

Publikationsreihe des John von Neumann-Instituts für Computing (NIC)
NIC-Serie

Band 19

John von Neumann-Institut für Computing (NIC)

Auswirkungen der Virtualisierung auf den IT-Betrieb

Wolfgang Gürich (Hrsg.)

GI-Fachtagung, 4. - 5. November 2003
Forschungszentrum Jülich
Proceedings

NIC-Serie

Band 19

ISBN 3-00-009100-9

Die Deutsche Bibliothek - CIP-Einheitsaufnahme
Ein Titeldatensatz für diese Publikation ist bei
Der Deutschen Bibliothek erhältlich.

Herausgeber: NIC-Direktorium
Vertrieb NIC-Sekretariat
Forschungszentrum Jülich
52425 Jülich
Deutschland
Internet: www.fz-juelich.de/nic
Graphische Betriebe, Forschungszentrum Jülich

© 2003 John von Neumann-Institut für Computing

Es ist erlaubt, dieses Werk oder Teile davon digital oder auf Papier zum persönlichen Gebrauch oder zu Lehrzwecken zu vervielfältigen, vorausgesetzt die Kopien werden nicht kommerziell genutzt. Kopien müssen diese Copyright-Notiz und das volle Zitat auf ihrer Titelseite enthalten. Andere Vervielfältigung bedarf der vorherigen schriftlichen Genehmigung des oben genannten Herausgebers.

NIC-Serie Band 19
ISBN 3-00-009100-9

Vorwort

Dieser Band erscheint zur Tagung BIK2003 der Fachgruppe „Betrieb von Informations- und Kommunikationssystemen (BIK)“ der Gesellschaft für Informatik. Es ist bereits die 15. Tagung dieser Reihe, die im zweijährigen Abstand die Entwicklungsgeschichte von Rechen- und Kommunikationszentren seit 30 Jahren begleitet. Neben dem beständigen schnellen Wandel bei der Technik der Informationssysteme, beeinflusste die zunehmende Ferne der Benutzer durch Nutzung von Kommunikationsnetzen und durch den Aufbau eigenständiger dezentralisierter Einheiten die Arbeit und den Betrieb der klassischen Rechenzentren. Diese Ferne der Nutzer ist, unterstützt durch die räumliche Entfernung, in erster Linie verbunden mit divergierenden Zielvorstellungen. Die Erfüllbarkeit von Forderungen nach preiswerten Diensten hoher Funktionalität, Aktualität und ständiger Verfügbarkeit muss mit dem Anspruch der Zentren bezüglich Betriebssicherheit, Wartbarkeit, Nachhaltigkeit und personellem Einsatz in Einklang gebracht werden.

Das Thema dieser Tagung stellt die Auswirkungen auf die sich abzeichnende Virtualisierung der IT heraus. Die Virtualisierung bringt einerseits für die Nutzer eine Vereinfachung im Verständnis von Funktionen und Anwendungen, führt aber andererseits für den Betreiber zu einer erheblichen Erhöhung der Komplexität des Gesamtsystems. (Diesen Zusammenhang habe ich in einem eigenen Beitrag zum Tagungsband ausgeführt, da diese Erörterung den Rahmen eines Vorwortes sprengen würde.) Die resultierenden Herausforderungen werden die Arbeiten der Informations- und Kommunikationszentren in den nächsten Jahren bestimmen.

Ziel dieser Tagung, wie auch ihrer Vorgängerveranstaltungen, ist es, den Dialog zu fördern zwischen den für den Betrieb Verantwortlichen und den Anbietern von Lösungen, Konzepten und Produkten für den Einsatz in Informations- und Kommunikationszentren.

Die Ausrichter dieser Tagungsreihe haben sich in den letzten Jahren verstärkt bemüht, Beiträge aus Industrie und Wirtschaft in den Erfahrungsaustausch einzubeziehen. Dabei nahmen sie in Kauf, dass sich der Charakter der Tagung von dem einer „wissenschaftlichen“ Fachtagung wegbewegte. Das Wesen der Wissenschaftlichkeit ist die Abwägung des Für und Wider der Argumente innerhalb eines einzelnen Beitrages. Es dürfte klar sein, dass bei einem Firmenbeitrag, der ein kommerzielles Produkt zum Inhalt hat, diese Art der wissenschaftlichen Selbstverleugnung nicht gefordert werden kann. Hier ist es Aufgabe der Teilnehmer bzw. der Leser, die fehlende innere Dialektik durch den Vergleich der Beiträge zum selben Themenkreis zu ersetzen, ähnlich wie sich

eine abgewogene Meinungsbildung nicht durch Lesen einer Tageszeitung, sondern aus dem Vergleich mehrerer Quellen speist.

Die Durchführung einer Tagung ist nur mit vielfältiger Unterstützung möglich. Dafür, dass mir diese in erheblichem Umfang gewährt wurde, möchte ich mich herzlich bedanken bei den Autoren der Beiträge zum Tagungsband und der Vorträge, bei den unterstützenden Institutionen und Sponsoren, beim Tagungsbüro des Forschungszentrums und bei den Helfern aus dem Zentralinstitut für Angewandte Mathematik, hier insbesondere bei Frau Elke Bielitz. Mein besonderer Dank gilt Frau Beate Herrmann, die maßgeblich am Zustandekommen dieses Tagungsbands beteiligt war.

Jülich, November 2003

W. Gürich

Wege der Virtualisierung in der Informationstechnik <i>Wolfgang Gürich</i>	1
On Demand Business und On Demand Computing: Eine technologische Perspektive <i>Walter Hehl</i>	9
Web Services – Das Netz ist der Rechner <i>Karl-Heinz Bondick</i>	15
Value Computing - Am Beginn einer neuen Ära <i>Franz N. Busch</i>	23
Grids and Grid-Services <i>Joan Massó, Wolfgang Rühmer</i>	29
Grid Computing und virtuelle Organisationen <i>Dietmar Erwin</i>	37
Storage Virtualisation What is it? What does it do? How will we Benefit? <i>Michael D. Senior</i>	47
Beiträge der RWTH Aachen zur kooperativen IT-Versorgung im Lande NRW <i>Christian Bischof, Klaus Brühl, Thomas Eifert</i>	53
Auswirkungen der Virtualisierung auf Transparenz und Fehlerdiagnose in lokalen Netzen <i>Andreas Binzenhöfer, Kurt Tutschku</i>	65
Anwendung von NAS-Systemen im Deutschen Wetterdienst <i>Alexander Harth</i>	75
Hochverfügbarkeit mit Oracle-Datenbanken <i>Torsten Rosenwald, Ingo Frost</i>	85
DBShadow <i>Uwe Stocker</i>	93
Virtuelles SAN für einen Supercomputer <i>Lothar Wollschläger</i>	101
Mobile Hotspots – Status Quo auf dem Weg zum mobilen Computing <i>Andy Schweiger</i>	107

Wege der Virtualisierung in der Informationstechnik

Wolfgang Gürich

Forschungszentrum Jülich GmbH
Zentralinstitut für Angewandte Mathematik
52425 Jülich
E-Mail: w.guerich@fz-juelich.de

Es wird der subjektive Versuch unternommen, den Begriff der Virtualisierung in der IT einzugrenzen und zu definieren. Dabei gibt eine historische Betrachtung der Virtualisierungsbemühungen in der Rechnerentwicklung und der Vergleich mit der virtuellen Energieversorgung Hinweise auf Ziele der Virtualisierung aber auch Einsichten bezüglich ihrer Beherrschbarkeit.

1 Einführung

Obwohl in den einschlägigen Fachzeitschriften derzeit fast in jeder Ausgabe mindestens ein Artikel erscheint, der sich mit Tendenzen einer zunehmenden Virtualisierung in der IT befasst, wird das Thema im Alltag der Rechenzentren nur sehr zögerlich behandelt. Dies liegt sicher zum großen Teil an der noch nicht abgeschlossenen Meinungsbildung, die zu einem gemeinsamen Verständnis über den Begriff Virtualisierung führen muss.

Hier soll der Versuch unternommen werden, zu diesem Klärungsprozess beizutragen und ihn voranzubringen.

2 Die Wurzeln

Ähnlich wie bei Aufsätzen über die Geschichte der Datenverarbeitung, die die Disziplin mit der Erfindung des Abakus beginnen lassen, wird auch bei der Virtualisierung z.T. tief in die Geschichte zurückgegangen. So bezeichnet Margaret Wertheim [1] die Ausgestaltung mittelalterlicher Kirchen mit Fresken, die der Architektur eine zusätzliche räumliche Tiefe geben als eine Form der Virtualisierung. Ein Beispiel zur Illustration dieser Auffassung ist die Bibliothek des Escorial-Palastes in Spanien (Abb. 1). Die Gemälde fügen der Decke eine Kassettenstruktur hinzu, die nur „virtuell“ vorhanden ist.

Eine weitere Meta-Ebene der Virtualisierung wird heute erreicht, wenn Architektur mit VR-Methoden konstruiert und „begehrbar“ wird [2].



Abb. 1: Bibliothek, El Escorial, Spanien

Diese Art der Virtualisierung, die sich auf eine Modellierung und Simulation der realen Welt bezieht, möchte ich, obwohl hier Datenverarbeitung zum Einsatz kommt, ausdrücklich nicht in die weitere Erörterung des Themas einschließen, sondern die Virtualisierung ausschließlich auf das *Instrument* Datenverarbeitung beschränken und sich den daraus ergebenden Nutzen für die Anwendungen.

In der Entwicklungsgeschichte der Rechner tauchte der Begriff und die Technik des Virtuellen in Form des virtuellen (Haupt-) Speichers um die Jahre 1962 bis 1964 auf. Gleichzeitig gab es mit CP-40 (später CP-67) erste akademische Versionen eines Betriebssystems zur Unterstützung des Konzepts der „virtuellen Maschine“. Den Entwicklern um Gene Amdahl und Geritt Blaauw [3] der sehr erfolgreichen Rechnerfamilie IBM/360 waren diese Konzepte bekannt, wurden aber von IBM bis auf eine Ausnahme noch nicht aufgegriffen. Mit dem Hinzufügen der sog. Blaauw-Box konnten auf dem IBM/360-67 erste Erfahrungen mit den neuen virtuellen Techniken gesammelt werden.

Zum festen Architekturbestandteil wurde der virtuelle Speicher ab 1970 mit der neuen Rechnerlinie IBM/370 und ab 1972 wurde die Weiterentwicklung des Betriebssystems CP-67 als VM/370 (Abb. 2) allgemein verfügbar [4].

Die Unterstützung für das VM-Betriebssystem wurde von IBM 1990 eingestellt, das Prinzip der Aufteilung realer Rechnerressourcen in virtuellen Maschinen existiert aber weiter in dem LPAR-Konzept und z.B. in Produkten der Firma VMware.



Abb. 2: Login bei IBM /370-Rechnerfamilie mit VM-Betriebssystem

Die Virtualisierung bei den Speichersystemen in der Rechnerperipherie begann mit der Entwicklung von Bandrobotersystemen, 1973 zunächst mit dem IBM Mass Storage System 3850 (s. Abb 3a) und später mit dem auf Standard-Kassetten basierenden System von StorageTek (Abb. 3b).



Abb. 3a: IBM Mass Storage System



Abb. 3b: Tape Library von StorageTek

Die Nutzung von automatischen Bandlaufwerken rund um die Uhr ohne Bedienungspersonal wäre an sich schon ein herauszuhebender technischer Fortschritt

gewesen, aber es wurde zusätzlich in Kombination mit Staging-Plattenlaufwerken und entsprechender Software das Prinzip des virtuellen Plattenspeichers eingeführt – aus Sicht des Benutzers/Programmierers also ein nahezu „unendlich großer“ Plattenspeicher, wobei das Bandsystem vor dem Nutzer versteckt bleibt. Interessanterweise wird die umgekehrte Kombination, also „virtual tapes“ erst Jahrzehnte später eingeführt.

3 Klärung des Begriffs

Aus den wenigen Beispielen ist ersichtlich, dass der Begriff der Virtualisierung in der IT immer als zwei Sichtweisen einer Methode, einer Technik oder einer Funktion beschrieben werden kann:

Die erste Sicht, die Detailsicht, ist die eines Entwicklers (oder später eines Betreuers) dieser neuen Technik, die in vielen Fällen aus der neuen Kombination vorhandener Hard- und Software aufbaut unter Hinzufügung weiterer neuer Elemente. Diese Sichtweise zeigt alle Details und wird bei Installation, Konfiguration, Performance-Optimierung und Fehlersuche benötigt.

Die zweite Sicht, die Nutzersicht, betrachtet eine sehr vereinfachte, modellhafte, abstrakte Beschreibung der „virtuellen Funktion“. Diese Beschreibung kann die Beschreibung einer Schnittstelle sein, hinter der die Komplexität der virtuellen Funktion vor dem Nutzer versteckt wird.

Die folgende Abbildung soll dies am Beispiel des virtuellen Hauptspeichers verdeutlichen.

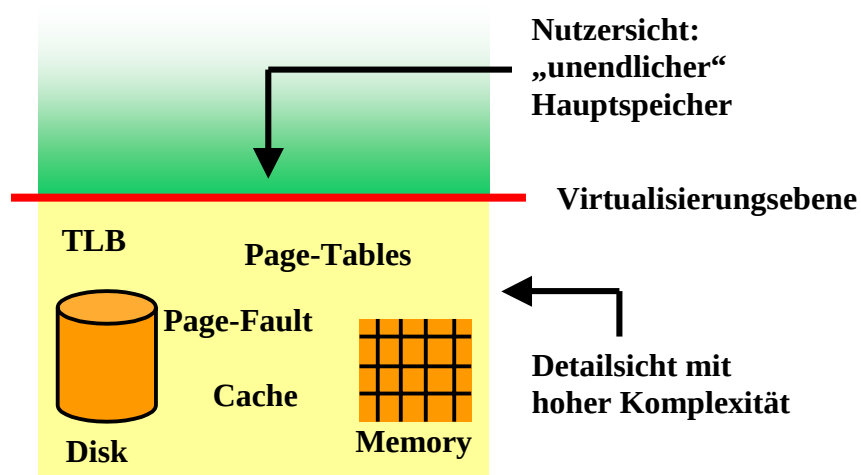


Abb. 4: Virtueller Hauptspeicher

Damit entspricht diese Interpretation von Virtualisierung dem ISO/OSI-Schichtenmodell, wobei die Trennung in Virtualisierungsschichten klare Schnittstellen erfordert, auf die kompatible Produkte und Applikationen unterschiedlicher Hersteller aufbauen können.

4 Neue Dimensionen durch Globalisierung

Die örtlich fixierte Betrachtung auf ein Rechnersystem oder Speichersysteme in einem Rechenzentrum muss seit Erfindung des Internet [5] (TCP: 1974, Übergang von ARPANET zu Internet: 1983-85) aufgegeben und auf den gesamten Globus ausgedehnt werden. Dabei ist das Netz nur eine notwendige Bedingung für globale Virtualisierungsmöglichkeiten. Solange das vermaschte Netz nur zum Aufbau von temporären oder permanenten Punkt-zu-Punkt-Verbindungen genutzt wird, um an entfernten Rechnersystemen Applikationen aufzurufen, führt dies nicht zu einer neuen Qualität gemäß der o.a. Begriffsbestimmung. Auch die Suche nach Informationen mit Hilfe von Suchmaschinen versteckt die Komplexität des zugrunde liegenden Prozesses nicht vor dem Nutzer.

