zeigt ein LAN, der rechte die Verbindung vom Forschungszentrum Karlsruhe zur Universität Bochum. Im WAN liegen die Werte meist bei 18 bis 20 Millisekunden, im LAN zwischen 0,11 und 0,16.

Die eher monolithisch strukturierte Version 2 des Globus-Toolkits (GT 2) darf sich vieler Installationen rühmen. Mit Version 3 (GT 3) führt Globus die Open Grid Service Architecture (OGSA) ein, eine auf so genannten Grid Services aufgebaute Infrastruktur. Der Paradigmenwechsel induziert bei den Planern schon laufender Grid-Projekte allerdings viel Unsicherheit.

Paradigmenwechsel bei Globus-Toolkits

Die Adaptionen an GT 3 waren bereits in vollem Gange, als sich zu Beginn dieses Jahres ein erneuter Versionswechsel abzuzeichnen begann. GT 4 versucht nun, eine Kompatibilität mit herkömmlichen Webservices herzustellen. Details zu der mittlerweile final vorliegenden Version 3 und Hintergründe zu GT 4 verrät der Beitrag ab Seite 54.

Der Artikel „Safer Grid“ ab Seite 51 behandelt einen sehr wichtigen Aspekt des Grid Computings: die Sicherheit. Wenn sich weltweit Tausende von Rechner

übers Internet vernetzen, ist das natürlich eine Herausforderung für die ebenso lästige wie unternehmungslustige Skript-Kiddie-Fraktion. Am Beispiel der Globus-Komponente Grid Security Infrastructure (GSI) zeigt der Beitrag die benutzten Authentifizierungs- und Autorisierungsprinzipien.

Das European Data Grid (EDG) stellt mit seiner Middleware solche Dienste bereit, die Globus bei seinen Toolkits bewusst ausgeklammert hat. Hierzu zählt der so genannte Ressource Broker, der Grid-Applikationen transparent auf passende Rechenressourcen verteilt (siehe Artikel ab Seite 58). EDG ist ein echtes europäisches Gemeinschaftsprojekt und hat umfangreiche Geldmittel und Manpower zu seiner Verfügung. Das Projekt wird allerdings in naher Zukunft durch EGEE (Enabling Grids for E-Science in Europe) abgelöst, das auf den Ergebnissen des European Data Grid aufbaut.

Alien (Alice Environment, [5]), ein Produkt des Alice-Experiments am Large Hadron Collider, ist ein Paradebeispiel für die Macht von Open Source. Anders als das EDG ist es nämlich keine komplette Neuentwicklung, sondern benutzt – wo immer es geht – schon vorhandene Perl-Module. Mit einer nur wenige

Entwickler umfassenden Mannschaft hat Alice es mit „Extreme Programming“-Techniken geschafft, eine funktionierende Grid-Umgebung zu programmieren, die sich in der Funktionalität durchaus mit EDG messen kann. Der Artikel „Alien im Wunderland“ stellt das ambitionierte Projekt vor.

Auf die Ergebnisse des deutschen Gemeinschaftsprojekts Unicore (Uniform Interfaces to Computing Resources, [2]) greifen insbesondere Supercomputing-Anwendungen zurück, beispielsweise der Deutsche Wetterdienst. Wie Globus und EDG setzt Unicore auf eine Art verteiltes Batchsystem. Neben den gerade genannten Middlewares trifft man noch häufiger auf Namen wie Cactus [3], Legion [4] und Condor [6].

Standards, Grid Forum und D-Grid

Die Vielzahl und die Eigenständigkeit der einzelnen Projekte machen klar, dass das Grid Computing einheitlicher Standards und Protokolle dringend bedarf. Als Standardisierungsgremium etabliert sich das Global Grid Forum (GGF). Es soll einmal eine ähnliche Stellung einnehmen wie die IETF (Internet Engineering Taskforce) für das Internet. Die Grid-Hochkaräter treffen sich mehrmals im Jahr an wechselnden Orten, um sich gegenseitig auf dem Laufenden zu halten. Das 10. Global Grid Forum fand

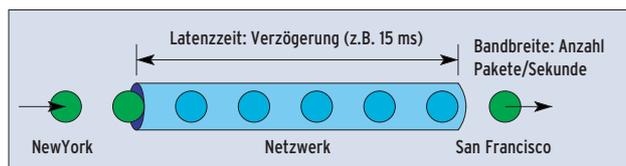


Abbildung 4: Latenzzeit und Bandbreite eines Netzwerks setzen dem Grid Computing Grenzen. Insbesondere die Latenzzeit, also die Verweildauer eines Datenpakets im Netzwerk, lässt sich nicht beliebig verkürzen.