Ein Beispiel, bei dem die Virtualisierung vollständig verwirklicht ist, findet sich bei jeder Steckdose und der Spruch „der Strom kommt aus der Steckdose“ kennzeichnet das Unwissen über die dahinter versteckte Komplexität.



Abb. 5a: Stromverbundnetz (Haupttrassen)

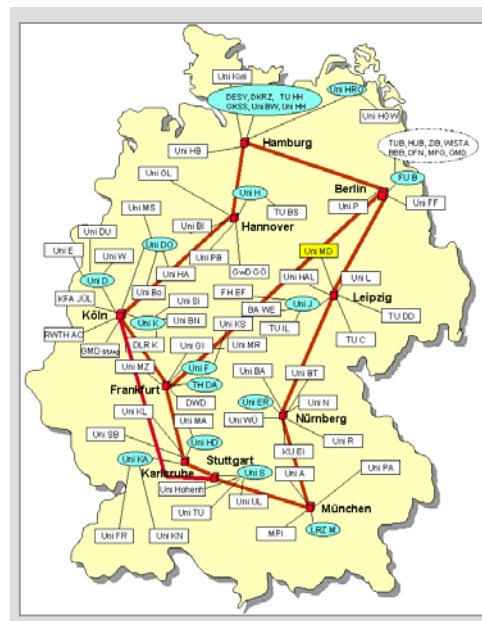


Abb. 5b: Breitbandnetz des Deutschen Forschungsnetzes

Bei dieser Virtualisierung spielt das Stromverbundnetz (Abb. 5a) eine entscheidende Rolle. Das Netz entstand in Deutschland um 1920 aus der genialen Idee, Steinkohle des Ruhrgebiets nicht mit Fuhrwerken an die Haushalte zu liefern, sondern in Form von Strom zu verteilen. Das heute europaweite Stromverbundnetz erzwingt einheitliche Standards bei der Einspeisung des Stroms durch die unterschiedlichsten Produzenten und Primärenergieformen.

Der Grad der Virtualisierung ist so weit entwickelt, dass kein Nutzer feststellen kann, woher sein Strom aus der Steckdose kommt. Selbst der ausdrückliche Vertrag mit einem Windmühlenbetreiber verhindert nicht, dass sich belgischer Atomstrom untermischt. Die Abrechnung des Lieferanten mit dem Kunden und die Bilanzierung zwischen den Lieferanten ist vom Strombezug abgekoppelt und erfordert neben dem Stromverbund ein weiteres Netzwerk von Stromhändlern.

Natürlich kann das Beispiel nur in der Tendenz zum Vergleich mit der globalen Virtualisierung in der IT herangezogen werden. Aus der Steckdose erhält man nur ein Produkt, aber mit hoher Qualität (Frequenzstabilität) und hoher Verfügbarkeit (s.u.). Über das Datennetz (Abb. 5b) werden oberhalb von TCP/IP unterschiedliche Kommunikationsprotokolle (z.B. smtp, http, ssh, scp) eingesetzt, die wiederum von unterschiedlichen Applikationen genutzt werden. Damit bleibt die sichtbare Komplexität für den Nutzer der „Datensteckdose“ sehr hoch. Auch die Werbung mit dem Tennisstar Boris Becker, „oh, ich bin ja schon drin“, kann darüber nicht hinwegtäuschen, denn jeder weiß ja, dass dann die Probleme erst anfangen.

Mit dem Grid-Computing [6,7] wird nun sowohl im akademischen Bereich als auch im kommerziellen Umfeld der Versuch unternommen, eine neue Virtualisierungsebene zwischen der Vielfalt der Basisprotokolle und den Applikationen einzuschieben. Dies kann nur dann dauerhaft erfolgreich werden, wenn die Schnittstelle einfach in der Nutzung ist (ähnlich wie beim Surfen auf der Basis von http) und weltweit akzeptiert wird. Daher müssen alle Beteiligten daran arbeiten, dass unabhängig von Einzelinteressen die Normierungs- und Standardisierungsbemühungen mit Nachdruck vorangetrieben werden.

5 Auswirkungen auf Anbieter und Nutzer

Das Hypertext Transfer Protocol (http) mit der Applikation WEB-Server und den in den Funktionen fast gleichartigen Clients, den Browsern, hat zu einer enormen Popularisierung des Internet geführt, so dass viele den Begriff Internet ausschließlich mit dem Surfen verbinden. Die überwiegende Nutzungsart ist dabei die Suche nach Informationen und dann die Präsentation der Information. Die Präsentation der Information ist auf den visuellen Anreiz ausgerichtet, d.h. die Botschaft soll beim Benutzer „ankommen“. Durch diese Verpackung ist die

Möglichkeit der Weiterverarbeitung der gewonnenen Information in eigenen Anwendungen sehr stark eingeschränkt. Grid-Computing könnte hier neben der Präsentation eine einheitliche Austausch- und Kombinationsmöglichkeit schaffen, mit der z.B. www.hrs.de und www.dwd.de von einem dritten Anbieter so gekoppelt werden können, dass man bei der Suche nach einem Urlaubshotel nur die Orte mit vorhergesagtem Sonnenschein erhält.

Es werden zukünftig viele nützliche aber auch weniger sinnvolle Anwendungen des Grid-Computing entwickelt werden deren Nutzung und deren Betrieb nicht ohne Risiko ist. Die Betreiber der neuentstandenen Dienste, die netzwerkweit auf vorhandene Funktionen anderer Betreiber aufbauen, werden dabei noch stärker abhängig vom technischen und wirtschaftlichen Erfolg der Partner, als es jetzt schon mit dem Abhängigkeitsverhältnis zu Lieferanten und Wartungsfirmen gegeben ist. Eine weitere Herausforderung wird die Fehlersuche und die Fehlerbehebung von Grid-Anwendungen mit sich bringen. Dabei muß dann die versteckte Komplexität unterhalb der Virtualisierungsebene verstanden werden. Beim großen Blackout im Staat New York in diesem Jahr hat es ca. eine Woche gedauert, bis die Ursache lokalisiert werden konnte.

Referenzen

- [1] M. Wertheim, *Die Himmelstür zum Cyberspace – Eine Geschichte des Raumes von Dante zum Internet*, Piper Verlag, München (2002)
- [2] J. Behr, T. Fröhlich, Ch. Knöpfle, W. Kresse, B. Lutz, D. Reiners, F. Schöffel, *The Digital Cathedral of Siena - Innovative Concepts for Interactive and Immersive Presentation of Cultural Heritage Sites*, in Bearman, David (Ed.) u.a., *Proceedings of the Cultural Heritage and Technologies in the Third Millennium*, Mailand, 57-71 (2001)
- [3] G. M. Amdahl, G. A. Blaauw, and F. P. Brooks, *Architecture of the IBM System/360*, IBM J. Research and Development, 8, 87–101 (1964)
- [4] M. Varian, *VM and the VM Community: Past, Present, and Future*, SHARE 89, (1997)
- [5] V. G. Cerf, Kahn, R. E., *A protocol for packet network interconnection*, IEEE Trans. Comm. Tech., 22, 627-641 (1974)
- [6] I. Foster and C. Kesselman, ed. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufman Publishers (1998)
- [7] D. W. Erwin and D. F. Snelling, *UNICORE: A Grid Computing Environment in Proceedings of Euro-Par 2001*, Springer LNCS 2150, 825-834 (2001)

On Demand Business und On Demand Computing: Eine technologische Perspektive

Walter Hehl

IBM Forschungslabor Zürich
Industry Solutions Lab
CH-8803 Rüschlikon
Switzerland
E-mail: whe@zurich.ibm.com

Die Vernetzung von Geräten, Menschen und Institutionen ist dabei, mit Macht in eine Phase einzutreten, die man als On-Demand-Ära bezeichnen kann. Das Referat zeigt auf, wie technische Perspektiven und wirtschaftliche Zwänge zusammenpassen und als Ergebnis in vielen Bereichen On-Demand-Systeme entstehen:

In der Produktion mag es die Digitale Fabrik sein, im Handel der integrierte Produzent/Verteiler und Händler, die Militärs nennen es das "Intelligent Battle Field", und so fort, ganz allgemein das Intelligente Echtzeit-Unternehmen.

1 Basistechnologien

Wir heben drei Grundtechnologien hervor:

- Prozessoren,
- Drahtlose Kommunikation,
- Software.

Wie bei den anderen großen Stufen in der Entwicklung der angewandten Informatik, ist die letztens treibende technische Kraft die Entwicklung der **Prozessorentechnologie**. Auf der Grundlage der Entwicklung der Chip-technologie – heute und noch für Jahre – in CMOS-Technologie, bestimmt das Moore'sche Gesetz den Haupttrend, und fasst die Entwicklung auf physikalischer Ebene, im architektonischem Bereich und der wirtschaftlichen Triebkräfte zusammen. Wie schon in den letzten Jahren, sind die Erwartungen der Ingenieure etwas konservativer als die der berufsmäßigen Futurologen mit einer Verdopplung alle 18 Monate, aber es ist durchaus möglich, dass die gegenwärtige Verdopplung der Prozessorenleistung pro Jahr weiter anhält.

Typisch für ein derartiges Wachstum ist der Selbstbeschleunigungs-Effekt durch die Technologie selbst: Je mehr Schaltkreise auf einem Chip realisiert werden können, um so mehr Prozessoren finden auf dem Chip Platz, um so größer können Caches und andere Hilfsstrukturen sein, um so mehr Kommunikation wandelt sich von mühsamer Off-Chip-Kommunikation zu schneller interner Kommunikation. Eine beobachtete Änderung im Trend und in der Risikobewertung ist die Abhängigkeit von erfolgreichen physikalischen Prozessänderungen durch neue, noch nicht bekannte Materialien. Es ist offensichtlich, dass neue Materialien – wie etwa Stoffe mit niedriger, aber auch mit hoher Dielektrizitätskonstante – in die komplexen Prozesse eingeführt werden müssen. In der Tat wird hierdurch das Risiko der Vorhersagen erhöht.

Auch die **drahtlose Kommunikation** profitiert von dieser Entwicklung: Protokolle wie das WLAN-Protokoll können zum Nahezu-Nulltarif in den Prozessor integriert werden. Der Stromverbrauch pro Transistor geht weiter drastisch zurück, und ganz normale CMOS-Chips werden die hohen Frequenzen selbst der WLAN-Verfahren der (nahen) Zukunft erreichen mit Datenraten von 56 bis 100 Mbits/sec. Benötigt man nur wenig Intelligenz im Chip – etwa nur die Rückgabe einer Identifikation –, so reichen einige Milligramm von Silizium aus, um diese Funktion zu liefern. Die vollständige mobile Kommunikation inklusive der Lösung des Problems des Wirrwarr der Standards wird durch das „software-definierte Radio“ ermöglicht: Die große logische Kapazität des Chips ermöglicht die Anpassung an die Ungereimtheiten und Inkompatibilitäten der jetzigen und der kommenden Standards. Analog und digital koexistiert auf dem Chip, und die Signale werden ab der Antenne digitalisiert und mit der passenden Software verarbeitet. Als Konsequenz haben wir Prozessoren überall, nicht nur „anywhere“ sondern „everywhere“ und begleitet von Sensoren überall, von einfachen und universellen RFIDs bis zu mit Computern gekoppelten Videokameras.

Softwaretechnologie ist eine weitere Grundtechnologie, in mancher Hinsicht sogar noch bedeutsamer, wenn auch unspektakulärer, als die der Hardware. In einigen Branchen, etwa im Automobilbereich und in der elektronischen Industrie, wird die integrierte und hochqualitative Softwareentwicklung zum unternehmenskritischen technischen Feld. Bis zu 90% aller Innovationen im Fahrzeug werden von und mit der IT getrieben, und der Umfang an Firmware steigt exponentiell an. Die menschlichen Aspekte (und Probleme) des Prozesses der Softwareentwicklung dagegen sind seit den ersten Großentwicklungen von Betriebssystemen vor nahezu 40 Jahren bekannt, aber nur wenig verbessert worden. Im Umfeld des On Demand Computing sind hier etwa die Verfahren der „extremen“ oder „agilen“ Softwareentwicklung zu nennen.

Für die IT-Welt insgesamt ist seit zwei Jahrzehnten die Erstellung und Verwendung verteilter und integrierter Systeme das dominante Thema. Nach der Generation der relativ „schweren“ verteilten Objekte im Sinne von CORBA haben praktisch alle Softwareunternehmen die Web Services als fundamentales „leichtes“ Softwarekonzept akzeptiert (z. Bsp. [1]). Web Services als virtuelle Komponenten erlauben den Zugriff auf Datenbanken, beschreiben physikalische Komponenten der IT-Infrastruktur oder geben den Zugriff auf Geschäftsanwendungen. Die technische Verschiebung von Objekten zu Diensten verspricht genau die organisatorische Dynamik in der Integration, die das On-Demand-Konzept benötigt [2].

2 On Demand Überall

Wechselt man von der technischen Sicht der IT-Trends zu der Sicht der Ökonomie und Betriebsführung, so erhält man diese Grundaufgaben für die Unternehmen:

- Konzentration auf Kernkompetenzen (mit der Konsequenz enger und loser Kooperationen),
- Schnelle, flexible und wirtschaftliche Reaktion auf Änderungen im wirtschaftlichen Umfeld,
- Stabilität und Zuverlässigkeit des Unternehmens als Ganzes.

Diese betrieblichen Anforderungen lassen sich auf technische Forderungen abbilden, die sich ungezwungen in den wesentlichen aktuellen Trends der IT wiederfinden, wie

- Offene Standards zur Beschreibung von Diensten, Ressourcen und Arbeitsabläufen,
- Virtualisierung von Ressourcen,
- Integration und leichtere Kombination von Prozessen im Unternehmen, von der Prozessbeschreibung über den Programmentwurf hin zur Implementierung, vom Datamining hin zur Unterstützung von operativen Entscheidungen,
- Reduktion der Komplexität wo immer möglich – auf Netzwerk, System oder Subsystem und Komponentenebene – hin zur Autonomie [3,4]

Die Quellen der Technologien für diese Fülle technischer Arbeit sind weit gestreut und reichen von der Community der Hochenergiephysik mit Grid und OGSA, den Erfahrungen mit mehreren Jahrzehnten Mainframe-Systemdesigns, hin zu Softwareunternehmen wie Rational/IBM mit der verbreiteten „Entwurfssprache“ UML. Im Hause IBM stehen bei nahezu allen Entwicklern und Beratern Aufgaben zum On Demand im Pflichtenheft und deren Integration in das On-Demand-Gesamtbild.

Die enge Verknüpfung von Technologie und ökonomischen Prinzipien im On-Demand-Bereich erzeugt neue Aufgabenstellungen für die angewandte Informatik, etwa die Beschreibung von IT-Strukturen als Märkte mit den Methoden der Finanzmathematik und die Entwicklung von Werkzeugen zur Analyse geplanter Prozessänderungen [5]. In dieser Refokussierung der angewandten Forschung und Entwicklung durch On Demand spiegelt sich auch unmittelbar die wachsende Bedeutung von anspruchsvollen Dienstleistungen im IT-Bereich. Diese anwendungs- und prozessnahen Technologien werden für Dienstleister zur eigenen Kernkompetenz.

Zwei Technologiebeispiele aus der Forschung zu On Demand, ein systemtechnisches und ein „business-technisches“ Projekt:

1. Das autonome Speichersystem „Collective Intelligent Bricks“ [6]:

Dieser hochkompakte RAID-Systemprototyp besteht aus einer Menge von Würfeln, die selbst Sätze von Plattensystemen enthalten und drahtlos mit ihren Nachbarn kommunizieren. Daten von defekten Platten werden transparent umgelagert – kein Wartungstechniker sollte das System je demontieren müssen.

2. Das Datenanalyse-System „WebFountain“ für heterogene Daten [7]:

WebFountain liefert eine Plattform für die Textanalyse auch von sehr großen Datenbeständen (semistrukturiert oder unstrukturiert), zur Zeit auf etwa 1.3 Milliarden gespeicherte Seiten. Anwendungen sind etwa Marktanalysen, das Risikomanagement eines Produkts oder das Auftreten neuer Produkte.

3 Ausblick

Das On-Demand-Konzept hat nicht nur die Unterstützung der IBM, sondern auch sinngemäß die anderer großer Hersteller von Systemen und von Beratungsfirmen, etwa als „Real Time Enterprise“ (Gartner Group) oder „Organic IT“ (Forrester Research).

Viele der Entwicklungen von „On Demand“ sind nicht neu und die einzelnen Funktionen zeichneten sich schon früh ab, etwa das Partitioning und Load Balancing bei den klassischen Mainframes und partielle Optimierungen bei verteilten Systemen. Nun kommen die benötigten Hardwarekomponenten (etwa RFIDs oder intelligente Speichersysteme), Softwarestandards (wie Web Services) und Business Trends (wie der Zwang zur Kapitalminimalisierung) synergetisch zusammen und erweitern die Systemgrenzen und die Dynamik auf das Gesamtunternehmen, auf die Partner und Kunden, ja auf die gesamte Wirtschaft.

Referenzen

Anbetrachts des umfassenden Charakters der Thematik, kann hier nur eine sehr beschränkte Literatúrauswahl zum Referat gegeben werden:

- [1] M. Champion, Ch. Ferris, E. Newcomer und D. Orchard, *Web Services Architecture*, <http://www.w3.org/TR/ws-arch/>
- [2] IBM, *Living in an on demand world*, IBM Corporation (2002), <http://www.ibm.com/e-business/doc/content/feature/offers/whitepaper.pdf>
- [3] P. Horn, *Autonomic Computing: IBM's Perspective on the State of Information Technology*, IBM Corporation (2001), auch unter http://www.research.ibm.com/autonomic/manifesto/autonomic_computing.pdf
- [4] A. G. Ganek, T. A. Corbi, *The dawning of the autonomic computing era*, IBM Systems Journal, Vol. 42, No 1, 2003
- [5] C. Kenyon, G. Cheliotis, *Grid Resource Commercialization: Economic Engineering for Delivery Scenarios*, in *Grid Resource Management*, hrsg. von J. Nabrzyski, J. Schopf, J. Weglarz, Kluwer, (2003)
- [6] IBM, *Collective Intelligent Bricks*, IBM Corporation (2003), http://www.almaden.ibm.com/StorageSystems/autonomic_storage/CIB/index.shtml
- [7] IBM, *WebFountain*, IBM Corporation (2003), <http://www.almaden.ibm.com/WebFountain/publications/pdf>

Web Services - Das Netz ist der Rechner

Karl-Heinz Bondick

T-Systems CDS GmbH
web services
Lübecker Straße 2
39124 Magdeburg
Germany
E-mail: Karl-Heinz.Bondick@t-systems.com

Web Services im Sinne von Web und Services als innovative, elementare und integrale Basis für erfolgreiche virtuelle Strukturen – ein Beitrag für den Kunden.

1 Das Netz ist der Rechner

1.1 Vorwort: Das Internet ändert das Konzept der Applikationsnutzung

Das Internet wird langsam aber sicher zu einem alltäglichen und selbstverständlichen Medium.

Auch in Zukunft wird das Internet weitestgehend so aussehen wie heute - allerdings wird es um einiges intelligenter sein.

Nicht nur im Privatbereich, auch im Business-Segment hält das Internet Einzug und hat Einfluss auf Abläufe und Organisationen.

Beispiele für Individualkunden

- Der leere Kühlschrank, der eine e-Mail mit einer Bestellung für frische Milch an das Lebensmittelgeschäft schickt. Die Milch wird vor dem nächsten Morgen vor die Tür geliefert.
- Ihr Auto stellt eine Anfrage an alle Tankstellen innerhalb von zehn Kilometer und stellt so fest, welche die günstigsten Preise bietet. Das Navigationssystem des Autos wird Sie daraufhin zu dieser Tankstelle navigieren.
- Mittels des Funktelefons ist es Ihnen möglich, ein Getränk an einem Automaten statt mit Münzen zu bezahlen.

Beispiele für Unternehmen

- Unternehmen gehen ins Netz: Mitarbeiter aus Casablanca, Berlin, New York und Tokio arbeiten in einem Team.
- Mobiler Vertrieb: Vertriebsmitarbeiter erstellen und versenden Angebote und Verträge unterwegs.
- Kunden bestellen online, prüfen und verfolgen den Auftragsstand und zahlen via Web.
- Der Lagerbestand eines Unternehmens wird elektronisch geführt. Bei Erreichen des Meldebestandes wird eine Bestellung an den jeweils preiswertesten Lieferanten automatisch ausgelöst. Die Waren / das Material werden am nächsten Tag versandt.
- Rechnungen, Zahlungserinnerungen werden automatisch überwacht und per e-Mail an die Kunden versandt.

1.2 Das Netz ist der Rechner

Die Globalisierung des Wettbewerbs nimmt immer mehr zu. Märkte werden nicht mehr durch nationale Grenzen definiert. Vielmehr zählt die Fähigkeit, Zielgruppen global anzusprechen und zufriedenzustellen.

Jeder ist davon betroffen: kleine und mittlere bis hin zum großen Unternehmen.

Dazu bedarf es globaler Dienstleistungen, die dabei helfen, Grenzen zu überwinden und unabhängig von Ort, Zeit und Sprache dem Kunden zur Verfügung stehen.

Vor allem im Business-Bereich wird der Bedarf an globalen Services enorm zunehmen.

Dies zeigt sich verstärkt im Bereich der Kommunikation, wo bereits heute viele der internationalen Unternehmen ihre Standorte zusammengeschaltet haben, so dass alle Mitarbeiter unabhängig vom jeweiligen Standort erreichbar sind.

Flexibilität, Geschwindigkeit, Internationalität bei hoher Qualität sind die daraus folgenden Anforderungen.

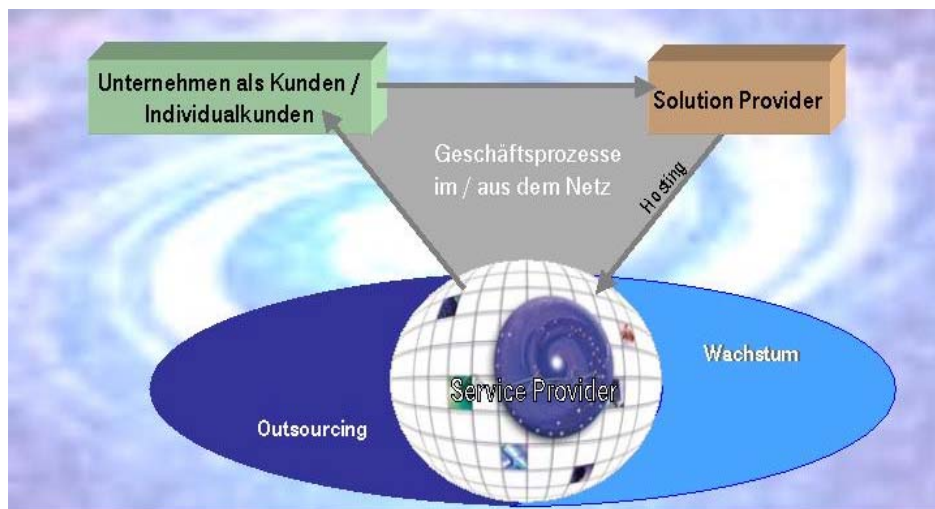


Abb. 1: Das Netz ist der Rechner. Outsourcing und Wachstum.

1.3 Wandel des Outsourcings

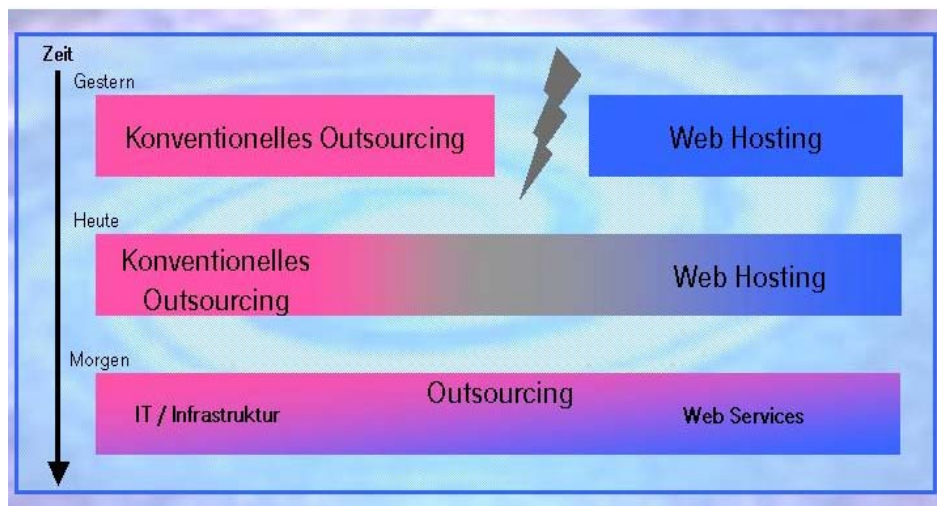


Abb. 2: Outsourcing Gestern – Heute – Morgen

These 1:

Das Web-Geschäft ist anders als die konventionelle IT:

- viele Beteiligte
- für Unternehmensfremde offen
- Verfügbarkeit weltweit, rund um die Uhr
- Anforderungen an Schnelligkeit

These 2:

Die interne IT und unternehmerischen Organisationsstrukturen sind nicht passend aufgestellt

- bestehende Systemlandschaft
- vorhandene Mitarbeiterverteilung und -kenntnisse
- eingeführte Prozesse, Prinzipien und Werkzeuge
- Netzseite ist oftmals ungenügend abgedeckt.

Ergo: Der Outsourcing-Gedanke erhält im Web-Geschäft eine neue Bedeutung.

Auch das Outsourcing entwickelt sich zunehmend zum globalen Dienstleistungsangebot. Outsourcing-Kunden möchten einheitlichen Service aus einer Hand, unabhängig vom jeweiligen Land, wo die Leistung zu erbringen ist.

In Zukunft werden Web Services zweifellos neue Geschäftsmodelle erzeugen, die festgefügt IT-Strukturen ablösen und lose gekoppelte Collaborative-Business-Netzwerke ermöglichen.

Gegenwärtig findet die Diskussion über Web Services hauptsächlich auf einer technologischen Ebene statt. Fragen nach der Plattform, den verwendeten Standards, Architekturen, Sicherheitsfragen und Service Level Agreements stehen im Vordergrund.

Web Services erleichtern es Unternehmen, Fremdanbieter zu finden, die spezifische Geschäftsprozesse kosteneffizient übernehmen können. Bisher wurden über das Internet nur Informationen ausgetauscht, zukünftig können selbst komplexe Geschäftsprozesse über das Netz aufgeteilt und von verschiedenen Partnerunternehmen gemeinsam abgearbeitet werden.

1.4 Kundenbindung durch systematische standardisierte Wertschöpfungsstrategien



Abb. 3: Wachstumsmotor

Die Stabilisierung und Erweiterung des Kundenbindungsprozesses wird durch den Wachstumsmotor (perpetuum mobile) erreicht.

Basis und Mittelpunkt des Wachstumsmotors ist der Kunde mit seinen Bedürfnissen sowie seinen Prozessen und Applikationen. Um wettbewerbsfähig zu sein und zu bleiben, muss ein Unternehmen (der Kunde) diese in das Netz bringen.

Der Treibriemen (die Energie des Wachstumsmotors) wird aus dem Know-how der Mitarbeiter in Zusammenarbeit mit den Partnern gewonnen, die zielgerichtet und permanent durch Nutzung angewandter Innovationen Standardisierungseffekte erzielen. Diese Effekte bewirken die Verbesserung der Wertschöpfung und veranlassen den Kunden, neue Aufgaben zu beauftragen.

Die proaktive Beschäftigung mit dem Kunden und seiner Business Prozesse unter Nutzung der Applikationen schafft ein positives Umfeld und Beständigkeit in der Zusammenarbeit.

Web Services (Web und Services) sind daher mit den drei strategischen Aussagen („Das Netz ist der Rechner“ + Die neue Welt des Outsourcings + Der Wachstumsmotor) dominanter Bestandteil der zukünftigen virtuellen IT-Dienstleistungen und ein neuer Wertschöpfungsansatz auch für den Kunden.

web services - Unsere Referenzen.



.....T.....Systems.....

Das Netz ist der Rechner.
Karl-Heinz Bondick

Referenzen

- [1] Thomas Mörsdorf, *Neue Anforderungen an Informations- und Kommunikationsdienstleistungen*
<http://www.innovation-aktuell.de/kv0705.htm>, 05.09.03
- [2] Web Services in Europa; SAP SI; 2003

Unternehmensportrait

T-Systems – ein starker Partner

T-Systems ist einer der führenden Dienstleister für Informations- und Kommunikationstechnologie (engl.: ICT) in Europa. Im Konzern Deutsche Telekom betreut das Unternehmen das Segment der größten Geschäftskunden. T-Systems gehört zum Kerngeschäft der Deutschen Telekom und trägt zu den Ergebnissen des gesamten Konzerns bei.

In über 20 Ländern beschäftigt T-Systems rund 43.500 Mitarbeiter. Die Telekom-Tochter erwirtschaftete 2002 einen Umsatz von rund 11,3 Milliarden Euro und ein bereinigtes EBITDA (Ergebnis vor Zinsen, Steuern, Abschreibungen und Amortisationen) von 1,15 Milliarden Euro. Das entspricht einer Steigerung gegenüber dem Vorjahr um rund 30 Prozent.