Die Suche nach dem Unteilbaren

Die Elementarteilchenphysik sucht nach den Grundbausteinen der Materie und will die Kräfte beschreiben, die zwischen ihnen wirken. Ein elementares Teilchen weist nach heutigem Verständnis keine innere Struktur auf. Freilich glaubten Forscher schon häufig, ein gerade gefundenes Teilchen sei elementar. Diese Feststellung erwies sich bei genauerem Hinsehen fast ebenso oft als Trugschluss. Atome sind nach heutigen Maßstäben schon fast riesige, komplexe Objekte. Sie bestehen aus Elektronen, Protonen und Neutronen, Letztere wiederum aus Quarks, und das ist noch nicht das Ende.

Riesige Ringe für kleinste Teilchen

Zur Suche nach neuen Teilchen und Substrukturen in bekannten Teilchen setzt die Forschung große Beschleunigersysteme ein, etwa den PEP-II-Ring am Stanford Linear Accelerator Centre (SLAC) in Kalifornien oder die Beschleunigerringe des Cern in Genf. In diesen Beschleunigern werden unterschiedliche Teilchensorten zur Kollision gebracht.

Das 1989 begonnene Large Electron Positron Colliders (LEP) am Cern beschleunigte Elektronen und ihre Antiteilchen, die Positronen, in einer 27 Kilometer umfassenden Kreisbahn und brachte sie an vier Wechselwirkungspunkten zur Kollision. Der LEP-Tunnel erstreckt sich unter Genf und dem schweizerisch/französischen Jura. Die Ausmaße dieses Ringes sind so groß, dass seine Kalibration neben den üblichen Gezeiteffekten auch vom saisonalen Wasserstand des Genfer Sees abhing, der das Umland verschieden stark senkt.

LEP wurde vor kurzem zugunsten eines noch größeren Nachfolgeprojekts stillgelegt. Der neue Large Hadron Collider (LHC) wird derzeit im alten LEP-Tunnel errichtet. Zusammen mit seinen Experimenten – darunter Atlas und Alice – soll der Beschleuniger ab 2007 bislang unerreichte Energiebereiche erschließen. Um zu kleineren Maßstäben vorzudringen, ist immer höhere Energie nötig. Manche Teilchensorten sind so schwer (und nach Einsteins $E=mc^2$ auch entsprechend energiereich), dass man sie mit bisher erreichbaren Energien nicht erzeugen kann. Ein Beispiel dafür ist das lang gesuchte Higgs-Boson.

Füllt alle 16 Sekunden eine 80-GByte-Platte

Die höhere Energie hat auch höhere Datenraten zur Folge, die die Wissenschaftler speichern und verarbeiten müssen. Beispielsweise wird das Alice-Experiment schwere Ionen miteinander kollidieren lassen. Für jede Kollision sind Tausende von Spuren geladener und neutraler Teilchen zu rekonstruieren. Die Daten dazu stammen aus den Signalen verschiedener wie Zwiebelschalen angeordneter Subdetektoren. Wegen der hohen Kollisionsrate produzieren die LHC-Experimente zusammen pro Sekunde etwa 40 GBit Daten. Sie füllen also alle 16 Sekunden eine handelsübliche 80-GByte-Platte. Die Experimente sollen über viele Jahre laufen.

Mit den Datenraten hat sich auch die Rechnerinfrastruktur in der Teilchenphysik verändert. Wo noch vor zehn Jahren Unix- und VMS-Maschinen sowie gelegentlich Mainframes die Daten verarbeiteten, kommen heute fast aus-

schließlich Rechnerfarmen auf Linux-Basis zum Einsatz. Der Siegeszug des freien Betriebssystems startete am Cern um 1995, als die ersten Physiker Linux auf ihren Arbeitsplatzrechnern einsetzten.