Wertschöpfung steigern mit IT und Telekommunikation

Die Lösungen von T-Systems zielen darauf, das Geschäft von Großkunden aus den Branchen Telecommunications, Services & Finance, Public & Healthcare sowie Manufacturing nachhaltig zu verbessern. Das Unternehmen optimiert für seine Kunden die Prozesse, senkt die Kosten und verbessert ihre Ergebnissituation durch den Einsatz modernster Technologien. Die Fachleute von T-Systems verfügen über spezifisches Branchen- und Prozess-Know-how.

Das Spektrum der Services reicht von der Integration neuer ICT-Lösungen in bestehende Kundensysteme über den Aufbau und Betrieb von Arbeitsplatzsystemen, Rechenzentren und Netzwerken bis hin zum Telekommunikationsgeschäft mit internationalen Carriern.

Value Computing – Am Beginn einer neuen Ära

Franz N. Busch

T-Systems CDS GmbH
Pallaswiesenstr. 182
64293 Darmstadt
E-mail: Franz.Busch@t-systems.com

Die sich wandelnden Anforderungen aus der Globalisierung der Wirtschaft werden auch zu einem drastischen Wandel in der Informationstechnologie führen. Weg von singulären, monolithischen Anwendungen, denen dezidiert Ressourcen zugeordnet sind, hin zu einer Virtualisierung, die alle Bereiche der IT-Infrastruktur umfassen wird. Ressourcen werden als Pool bereitstehen und bedarfsgerecht genutzt, dies wird zu flexiblen und verbrauchsgerechten Abrechnungsmodellen führen. Erste Schritte sind u. a. die Adaptation und Weiterentwicklung der bisher nur im technisch-wissenschaftlichen Umfeld eingesetzten Grid-Technologie in Richtung Einsatz im kommerziellen Umfeld, dem „Commercial-Grid“. Der Weg führt dabei zum virtuellen Rechenzentrum, das weitgehend autonom und flexibel auf Laständerungen und Störungen reagiert und so den gewandelten Anforderungen an die Informationstechnologie Rechnung trägt.

1 Wandel in der Informationstechnologie

Die Art und Weise, wie heute noch IT-Services entwickelt, genutzt und die zugehörigen Infrastrukturen betrieben werden, steht mit hoher Wahrscheinlichkeit vor durchgreifenden Veränderungen. Viele kleinere Innovationen, jede für sich genommen nur von geringer Bedeutung türmen sich, angetrieben und zusammengeführt durch aktuelle Trends, zu einer Innovationswelle auf.

IT-Services werden zunehmend als Bestandteil einer Grundversorgung verstanden, ähnlich wie Strom und Wasser. Eine jederzeit gesicherte Verfügbarkeit, hohe Qualität, Sicherheit und niedrige Kosten werden als selbstverständlich angenommen und die Frage nach dem Lieferanten oft nur

über den Preis entschieden. Gleichzeitig fragt der Markt aber immer dringlicher nach flexibel gestaltbaren, jederzeit kurzfristig abzurufenden und nach Verbrauch zu bezahlenden IT-Leistungen. Hinter diesen Forderungen steht die hohe Dynamik der Märkte, die Unternehmen zu größerer Flexibilität und höherer Reaktionsgeschwindigkeit zwingt, um im globalen Wettbewerb zu bestehen. Neue Geschäftsmodelle erfordern eine angepasste Informationstechnologie und genau darauf abgestimmte Prozesse. Von der IT als Geschäftsprozesse unterstützendes oder gar tragendes Hilfsmittel wird mindestens die gleiche Flexibilität und Reaktionsgeschwindigkeit verlangt. IT-Services, die diesen Ansprüchen gerecht werden, tragen dann unmittelbar zu einer gesteigerten Wertschöpfung für ihre Nutzer bei – wir nennen sie daher „Value Computing“.

Der heutigen IT haftet, oft zu Recht, das Stigma der Rigidität, mangelnder Reaktionsgeschwindigkeit sowie zu hoher Kosten an. Dies resultiert aus einer zu hohen Komplexität, die wiederum auf die Vielzahl individueller Lösungen, dedizierter IT-Ressourcen und Skills zurückzuführen ist. Ein weiterer Nachteil zulasten der erwünschten Flexibilität ist die enge vertikale Verzahnung: von den Besonderheiten der Hardware, über Betriebssysteme und Middleware bis hin zur individuell gestalteten Applikation.

Die „klassische“ Vorgehensweise, permanent die Komplexität zu reduzieren um sie beherrschbar zu halten, wird alleine nicht den Durchbruch in Richtung „Value Computing“ bringen. Vielmehr sind neue, an den Grundlagen ansetzende Wege erforderlich.

2 Value Computing

Die „Value Computing“ voraussetzenden Technologien sind zum Teil bereits vorhanden oder reifen gerade heran. Die vielleicht wichtigsten sind Grid-Computing, XML und Web-Services. Zudem sind zwei sich immer deutlicher abzeichnende Trends relevant, zum einen eine Virtualisierung, d. h. die Trennung der ITC-Komponenten, um so eine größere Re-Kombinierbarkeit zu erreichen, und zum anderen eine Dynamisierung, also die Nutzung der Mechanismen des Grid-Computings in der Legacy Welt.

Tatsächlich haben bei dem von uns erarbeiteten Modell des „Value Computing“ die vielfältigen Grid-Aktivitäten, die bislang ausschließlich im technisch-wissenschaftlichen Umfeld mit großem Erfolg im Einsatz sind, Pate gestanden. Wir reden daher im Zusammenhang mit „Value Computing“ auch von „Commercial Grids“.

Grid-Computing ist aus der Not der Wissenschaft nach höchsten Rechenleistungen bei geringen finanziellen Mitteln entstanden. Die Anwendungsbeispiele reichen von SETI@Home über die Experimentanalyse beim CERN bis hin zu Crash-Simulationen im Automobilbau. Schon einmal ist eine die Gesellschaft tiefgreifend verändernde Technologie aus wissenschaftlichen Aktivitäten entstanden: das World-Wide-Web. Wahrscheinlich geschieht nun etwas ähnliches mit Grid-Computing.

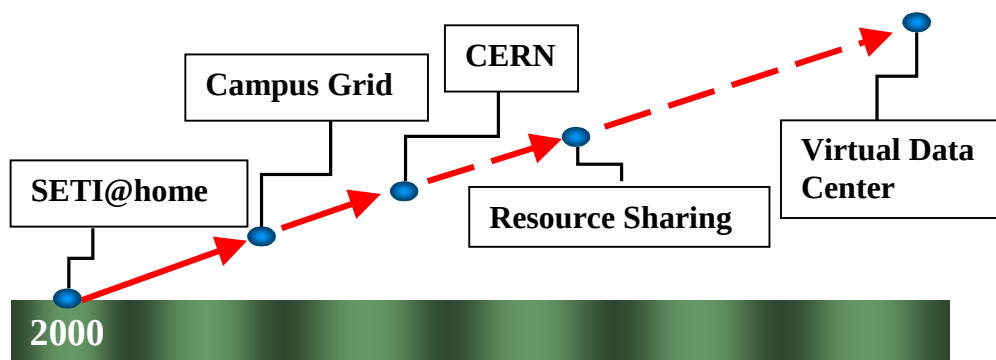


Abb. 1: Road Map Grid Computing / Value Computing.

„Value Computing“ ist ein komplexes Thema mit vielen Facetten, wie etwa der Software-Entwicklung, aber auch der zugrundeliegenden IT-Infrastruktur. Die IT-Infrastruktur von morgen wird virtualisiert sein. Virtualisierungswellen hat es in der IT schon viele gegeben und alle haben im Endeffekt zur Erhöhung von Flexibilität, Qualität und Leistung bei sinkenden Kosten beigetragen: virtueller Speicher, virtuelle Maschinen, virtuelle private Netze (VPN), virtueller Storage (SAN/NAS). „Value Computing“ und „Commercial Grids“ nutzen die nächste Welle, nämlich die Virtualisierung ganzer Rechenzentren, ja geographisch verteilter Verbünde von Rechenzentren.

3 Virtualisierung von Rechenzentren

Die Grundidee eines virtualisierten Rechenzentrums ist einfach: Das Rechenzentrum ist der Rechner. Die Prozessoreinheit und der Arbeitsspeicher dieses „Rechners“, nennen wir sie „Compute“-Einheit, besteht aus einem Pool standardisierter Server. Der Datenspeicher besteht aus einem Pool von in sich bereits virtualisierter SAN/NAS und Tape-Libraries. Das Rechenzentrums-interne Hochleistungskommunikationsnetz ist der System-Bus, die Kommunikation nach außen (und innen) wird über einen Pool von Firewalls, Switches und Sicherheitseinrichtungen dargestellt, den wir als

„Communication“-Einheit ansprechen. Das Ganze wird verwaltet durch eine Software-Komponente, dem Ressourcen-Manager, der im weitesten Sinne als das Betriebssystem des virtuellen Rechners (virtualisiertes Rechenzentrum) verstanden werden kann. Er bindet u. a. anforderungsgerecht und dynamisch die notwendigen Ressourcen zusammen, um die anstehenden Arbeiten zu bewältigen. Wegen der speziellen legalen und kontext- bzw. inhaltsbezogenen Anforderungen an die Archivierung von Informationen wird außerhalb der normalen Datenspeicherung zusätzlich eine eigene Archiv-Infrastruktur ausgeprägt. Das Ganze ist eine verallgemeinerte Ausprägung der Grid-Philosophie für den kommerziellen Einsatz, wobei viele Technologieelemente aus den technisch-wissenschaftlichen Grids Verwendung finden bzw. sich weiterentwickeln lassen.

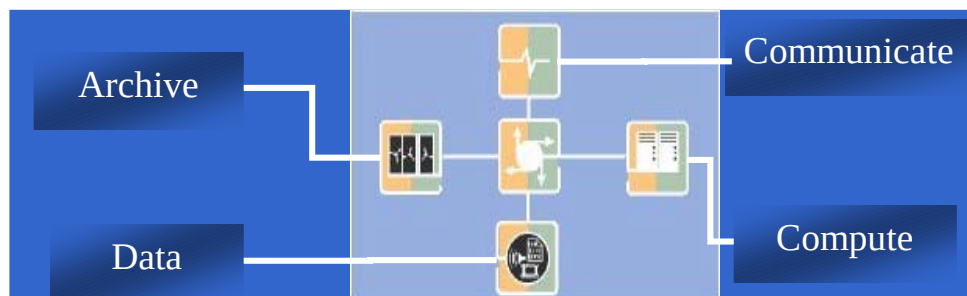


Abb. 2: Modellvorstellung eines virtualisierten Rechenzentrums

Die Vorteile eines virtualisierten Rechenzentrums sind immens. Die dynamische und flexible Nutzung von Ressourcen aus Pools führt zu einer besseren Auslastung der Infrastruktur. Durch Nivellierungseffekte im Ressourcen-Pool ist eine bessere Abfederung von Peak-Anforderungen im Gegensatz zu heute möglich, noch werden die meisten Systeme nach der Peak-Last und damit in der Regel viel zu groß dimensioniert. Die eingebaute Fehlertoleranz ermöglicht bei Störung einer Komponente, dass aus dem Pool automatisch und sehr schnell die nächste Komponente aktiviert und eingebunden wird, ohne dass Daten verloren gehen oder Transaktionen abgebrochen werden. Die starre Kopplung der Applikationen an physische Hardware wird aufgehoben, dadurch werden teure Insellösungen mit Detailausprägungen vermieden. Das Ressourcenmanagement vereinfacht sich dadurch dramatisch. Ein hoher Grad an Autonomie und Automation führt zu durch Service Level Agreements getriebenen automatischen Konfigurationen, Monitoring und Reaktion auf Störungen. So wird ein pro-aktives, gesamtheitliches Management der Ressourcen möglich. Geschäfts- und Abrechnungsmodelle werden sich anpassen, so können die bereitgestellten Leistungen nutzungsbezogen abgerechnet werden. Es ist der Wandel von der „Just-in-case“-Bereitstellung von IT-Services zu „Just-in-time“.

Ein virtuelles Rechenzentrum kann sich zudem aus mehreren geographisch verteilten virtualisierten Rechenzentren als Verbund darstellen, dies eröffnet zusätzliche neue Möglichkeiten, etwa die des Last- oder Ressourcenausgleichs. IT-Services können so noch flexibler auf geographischer Skala bereitgestellt werden.

Dass dies alles eine sehr ernsthafte Zielsetzung ist, zeigen die Aktivitäten und Programme einiger namhafter Hersteller. Das N1-Konzept von Sun, das Utility-Datacenter von HP und die „on demand“-Initiative von IBM beruhen auf sehr ähnlichen Zielsetzungen und Vorstellungen. Aber auch auf Seiten der Software-Hersteller ist Bewegung erkennbar, beispielsweise was Werkzeuge zur automatischen Übersetzung von Service Level Agreements in physische Konfiguration der benötigten Ressourcen betrifft. Hersteller kommerzieller (Standard-) Softwarepakete haben Programme aufgesetzt, ihre Produkte „Grid-fähig“ zu machen.

Auch auf Ebene der System-Integration, d. h. dem Bau von Applikations-Software greift die Erkenntnis Raum, dass monolithische kundenindividuelle Lösungen, die dazu auch noch tief mit der verwendeten System-Infrastruktur verknüpft sind, nicht Kosteneffizienz, Flexibilität und „time to value“ fördern. Stattdessen rücken Plattform-unabhängige Entwicklungsmethoden, „Model driven Architectures“ und wiederverwendbare Software-Bausteine verstärkt in den Mittelpunkt des Interesses. Ein weiterer Beleg ist die zunehmende Bedeutung von Java bei der Entwicklung von Applikationen. Dieser Trend ergänzt hervorragend die Virtualisierungsstrategie auf Infrastrukturebene.

Virtualisierung geht immer mit Bildung neuer Abstraktionsebenen einher, auf deren Basis sich wiederum neue Wertschöpfungsstufen realisieren lassen. Es ist offensichtlich, dass die heute eingesetzten Applikationen – von technisch-wissenschaftlichen abgesehen – auf einer solchen virtualisierten Infrastruktur erst sehr begrenzt einsatzfähig sind.

4 Eigene Arbeiten

T-Systems hat in einer Innovationspartnerschaft gemeinsam mit IBM einen Prototypen eines „Commercial Grids“ entwickelt und vorgestellt. Er demonstriert beispielhaft welche neue Qualität von Services prinzipiell realisierbar ist, etwa die sehr flexible und rasche Reaktion (im Bereich einiger Sekunden) auf Laständerungen aber auch auf Störungen. Wir werden zusammen mit unserem Partner das Thema weiter vorantreiben.

5 Erforderlicher Wandel

„Value Computing“ liegt im Trend und die demonstrierten Möglichkeiten sind vielversprechend. Dennoch setzt „Value Computing“ einen erheblichen Prozess der Veränderung bei Erstellung, Nutzung und Betrieb von IT-Services voraus. So wird die logische Abkopplung der Applikationsebene von der physischen Infrastruktur eine Umorientierung erfordern. Die Lösungsentwickler werden sich immer mehr daran gewöhnen, Plattform-unabhängig zu entwickeln und wiederverwendbare Bausteine zu integrieren. Die Nutzer werden sich daran gewöhnen, dass sie „ihre“ Server nicht mehr physisch identifizieren können, wohl aber Menge und Qualität der genutzten Ressourcen und der zugesicherten Service-Level. Der Systemadministrator wird sich immer weniger um die Parametrierung spezifischer Systeme kümmern müssen, sondern wird zum Ressourcen-Logistiker avancieren.

Nicht zu verschweigen ist, dass noch sehr viele technische Lücken zu schließen sind, bevor in größerem Umfange „Commercial Grids“ Wirklichkeit werden können. Stellvertretend seien zwei wichtige Themenkomplexe genannt. Commercial Grids werden nur dann erfolgreich sein können, wenn Security und Privacy in einer dynamisch flexibel genutzten Infrastruktur in vollem Umfang und unzweifelhaft gewährleistet werden können. Zudem ist bei der Steuerung der Ressourcen noch erheblicher Aufwand erforderlich. Eine automatische Konfiguration, die sensitiv auf Service-Level-Vereinbarungen, Kosten- und Leistungswünsche reagiert, ist heute erst in Teilen möglich. An all diesen kritischen Fragestellungen wird aber mit hohem Engagement gearbeitet.

6 Ausblick

Die Ära des „Value Computing“ steht erst an ihrem Anfang. Am Ende werden aber die neuen Möglichkeiten, allen voran die enorme Flexibilität und Reaktionsfähigkeit überzeugen. Wir rechnen fest damit, dass „Value Computing sich durchsetzen wird“. In etwa einem Jahr werden wir die ersten kommerziellen Grid-Applikationen sehen, in ungefähr drei, spätestens in fünf Jahren werden signifikante Anteile der IT auf „Commercial Grids“ basieren. T-Systems wird „Value Computing“ aktiv vorantreiben, getreu unserem Motto „we enable business flexibility“.

Grids und Grid-Services

Joan Massó
Wolfgang Rühmer

52425 Palma de Mallorca, Spanien
82166 Gräfelfing, Deutschland

E-mail: {jmasso, wruehmer} @ gridsystems.com

Nach einer kurzen Vorstellung des Unternehmens GridSystems werden die spezifischen Produkteigenschaften von Inner-/Outergrid dargestellt. Zum Abschluss wird auf die verschiedenen Standards zusammen mit den Herausforderungen der Zukunft eingegangen.

1 What defines a Grid ?

Ian Foster: „flexible, secure, coordinated resource-sharing among dynamic collections of individuals, institutions, and resources”

IBM: “Grids enable the virtualization of distributed computing and data resources”

Deutsche Bank (New Economy Report): “the GRID...to supercede the Web, and form the bedrock of the new economy”

Oracle: “the most flexible and cost-effective way to manage enterprise information and business applications.”

2 The Company Gridsystems

In February of the year 2000 the company GridSystems was founded by Joan Masso and established in a Campus area in the North-West of Palma de Mallorca. After 3 years of concentrated development there exists a large customer base in

Spain and the company is now expanding to new territories like Germany, France, Italy, UK, Japan and USA.



Abb. 1 The founder Joan Masso at the Campus

As of today Gridsystems employs 50 people and has a yearly revenue of about 5 MEuros.



Abb.2: A look into developer's office

2 The Product Innergrid

Introduction

InnerGrid is the software platform of GridSystems that enables your organization to easily take advantage of all your resources (data, applications, computers, and people) by creating your own grid. InnerGrid simplifies the cost and complexity of your IT infrastructure and the integration of your solutions. InnerGrid shatters technological constraints and gives you the time and freedom to concentrate on what really matters: your business.

InnerGrid allows you to distribute complex tasks among all the computers available in a Grid in order to speed up processes and to make possible new ways of working. In the following paragraphs we present an example to illustrate the InnerGrid capabilities.

All classic movies are being converted to digital format. Why? Because soon they will be available in DVD, but also... in your cellular phone! Once a film is in a digital format, it can be reconverted to adapt it to the final target. For instance, DVD requires high definition, but cellular phones require low resolution and very high compression. Why is this related to InnerGrid? Conversion of video formats is a very computational intensive task. In a single computer it would take a large amount of time to perform the conversion. InnerGrid splits the conversion task in little pieces (fragments of the whole film) and distributes them among the computers in the Grid. Each computer converts its fragment of the film. InnerGrid takes care of gathering all the converted fragments and joins them to obtain the film in the target format.

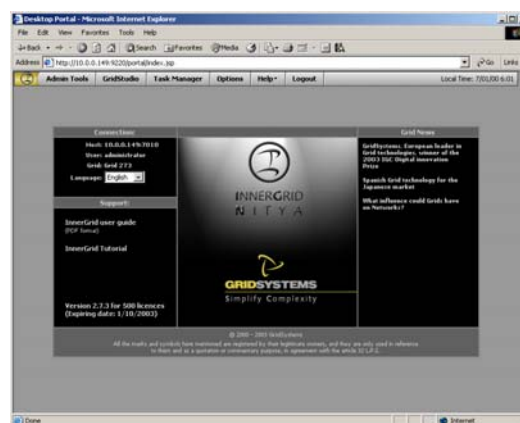


Abb. 3: Screenshot taken from InnerGrid Nitya

Technology in a nutshell

The InnerGrid components are the Servers and the Agents. In each InnerGrid installation there is at least one Server and several Agents. The Server is a key element in the architecture of InnerGrid. It manages the distribution of computing tasks towards node clients. Two or more Servers can work on different machines when fault tolerance is critical. The InnerGrid Agent is the piece of software that enables the nodes of the Grid to collaborate in the execution of tasks. The software runs continuously on each computer integrated in the platform and executes the set of actions required for its participation in the processes. The Agent is a very lightweight software, so its execution does

not affect the normal performance of the computer. It benefits from unused computational resources only.

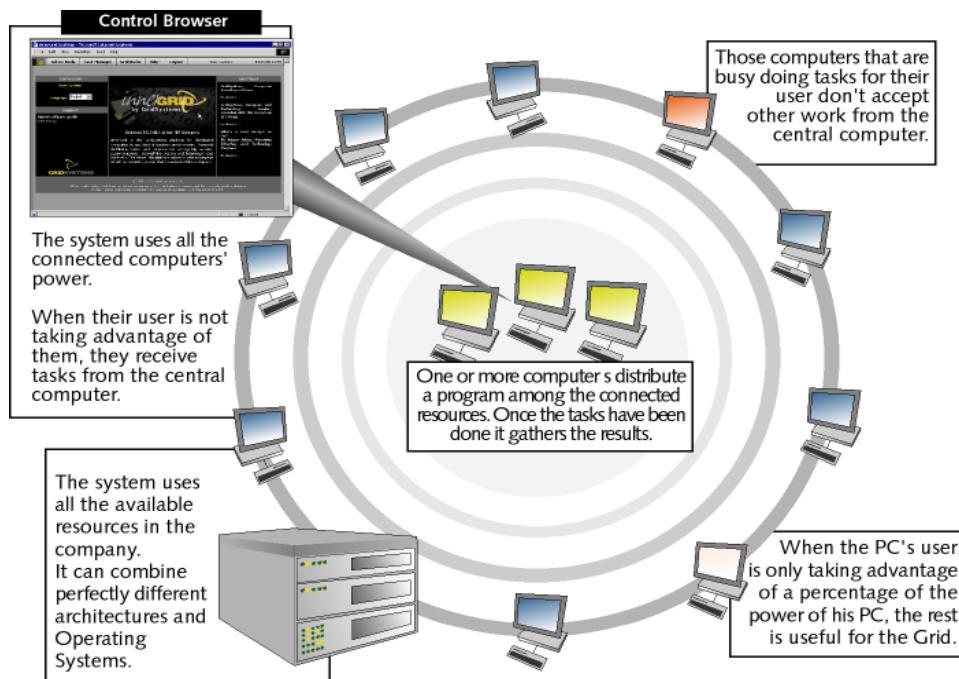


Abb. 4: Concept of how InnerGrid works

How does it work?

To show it let us explain the video conversion example giving the technical detail.

In a first step the original film is split in small fragments. The InnerGrid Server stores in its file system each one of these fragments, together with the binaries of the program used to make the conversion. When the user launches the video conversion task, the InnerGrid Server gives out a fragment of the film to each Agent that has computational resources available and fulfils the requirements needed to execute the conversion program. Each Agent saves "its" fragments of the film and the conversion program in its file system and runs the program to convert the fragments. Once an Agent has finished, the original fragment is deleted and its converted counterpart (the result file) is moved to the file system of the Server, which saves all the converted fragments.

Once the whole video conversion task is finished, the conversion program is deleted in all the Agents and the Server joins all the pieces to get the film converted to the desired format.

3 Key Properties of InnerGrid Nitya

InnerGrid is a Grid, but also more than that. It possesses other characteristics that make it an attractive choice among other possibilities. The aim of its construction has not only been distribution or computing, but also administration easiness, robustness, and reliability, so that the user can avoid the complexities of the problem and concentrate on his business. InnerGrid means to be an instrument to help you to do your job, not yet another costly (both in work and money) computing tool.

Easy

One of the characteristics of InnerGrid is that both the installation and configuration are very easy. In a few hours, heterogeneous and spread resources become under the control of the Server and provide you with new capacities.

Robust, Resilient and Reliable

InnerGrid is a robust product. It does not have a single point of failure. The business processes are carried out by many machines working together and when one fails to finish its part, this is assigned to another without any loss of efficiency. On the Server side, the possibility to have redundant Servers guarantees that even in the case of failure, another machine will immediately assume the control and continue the task. It is a resilient product. Not only it is unaffected by hardware or software failures of some of the components of the Grid, but you can change its composition while it works and it will adapt to the new situation. If you add a new computer, it will soon receive a task to do. If you remove it, anything it might be doing will be assigned to another computer.

It is a reliable product. When you run a task, your data and its processing are safely controlled at any moment. This ensures that the correct results reliably come to you.

Scalable

InnerGrid provides you with a computing resource that can grow to virtually no limits. Plugging another machine automatically adds more computing power to

your system. Your previous investments in hardware will not become obsolete when you want to increase your capacity in throughput. The scalability can be thought two-fold. From one point of view, increasing the number of machines in the Grid decreases the time spent in a job, as long as you can properly divide it in sufficiently small parts to distribute it among all the resources. From a second point of view, one can enlarge the task with little time cost as long as there are free resources that can accept the growing load.

Multiplatform

In a large company, intranets include many different machines, with different hardware and different operating systems. Most of them are supported by InnerGrid: Windows (NT, 2000 and XP), Mac OS X, and all UNIX systems (Linux, Solaris, AIX, IRIX...). No matter how heterogeneous your machines are, InnerGrid puts them all in a virtual box so you can see them as a single piece of hardware that will readily carry out your business processes.

Security

Protection mechanisms are a must nowadays, even in the intranet of a company. InnerGrid takes this into account and provides configurable multilevel security to fulfil the organization requirements, from those of a research group in a university to those in a great corporation. InnerGrid protects your data and your processes as they are distributed over your network. Communication among all the components of InnerGrid can be carried through SSL (Secure Sockets Layer), the most popular protocol that provides encrypted communications over a network. The Agents and the Server authenticate each other to make sure that non-trusted Agents cannot connect to the Server. So no external user can gain access to InnerGrid. InnerGrid has an authentication system that grants different permissions for defined users. For each user of InnerGrid two kinds of permissions are defined: those limiting the actions the user can perform within InnerGrid and those restricting the access to certain Grid resources, like computer power, the file system, distributed applications, and so on. These permissions are completely configurable to fulfil the requirements of any organization. For those Grid administrators not willing to spend time in configuring permissions, InnerGrid comes with three already defined levels of security for users.

5 The OuterGrid

OuterGrid is the concept to link to independent InnerGrid entities through secure links with each other. This can also mean sharing of data on distributed resources. Candidates for this concept are multinational research efforts or institutions as well as commercial companies willing to incorporate their international subsidiaries.

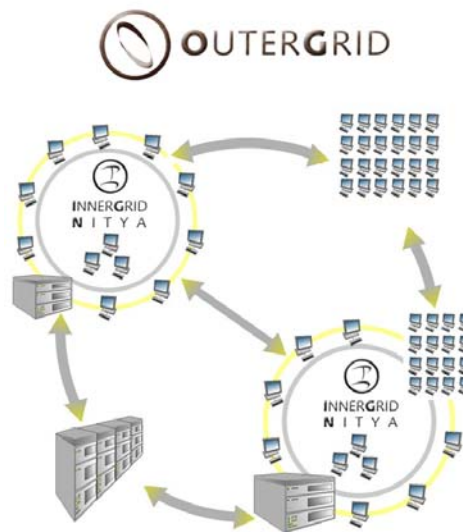


Abb.7: Schematic Concept of OuterGrid

6 Standards

One of the new paradigms in computer science is the web service one. The web service term can be described by dissecting the name:

"Web" indicates that it is a Web resource. That is anything that is identified by a Web address and can be accessed via Web protocols.

"Service" indicates that it is an application program that exposes its functionality through some type of Application Programming Interface (API).

The API describes to the programmer the list of things that he/she can do by accessing the service.

The Grid community has been developing specific standards for the web services that are specialized in offering Grid capabilities as a service. These are called Grid services. The current standardisation is centred on the OGSA (Open Grid Services Architecture), which is almost defined in its definitive way. GridSystems is committed to provide OGSA-compliant Grid services.

InnerGrid offers Grid services for the intranet. The SOAP (Simple Object Access Protocol) protocol is used as the underlying technology to implement the Grid services. SOAP is a set of conventions used to invoke code using XML (eXtensible Markup Language) over HTTP (the mainly used communication protocol in Internet).

Grid Computing und virtuelle Organisationen

Dietmar Erwin

Forschungszentrum Jülich GmbH
52425 Jülich
Germany
E-mail: d.erwin@fz-juelich.de

Grid Computing hat seit dem Erscheinen des Buches von Ian Foster und Carl Kesselman *The Grid: Blueprint for a New Computing Infrastructure* eine rasante Entwicklung erfahren. Das ursprünglich technologiegetriebene Ziel, Protokolle und Schnittstellen zu schaffen, um Benutzern den Zugriff auf verteilte Ressourcen (Rechenleistung, Daten, Archivspeicher, Instrumente, Ausgabemedien) in standardisierter Form zu ermöglichen, wurde erweitert. Grid-Technologie soll die Basis für virtuelle Organisationen bilden, in die Anbieter Ressourcen dynamisch einbringen und veränderlichen Nutzergruppen auf Zeit zur Verfügung stellen können. Anhand von zwei Grid-Lösungen, UNICORE und Globus, wird der Stand der Technik aufgezeigt und analysiert, in welchem Umfang die Ziele des Grid Computing heute realisiert sind. Insbesondere wird dargestellt, welche technischen und organisatorischen Herausforderungen noch zu bewältigen sind, um durch Grid Systeme virtuelle Organisationen in der Praxis umzusetzen. Dazu gehören Fragen wie Grid Security, Grid Economy und Grid Administration. Abschließend wird ein Ausblick auf die Entwicklungen von Standards zur Open Grid Service Architecture (OGSA) gegeben.