Der Wechsel zu Linux führte aber auch zu unerwarteten Problemen. Ein Physiker möchte, dass sein Programm immer dasselbe Ergebnis liefert, egal wo es läuft. Unterschiedliche Distribution und verschiedene Versionen einer Distribution verwenden unterschiedliche Versionen von Bibliotheken und Kernel. Das beeinflusst aber die Rechengenauigkeit, besonders die Mathlib hat sich als kritisch herausgestellt. Bei den einheitlicheren Unix-Plattformen war dieses Problem unbekannt.

Haupttriebfeder des Grid Computing

Die erwarteten Datenmengen des LHC stellen enorme Anforderungen an die Rechnerinfrastruktur. Die Applikationen brauchen schnelle Netze, wahlfreien Zugriff auf einzelne Datensätze und enorme Rechenkapazität. Statt alle Daten zentral zu speichern und zu verarbeiten, wollen die LHC-Wissenschaftler dezentral die existierenden (oder neu aufgebauten) Rechenressourcen der teilnehmenden Länder nutzen. Das Ziel ist es, die Rechen- und Speicherlast zu verteilen – daher treibt dieses Projekt das moderne Grid Computing voran.

In der Kombination aus Teilchenphysik und Grid Computing erweist sich Linux als Segen und Fluch zugleich: Es bewährt sich als günstige und stabile Plattform, führt durch seine Vielfalt aber zu beinahe babylonischen Verhältnissen in den Rechenzentren.

im März 2004 an der Humboldt-Universität in Berlin statt (siehe Beitrag ab Seite 66). Die Ortswahl mag verblüffen, denn unter den großen Grid-Initiativen, vom Malaysia-Grid bis hin zu halb militärischen Installationen, beispielsweise beim amerikanischen Department of Energy [7], sucht man eine deutsche vergebens. Doch just auf dem 10. GGF stellte die Ministerin für Wissenschaft und Forschung eine fach- und organisationsübergreifende deutsche Grid-Initiative vor – das D-Grid.

Das programmierte Chaos

Von „World Wide“ ist im Zusammenhang mit dem Grid heute noch nicht zu reden. Davon ist es genauso weit entfernt wie von einheitlichen Standards. Auf Konferenzen scherzen die Teilneh-

mer gelegentlich vom G-Wort (in Anlehnung an das schlimme F-Wort). Dabei mangelt es nicht mal an Geld oder Mitarbeitern. Das Problem ist eher, dass viele kluge Köpfe viele (oft kluge) Lösungen hervorbringen – wie überall sonst in der Forschung auch.

Gespannt darf man sein, welche Auswirkungen das zunehmende Interesse der Wirtschaft an der neuen Technologie haben wird. Zwar sind Insider über das gegenwärtig noch herrschende Chaos weder überrascht noch sonderlich besorgt. Doch kommerzielle Verwertung verlangt Standards – also Schluss mit dem Chaos. Vielleicht war die Gründung der fachübergreifenden deutschen Grid-Initiative auch in diesem Sinne überfällig und beschert Deutschland den Anschluss an die bereits etablierten Grid-Nationen. (jk) ■

Infos

- [1] Europäische Hochgeschwindigkeits-Netzwerk GÉANT: [<http://www.dante.net/geant>]
- [2] Unicore Middleware: [<http://www.unicore.org>]
- [3] Cactus Middleware: [<http://www.cactuscode.org>]
- [4] Legion Middleware: [<http://legion.virginia.edu>]
- [5] Alien: [<http://alien.cern.ch>]
- [6] Condor: [<http://www.cs.wisc.edu/condor>]
- [7] DoE Grid: [<http://doesciencegrid.org>]
- [8] Global Grid Forum: [<http://www.ggf.org>]

Der Autor

Rüdiger Berlich studierte Physik an der Universität Bochum. Ab 1998 arbeitete er bei Suse, unter anderem als Technical Manager (Support) der US-Filiale und als Geschäftsführer der britischen Niederlassung. Im Januar 2004 promovierte er an der Universität Bochum im Bereich der Teilchenphysik und des verteilten Rechnens.