1 Virtuelle Organisationen

Das Konzept virtueller Organisationen steckt – wie auch Grid Computing – noch weitgehend in den Kinderschuhen. Eine abschließende, allseits akzeptierte Definition existiert noch nicht. Betriebswirte und Organisationstheoretiker charakterisieren virtuelle Organisationen durch ihre Eigenschaften. Bultje und van Wijk [1] definieren vier Unterkonzepte des Begriffs ‚virtuell‘, aus denen sich verschiedene Aspekte von virtuellen Organisationen ableiten:

- a) nicht real aber real aussehend: eine virtuelle Organisation erscheint nach außen wie ein reales, traditionelles Unternehmen, besteht aber in Wirklichkeit nur als Konglomerat von unabhängigen Partnern;

- b) immateriell, durch IT unterstützt: etwas existiert nicht, es wird lediglich durch Daten im Rechner kreiert. Eine *virtuelle Shopping Mall* ist ein Beispiel hierfür;
- c) potentiell präsent: die Organisation existiert nicht, kann aber bei Bedarf aus den vorhandenen Komponenten geschaffen werden. (Beispiel: *virtuelles Cluster*);
- d) existierend aber veränderlich: Die Organisation existiert, wird jedoch ständig dem Bedarf angepasst. Auf der Ebene von Unternehmen, entspricht dies einem *virtuellen Unternehmen*, auf der Ebene der Mitarbeiter einem *virtuellen Team*.

Eigenschaften virtueller Organisationen sind nach Mertens, et. al. [2]:

- Überschreitung von Organisationsgrenzen;
- Komplementarität von Kernkompetenzen;
- Ressource-Pooling;
- Gemeinsame Nutzung von Wissen;
- Geographische Verteilung;
- Veränderliche Zusammensetzung;
- Gleichheit der Teilnehmer;
- Einsatz elektronischer Kommunikation.

Das Ziel virtueller Organisationen ist flexibler und effektiver auf veränderte Märkte und Kundenanforderungen reagieren zu können als traditionelle Organisationen. Durch Informations- und Kommunikationstechnik wird die Schaffung virtueller Organisationen ermöglicht. Eine essentielle Voraussetzung ist Vertrauen [3]. Weiterführende Informationen findet man u.a. im *Electronic Journal of Organisational Virtualness* [4].

2 Grid Computing

Nahezu zeitgleich wurde der Begriff *Grid* für eine künftige, verteilte Computing Infrastruktur für Science und Engineering [7] geprägt. Dem Grid-Konzept liegt koordiniertes Ressource-Sharing zugrunde sowie Problemlösung in dynamischen, multi-institutionellen virtuellen Organisationen. Ressource-Sharing bezieht sich nicht nur auf Rechenleistung oder Daten im Grid, sondern integriert Instrumente, wie Beschleuniger, Teleskope oder Topographen, Visualisierung, Netze, Software sowie die Nutzer der Ressourcen. Bild 1 zeigt eine schematische Darstellung der möglichen Komponenten eines Grid.

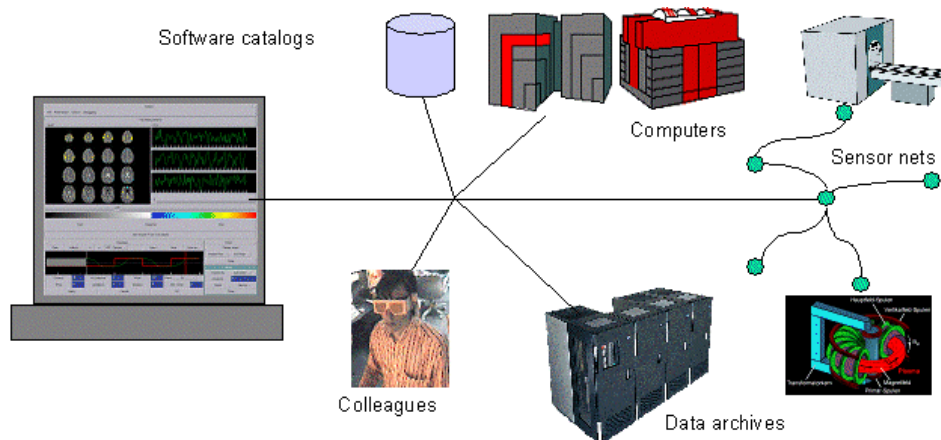


Abb. 1: Die Vision des Grid als Zusammenschluss aller relevanten Ressourcen, die zur Lösung eines Problems benötigt werden.

Die virtuellen Organisationen, die durch Grid Computing unterstützt werden, umfassen Anbieter und Nutzer von Ressourcen. Sie haben mit den Konzepten, die von der Betriebswirtschaftslehre geprägt sind, folgendes gemeinsam:

- Sie überschreiten Organisationsgrenzen: sowohl die Anbieter als auch die Nutzer der Ressourcen gehören i.A. verschiedenen Einrichtungen (Universitäten, Forschungseinrichtungen, Firmen) an;
- Komplementarität von Kernkompetenzen ist i.A. gegeben, z.B. Erzeugen von Daten an einem Beschleuniger, deren Auswertung an einem Universitätsrechenzentrum;
- Ressource-Pooling und Sharing ist ein primäres Ziel des Grid Computing;
- Gemeinsame Nutzung von Wissen ist zur Lösung komplexer Probleme essentiell;
- Geographische Verteilung ist bei wissenschaftlichen und technischen Kollaborationen i.A. gegeben;
- Veränderliche Zusammensetzung ist insbesondere eine Charakteristik der Nutzer des Grid;
- Einsatz elektronischer Kommunikation ist eine unabdingbare Voraussetzung für Grids.

Gleichheit der Teilnehmer ist keine notwendige Eigenschaft; es wird zwischen der Rolle von Anbietern und Nutzern unterschieden. Für Grids existiert ebenso wie für virtuelle Organisationen keine abschließende Definition. Vielmehr werden derzeit unterschiedliche Lösungen als Grids bezeichnet. Ein Extrem stellt zum Beispiel seti@home dar: Es existiert ein Benutzer (ein Problem soll gelöst werden) und eine sehr große Zahl von Providern (PC-Nutzern), die ihre freie

Rechenkapazität kostenlos zur Verfügung stellen. Am andern Ende des Spektrums werden Rechner, mit einheitlicher Administration und einheitlicher Zugangssoftware (z.B. Sun Grid Engine) als Grid bezeichnet.

3 Grids und virtuelle Organisationen

Die Lösung anspruchsvoller technischer und wissenschaftlicher Probleme unter Einsatz verteilter Ressourcen stellt an sich bereits hohe Anforderungen an die zugrunde liegende Software-Infrastruktur. Es müssen Protokolle und Schnittstellen existieren, die das Auffinden, Nutzen und Überwachen der verteilten Ressourcen ermöglicht. Nutzer müssen authentifiziert und autorisiert werden. Komplexe Aufgaben erfordern das Zusammenspiel und koordinierte Reservierung mehrerer Ressourcen. Hierzu sind Directory Services, Scheduling und Brokering, Daten-Replikation, Workload Management, Software Discovery und mehr erforderlich. Zu diesen technischen Anforderungen, die auch für verteilte Verarbeitung innerhalb eines Unternehmens relevant sind, kommen bei virtuellen Organisationen insbesondere die folgenden Aspekte hinzu, die sowohl technische als auch organisatorische Komponenten haben:

- Grid Economy
- Grid Security
- Grid Administration

Grid Economy: Die Ressourcen, die zur Lösung von Problemen im Grid benötigt werden, gehören in der Regel zu realen Organisationen. Die Ressourcen stehen den Benutzern nicht kostenlos zur Verfügung (Ausnahme *seti@home*); diese werden vielmehr mit den Kosten belastet, sei es durch explizite Rechnungsstellung, durch Belastung einer firmeninternen Kostenstelle, oder durch Abbuchung von einem ‚Rechenkontingent‘, das vorher von einer Vergabeorganisation bewilligt wurde.

Grid Computing erfordert, dass die erbrachten Leistungen gegenseitig verrechenbar sind, unabhängig von bestehenden technischen Accountingsystemen. Der Grid-Nutzer erwartet bestimmen zu können, wann und zu welchen Kosten die Aufgabe bearbeitet werden soll. Die Anbieter müssen andererseits sicher sein, dass die vereinbarten und gelieferten Leistungen auch bezahlt werden. Die dem Grid-Modell zugrunde liegende Analogie des ‚Stromnetzes‘ ist zwar eingängig, aber hierfür viel zu simpel. Forschungsarbeiten zum Thema Grid Economy haben gerade begonnen [5].

Grid Security: Die Nutzung verteilter Systeme setzt heute fast immer voraus, dass der Anwender in jedem System bekannt ist und einen z.B. durch Benutzerkennung und Passwort geschützten Zugang hat. Ziel im Grid ist, sich

lediglich an einer Stelle auszuweisen müssen (single sign-on); damit ist nicht notwendigerweise der Zugriff auf Ressourcen (Autorisierung) verbunden. Die übliche Technik ist der Einsatz von Zertifikaten nach dem X.509-Standard. Voraussetzung ist eine funktionierende Public Key Infrastruktur (PKI). Alle Teilnehmer in einer virtuellen Organisation müssen die Policy der ausstellenden Zertifizierungsinstanz akzeptieren. Falls mehrere Zertifizierungsinstanzen involviert sind, müssen sowohl alle Policies von allen Teilnehmern anerkannt werden, als auch die entsprechende Softwareunterstützung vorhanden sein. Dieser Aspekt wird in Grid-Projekten oft ignoriert.

Grid Administration: Bereits in traditionellen Organisationen ist die Administration von Rechnern eine nicht-triviale Aufgabe. Dies umfasst sowohl das Benutzermanagement, die Installation von Software, als auch den Betrieb und das Problemmanagement. Benutzermanagement kann innerhalb einer Firma zentralisiert werden oder nach einheitlichen Standards und mit einheitlichen Namenskonventionen durchgeführt werden. In einer virtuellen Organisation ist dies eine im Allgemeinen unlösbare Aufgabe. Angenommen die virtuelle Organisation V1 habe sich auf einen Standard S1 verständigt, während V2 den inkompatiblen Standard S2 nutzt, kann ein neuer Teilnehmer O3 nie gleichzeitig V1 und V2 angehören.

Software-Installation im Grid erfordert nicht nur, dass alle Teilnehmer kompatible Software-Stände einer Grid-Software einsetzen, sondern dass diese auch noch mit der lokal eingesetzten Software kompatibel sein muss. Die Anforderungen an die Entwickler von Grid-Software sind entsprechend hoch. Es genügt z. B. nicht, dass die Grid-Software nur mit einer bestimmten Java-Version zusammenarbeitet. Das Problem ist nicht neu, wird aber durch die organisatorische Unabhängigkeit der beteiligten Organisationen verschärft.

Grid-Lösungen stellen hohe Anforderungen an Erkennen und Beseitigen von Problemen. Administratoren sind in der Lage, Fehler auf den von ihnen betreuten Systemen zu diagnostizieren und zu beseitigen und so betroffenen Benutzern zu helfen. In einer virtuellen Organisation wird die Vergabe von Administrator-Rechten an ‚fremde‘ Administratoren die Ausnahme bleiben. Grid-Lösungen können heute bereits erkennen, in welchem der beteiligten entfernten Systeme ein Problem aufgetreten ist; die lokalen Administratoren sind jedoch auf die Zusammenarbeit und Unterstützung der dortigen Kollegen angewiesen.

4 Grid-Lösungen

In zahlreichen Forschungs- und Entwicklungsprojekten sind wichtige Beiträge zum Grid-Computing geschaffen worden. Es handelt sich dabei meist um Prototypen, um ausgewählte Aspekte des Grid-Computing zu demonstrieren. Als

Beispiele für umfassende Lösungen seien Globus [10] und UNICORE [11] aufgeführt, die komplementäre Ansätze verfolgen.

Das Globus Toolkit™ stellt den Anwendern Schnittstellen zur Verfügung, um Grid-Anwendungen zu entwickeln und Grid-Dienste zu nutzen. Die Globus-Architektur ist in Abbildung 2 dargestellt. Die Globus-Sicherheitsinfrastruktur basiert auf X509-Zertifikaten und verwendet das Prinzip der Delegation von Rechten, so dass eine Globus-Anwendung weitere Anwendungen dynamisch starten kann, fast als ob sie vom Benutzer selbst initiiert worden sind. Resource Management wird mittels GRAM (Globus Resource Allocation Manager) realisiert, mit dessen Hilfe der Benutzer Anwendungen auf entfernten Systemen starten und überwachen kann. Zugriff auf entfernte Daten geschieht mittels GASS (Globus Access to Secondary Storage), welches das Staging von Dateien durchführt. Grid-Informationsdienste werden von Globus über MDS (Monitoring und Directory Service) bereitgestellt. MDS stellt eine einheitliche Schnittstelle zu der statischen und dynamischen Information über einzelne Ressourcen, dem Grid Resource Information Service (GRIS), dar. Ein weiterer Bestandteil von MDS ist der Grid Index Information Service (GIIS), der die Gesamtheit der Ressource in einem Grid enthält.

Das Architekturdiagramm in Abbildung 2 zeigt den Kontrollfluss für die Jobs submission von einem Globus-Klienten (links) und für den vom Klienten initiierten Filetransfer mittels GridFTP zwischen zwei Servern (rechts).

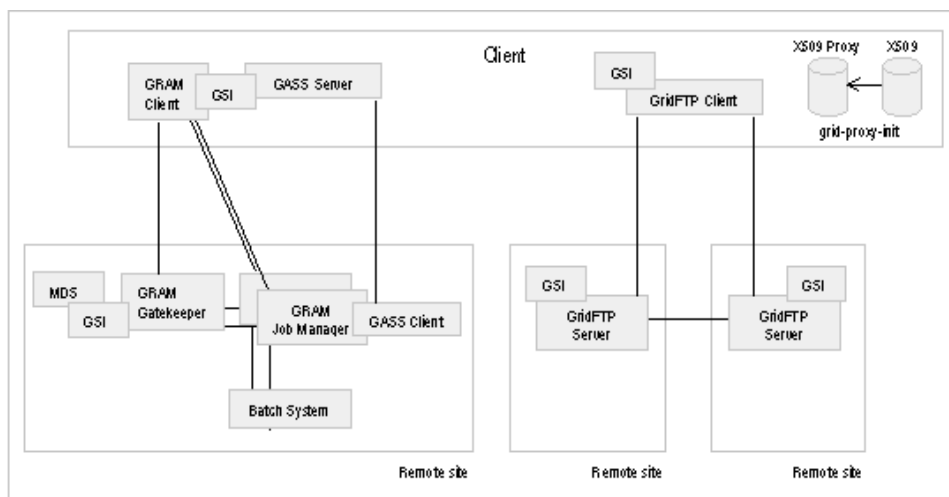


Abb. 2: Die Globus-Architektur

Anwendungen können diese als Bibliotheken verfügbaren Dienste direkt über Globus-APIs aus C, Java oder Python ansprechen. Condor-G oder Nimrod/G sind Grid-Scheduler, die Globus-Ressourcen nutzen und MPICH-G2 ist eine Grid-fähige Message-Passing-Interface-Bibliothek. Globus stellt Grid-Entwicklern Basisdienste und höherwertige Dienste zur Verfügung.

Im Gegensatz zu der horizontalen Integration bei Globus hat UNICORE den Weg gewählt, ein vertikal integriertes System für einen stoßkantenfreien („seamless“), sicheren und intuitiven Zugang zu entfernten Ressourcen zu schaffen.

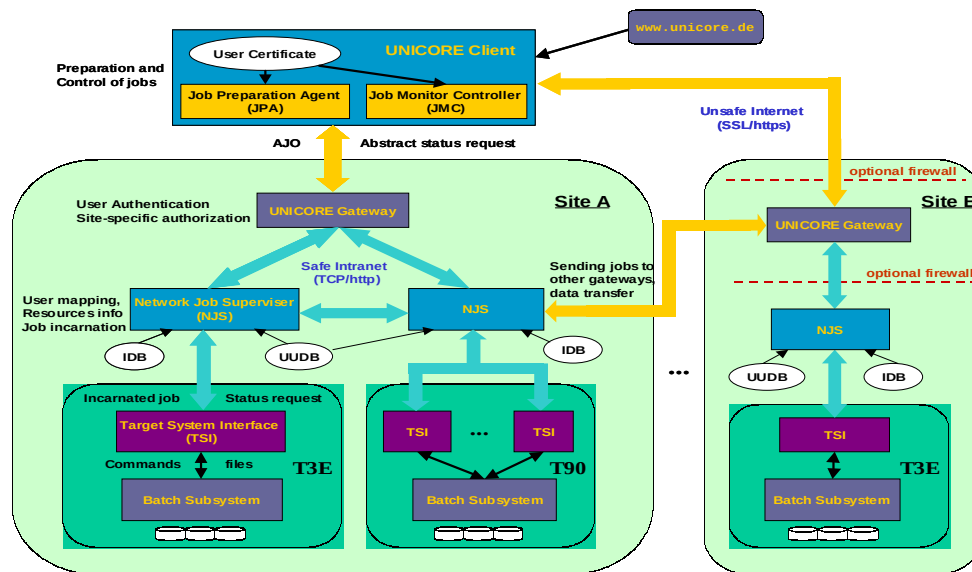


Abb. 3: Die UNICORE-Architektur

Über ein graphisches Interface, den UNICORE-Klienten, beschreibt der Benutzer die zu lösende Aufgabe und die dafür benötigten Ressourcen. Die Beschreibung umfasst die auszuführenden Programme, die Quelle von Eingabedateien, das Ziel für Ergebnisdaten sowie die Abhängigkeiten von Ausführungsschritten. Es können Schleifen, bedingte Abarbeitungszweige und Zeitsteuerung formuliert werden. Eine Aufgabe (UNICORE-Job) kann gleichzeitig mehrere Rechner in verschiedenen Zentren nutzen. In der aktuellen Produktionsversion selektiert der Benutzer explizit die Zielsysteme (Organisationen und Rechner), auf denen die einzelnen Teilschritte ausgeführt werden sollen. UNICORE prüft vor der Submission, ob die angeforderten Hardware- und Softwareressourcen verfügbar sind und warnt den Benutzer, falls dies nicht der Fall ist. Im Projekt GRIP [12] wird ein Broker entwickelt, der geeignete Zielsysteme im Grid findet. Der UNICORE-Klient erzeugt aus den Anforderungen eine abstrakte Darstellung des Jobs. Dieser kann auf dem Klienten gespeichert werden. Ein syntaktisch korrekter

Job wird zu einem UNICORE-Gateway submittiert, das die Identität des Benutzers anhand des Zertifikats prüft. Gültige Jobs werden an den Network Job Supervisor weitergeleitet, der den abstrakten Job in einen zielsystemspezifischen Job übersetzt. Der Benutzer muss nicht wissen, wie der Aufruf eines Fortran-Compilers lautet; dies ist in der entsprechenden Datenbank beim NJS festgehalten. Ändert ein Rechenzentrum das Verzeichnis, in dem sich der Compiler befindet, muss dies nur in der lokalen Datenbank nachgehalten werden. Benutzerjobs laufen ungeändert weiter. Ebenso übersetzt NJS das Zertifikat in den Benutzernamen auf dem Zielsystem. Dadurch ist es nicht erforderlich, bestehende Namenskonventionen zu ändern.

Teilaufgaben für ein anderes Zielsystem werden vom NJS an dieses weitergeleitet. NJS übernimmt für laufende Jobs die Synchronisation der Teilaufgaben, die Ablaufsteuerung, den Transfer von Dateien zwischen verschiedenen Systemen und die Zwischenspeicherung von Ergebnissen, bis sie vom Benutzer abgerufen werden. Über das Graphische Interface (GUI) hat der Benutzer jederzeit die Möglichkeit, den Status aller laufenden Jobs zu sehen, Ergebnisse auf dem Klienten zu speichern oder zu visualisieren, sowie Jobs abubrechen und zu löschen. Ausführliche Information zu Architektur und Funktionsumfang von UNICORE findet der Leser in [6].

5 Status und Ausblick

Die Theorie virtueller Organisationen setzt eine adäquate ICT-Infrastruktur voraus, ohne darauf einzugehen, wie diese realisiert ist. Grid-Computing adressierte zunächst die technischen Aspekte, um *distributed computing* und *resource sharing* zu ermöglichen, wie in der Vision des Grid formuliert. Die Terminologie und die Konzepte virtueller Organisationen erwiesen sich als geeignet, die organisatorischen Aspekte im Grid zu beschreiben. Grid-Lösungen wie UNICORE und Globus haben prototypisch nachgewiesen, dass Grids als Infrastruktur für virtuelle Organisationen geeignet sind. Wichtige Forschungs- und Entwicklungsarbeiten bleiben nach wie vor zu leisten, insbesondere auf den Gebieten Grid Economy und Grid Security. Einen deutlichen Schub werden diese Entwicklungen aus den gemeinsamen Arbeiten von Industrie und Forschung erhalten. Haben bisher kommerzielle Anbieter auf Web Services gesetzt, die Forschungs-Community auf Grids, so arbeiten beide seit etwa 2002 gemeinsam an der Open Grid Service Architecture [9]. Im Global Grid Forum [13] wurde hierzu als einer der ersten Schritte ein Entwurf für einen offenen Standard unter dem Titel *Open Grid Services Infrastructure* [14] erarbeitet. Grid-Lösungen, die diesen Standard implementieren, werden zur Akzeptanz des Grid Computing beitragen. Für UNICORE werden im EU-Projekt GRIP die ersten OGSA-konformen Komponenten entwickelt; diese Arbeiten sollen im beantragten Projekt UniGridS (Uniform Interface to Grid Services) abgeschlossen werden.

Globus stellt mit seinem Toolkit 3 (GTK3) ebenfalls OGSA-Schnittstellen zur Verfügung. Entwickler, Anbieter und Nutzer von Grids erwarten und arbeiten gemeinsam daran, dass die Open Grid Services Architecture der akzeptierte Standard für Informations- und Kommunikationstechnik für virtuelle Organisationen wird und dass die offenen Fragen gemeinsam von Industrie und Forschung zügig beantwortet werden.

Referenzen

- [1] Bultje, J. van Wijk, *Taxonomy of Virtual Organisations based on definitions, characteristics and typology*, VoNet: The Newsletter, <http://www.virtual-organizations.net>, 2(3), 7-20 (1998)
- [2] P. Mertens, J. Gries, D. Ehrenberg, *Virtuelle Unternehmen und Informationsverarbeitung*, Springer, Berlin, 64 ff (1998)
- [3] Ch. Lane, R. Bachmann, eds., *Trust Within and Between Organisations*, Oxford University Press (1998)
- [4] <http://www.virtual-organization.net>
- [5] R. Buyya, *Economic-based Distributed Resource Management and Scheduling for Grid Computing*, PhD Thesis, <http://www.cs.mu.oz.au/~raj/thesis> (2002)
- [6] Erwin, D., ed., *UNICORE Plus Final Report – Uniform Interface to Computing Resources*, <http://www.unicore.org> (2003)
- [7] I. Foster and C. Kesselman, ed. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufman Publishers (1998)
- [8] I. Foster and C. Kesselman, S. Tuecke, *The Anatomy of the Grid: Enabling Scalable Virtual Organizations*. International J. Supercomputer Applications, 15(3), (2001)
- [9] I. Foster and C. Kesselman, J. M. Nick, S. Tuecke, *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems*, www.globus.org/research/papers/ogsa.pdf (2002)
- [10] <http://www.globus.org>
- [11] <http://www.unicore.de>
- [12] <http://www.grid-interoperability.org>
- [13] <http://www.ggf.org>
- [14] S. Tuecke, et al, *Open Grid Services Infrastructure (OGSI)*, <http://sourceforge.net/projects/ggf/> (2003)

Storage Virtualisation

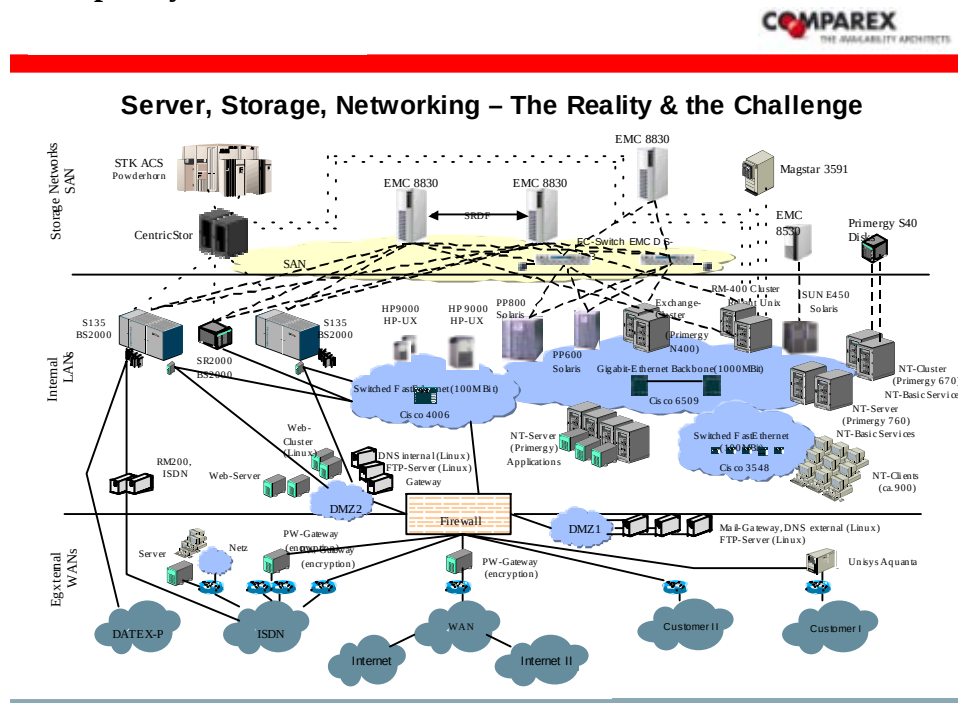
What is it? What does it do? How will we Benefit?

Michael D. Senior

COMPAREX Informationssysteme GmbH
68165 Mannheim
Germany
E-mail: michael.senior@comparex.de

Storage Virtualisation is threatening to revolutionize the way we manage and perceive disk storage today. Market analysts are prone to suggest that high-end and mid-range storage systems have become commodity products to such an extent that they are more or less interchangeable. Whether this is true or not is immaterial as the concept of storage virtualisation will help make it true. What does virtualisation promise? What, in brief, are the technologies involved? Where do we stand today and what might we expect in the not too distant future? Who are the vendors of today and tomorrow? These are some of the questions addressed, and hopefully answered, in the context of this presentation.

1 Complexity



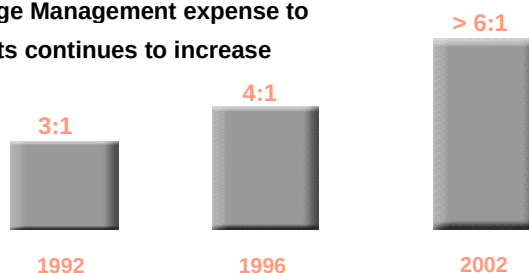
There was a time not too long ago, during the era of mainframe predominance, with big iron operating systems such as MVS, VSE, BS2000 that we believed the IT world was a somewhat complex environment. With the advent of a host of Unix flavours and Windows all advocating the world of Open Systems in their own unique and proprietary way, large data centres today have evolved into nightmares of physical mazes. The slide above provides an overview of an actual data centre at a customer site. Most organisations today have probably evolved such that all have some or several versions of Windows installed, at least one but probably two or more versions of Unix, Netware and possibly also a mainframe operating system to contend with. Even within product families exploiting a homogeneous Windows environment, transferring data across common resources from one application to another is not always simple. An example of this is the above slide which did not allow itself to be transferred from PowerPoint to Word without having to resort to insider tricks known only to those well-versed in graphics. This is just an extremely trivial example of the challenges facing users in heterogeneous environments. Despite all the benefits provided by today's drop and drag, plug and play technology, using a computer can still be frustrating and time-consuming. As storage networks gradually replace direct-attached storage environments, storage management becomes simpler, i.e. less complex, and can be more cost-effectively implemented. Without storage networking there would be no storage virtualisation and virtualising storage i.e. effectively hiding the physical back-end disk storage from the application will contribute in no small way to easing the storage manager's task enabling him to handle terabytes of disk capacity.

2 Costs, Management & Virtualisation



Management Costs Escalate...

Ratio of Storage Management expense to
Hardware Costs continues to increase



Source: Storage
Networking World

Today: Over **\$6** spent on Storage Management for every **\$1** spent on Hardware. Again, complexity is responsible for this.

"Recent studies have found that the initial purchase price for storage hardware accounts for only about 10 percent of the total cost of ownership."

Linda Sanford,
Senior Vice-President,
IBM Storage Systems Group

Is Virtualisation the Key ?

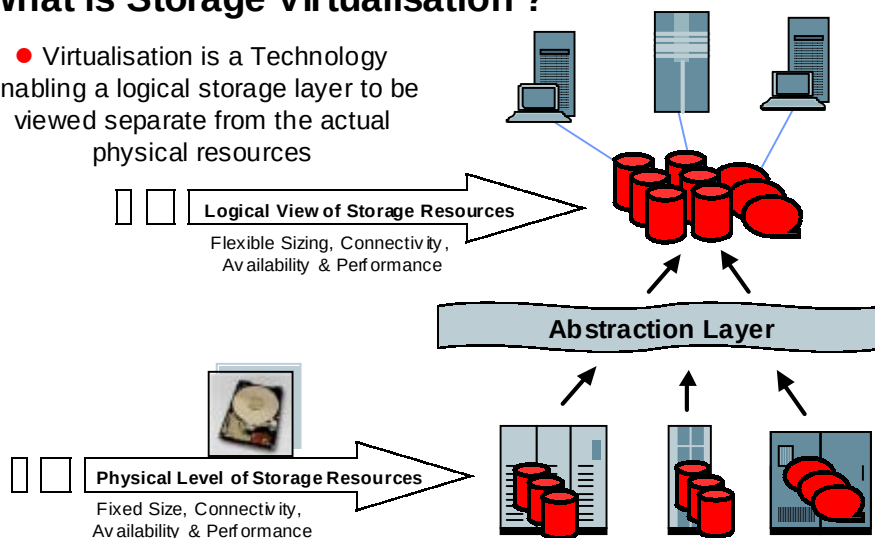
The actual statistics touted by industry analysts such as Gartner or MetaGroup as well as those of the vendors themselves may differ, but on the central issue they all agree: in a world where annual storage growth is seldom less than 50% and frequently higher than 100%, the costs of managing storage have long ago overtaken the cost of the storage hardware itself. In the above, simply as an example, Linda Sanford (Vice President at IBM until recently responsible for the OS/390 division and now heading up the Storage Systems Group) estimates non-hardware, management-related costs at 90% of TCO (Total cost of Ownership). The challenge, therefore, is not to strive for reduced hardware prices – this happens as a result of Moore’s Law as a matter of course – but to analyse the usage of personnel, examine procedures and overall efficiency across the storage environment. Centralisation and consolidation via a SAN will certainly enhance efficiency and reduce costs. If vendors are to be believed, the increasing availability of virtualisation technology should lead to further considerable management savings.

3 A Simple Concept



What is Storage Virtualisation ?

- Virtualisation is a Technology enabling a logical storage layer to be viewed separate from the actual physical resources



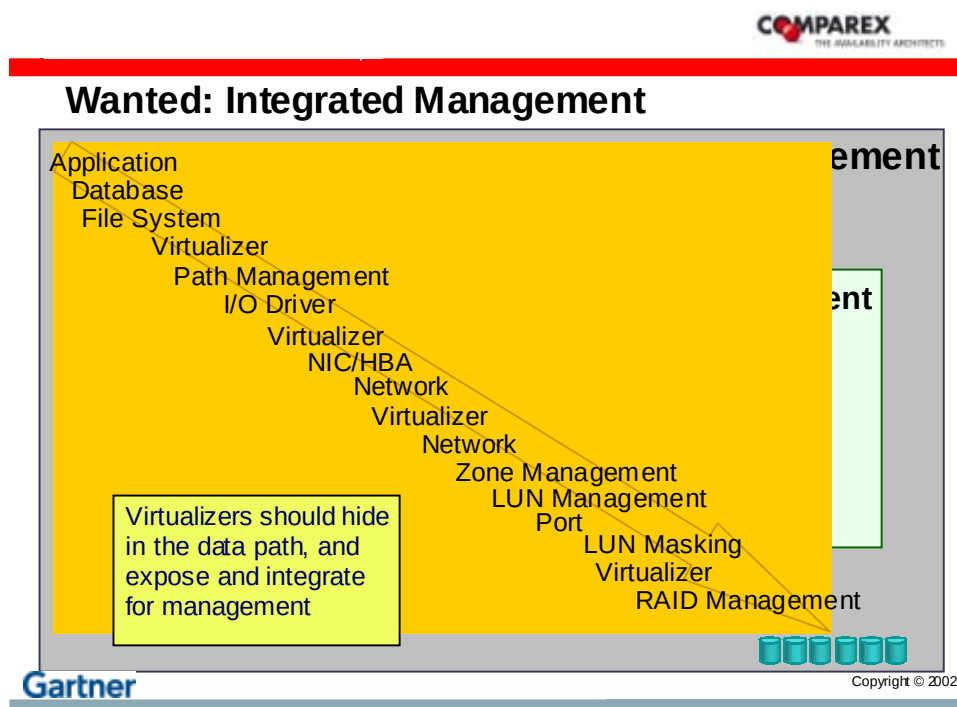
“The ability of abstract storage to present a common management interface for Storage Management software, and the ability to add data movement functions at the fabric level for such functionality as remote copy and point-in-time copy regardless of the type of back-end storage.”

Randy Kerns, Evaluator Group, 05/2003

In accordance with the acronym KISS – Keep it Simple Stupid, or perhaps better still, Keep IT Simple Stupid – virtualisation appears invitingly simple and threatens to take the complexity out of storage and its management. The new abstract virtualisation layer renders the underlying physical storage layer practically invisible to the application. Is it an enterprise ESS “Shark” system from IBM, an XP from hp, or a DMX from EMC? Or perhaps a mid-range EMC CLARiiON, an IBM FASTT or an hp EVA? Or any mix of these or similar systems? After virtualisation the question is superfluous. The application sees a virtual layer of disks – SCSI, FC, ATA – the type, the technology and the vendor are no longer relevant. Virtualisation looks after disk residence, copy and migration functionality no matter which systems are involved.

The benefits are manifold, not least the ability to introduce disk backup i.e. production data on expensive disks with high availability backing data up to e.g. cheap ATA disks for fast recovery.

4 Simple Concept - Complex Implementation

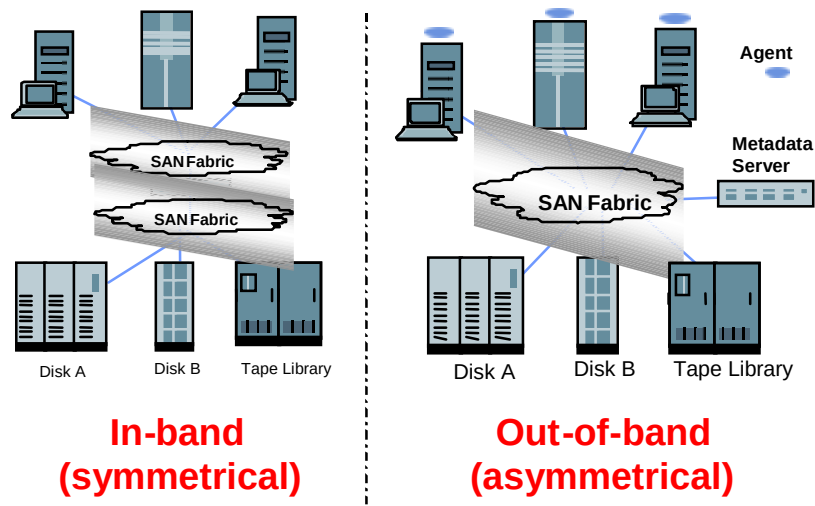


The virtualizer sits in the middle of the storage area network and captures every I/O initiated. Depending on the technology used, mapping tables are invoked to

reroute the data to wherever the system deems it ought to go. If we look at the data path exemplified above then we can see that this is no trivial matter. Central to the network the virtualiser needs to take into account all of the incidences along the data path and guarantee both the integrity of the data to be stored and the recovery of that data. Most systems claiming today to be in a position to be able to virtualise the storage network are simply benefiting from current market hype. Smaller companies have neither the resources nor the technology and larger companies have jumped on the virtualisation bandwagon either very late, and as such have not gone very far down the development road, or have as yet not even joined the fray.

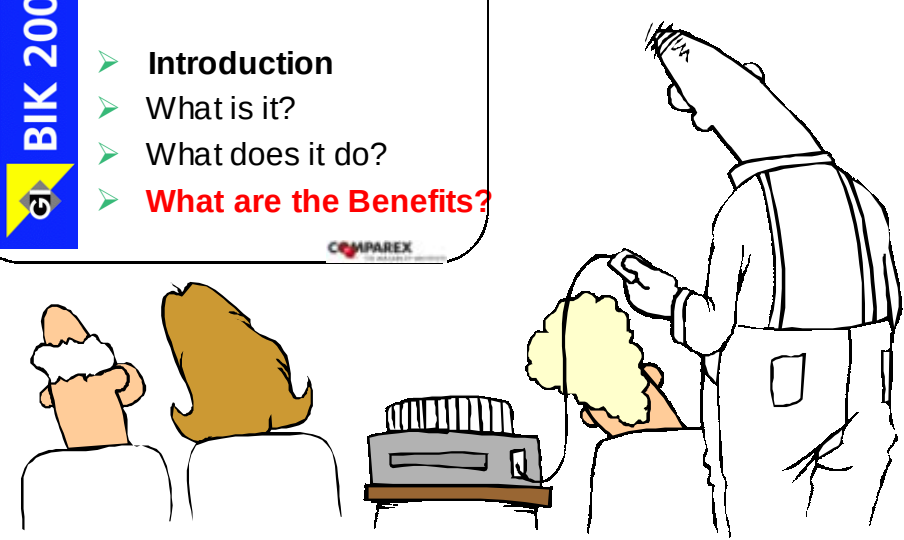
5 In-Band v. Out-of-Band

Storage Virtualisation – 2 Methods



Two technologies dominate: in-band (symmetrical) and out-of-band (asymmetrical). Most implementations have opted for the in-band solution. The following slides will examine how the two technologies differ, which companies have opted for which technology and where are the benefits and the downsides of the technologies.

6 Promises, Promises



BIK 2003

Storage Virtualisation:

- Introduction
- What is it?
- What does it do?
- **What are the Benefits?**

COMPAREX

This final section will examine the hype, the benefits, the pitfalls. Where do we stand today? What can we really expect from storage virtualisation? And a personal opinion on which vendors are likely to survive beyond the year 2004.

Beiträge der RWTH Aachen zur kooperativen IT-Versorgung im Lande NRW

**Christian Bischof
Klaus Brühl
Thomas Eifert**

Rechen- und Kommunikationszentrum der RWTH Aachen
52074 Aachen
Germany
E-mail: {bischof, bruehl, eifert} @rz.rwth-aachen.de

Das Land NRW hat im Januar 2002 ein Strategiepapier zur kooperativen Versorgung der Hochschulen des Landes im Bereich des Hoch- und Höchstleistungsrechnen erarbeitet. In diesem Papier wird die Bildung von überuniversitären Verbünden zur Lösung der spezifischen infrastrukturellen Problemen propagiert.

Der Rechnerverbund NRW (RV-NRW) wird hierbei als adäquates Mittel zur effizienteren kooperativen Nutzung von Ressourcen, aber auch als Möglichkeit zum Aufbau eines gemeinsamen Kompetenznetzwerkes zwischen den Hochschulen gesehen.

Das Rechen- und Kommunikationszentrum der RWTH Aachen bietet im Rahmen dieses Verbundes Landesdienste für die Hochschulen in NRW an. Neben den eher traditionellen HPC-Rechendiensten werden demnächst zusätzlich ein CPC-System (Collaborative Product Commerce) zum verteilten Konstruieren bzw. zum Prozessmanagement von komplexen technischen Systemen sowie Archivdienste landesweit angeboten.

1 Das kooperative Versorgungskonzept des Landes NRW

Für Hochschulen ist eine leistungsfähige Informations- und Kommunikationsinfrastruktur (I&K), angefangen von der Basisvernetzung bis hin zum High Performance Computing als Differenzierungsfaktor im nationalen und internationalen Wettbewerb, mittlerweile ein unverzichtbares Instrument. Auch wenn die Diskussion der Wettbewerbsfähigkeit in einer modernen und globalen Bildungslandschaft stark von anwendungs- und inhalteorientierten Überlegungen geprägt ist, darf die kontinuierliche Weiterentwicklung einer adäquaten I&K-Infrastruktur nicht vergessen werden; bildet sie doch das Rückgrat für programmatische Aktivitäten der Hochschulen.

Durch die rasche technologische Weiterentwicklung der I&K-Infrastruktur und die immer stärker werdende Abhängigkeit davon hat sich in den letzten Jahren ein enormer Investitionsbedarf aufgebaut. Deshalb ist es notwendig, durch die Verankerung von nachhaltigen Planungskonzepten in den Hochschulen eine leistungsfähige und flächendeckende I&K-Infrastruktur aufzubauen, sowie diese kontinuierlich und bedarfsgerecht weiterzuentwickeln.

Planungen für den I&K-Bereich sind in NRW seit dem „ADV-Gesamtplan für die Hochschulen des Landes“ aus dem Jahre 1975 bekannt und üblich. Aus dieser Kontinuität heraus wurde eine Anfrage des Wissenschaftsrat an die Länder zur weiteren Planung der Bundeshöchstleistungsrechenzentren im August 2001 zum Anlass genommen, eine entsprechende Konzeption für das Hoch- und Höchstleistungsrechnen auf Landesebene zu entwickeln.

Unter dem Titel „High Performance Computing in Nordrhein-Westfalen – Kooperatives Versorgungskonzept für das Hoch- und Höchstleistungsrechnen in den Hochschulen des Landes (HIPEC NRW)“ wurde im Januar 2002 von einer Arbeitsgruppe der Hochschulen unter Federführung des Ministeriums für Wissenschaft und Forschung (MWF) ein Strategiepapier für ein kooperatives Versorgungskonzept für das Hoch- und Höchstleistungsrechnen in den Hochschulen des Landes erarbeitet [1]. Ziel des Papiers ist es, Teilbereiche der jeweiligen informations- und kommunikationstechnischen Infrastruktur der Hochschulen mittels einer kooperativen Lösung nachhaltig auszurichten.

2 Die Situation der Rechenzentren in den Hochschulen

Die Rechenzentren sind die zentralen Dienstleister für den I&K-Bereich in den Hochschulen. Die gestiegenen qualitativen und quantitativen Anforderungen an die Infrastruktur und die damit erbrachten Dienste müssen trotz angespannter Finanz- und Personallage sichergestellt werden. Die Benutzer verlangen zuverlässige Mail-, Web-, Rechen-, und Datendienste sowie Beratung bei der Konfiguration der i.A. dezentral über Webschnittstellen verwaltenden Dienste. Darüber hinaus führt der ansteigende Einsatz von Simulationswerkzeugen zu einer stetig wachsenden Nachfrage an Anwenderberatung, Schulung und methodischer Unterstützung. Dies erfordert, da lokale Produktivitätssteigerungen in den RZ meist nur noch begrenzt möglich sind, aufgrund der personellen und finanziellen Restriktionen eine andere Organisationsform, welche über die traditionellen Universitätsgrenzen hinausgeht. Hier bieten sich kooperativen Lösungen wie der Rechnerverbund NRW (RV-NRW) als ein möglicher Ausweg an.

Ziel eines solchen Verbundes muss es sein, durch die Organisation der Ressourcen die Dienstleistungen der einzelnen Hochschulrechenzentren zu

einem neuen Ganzen zu ergänzen und das Niveau der IT-Versorgung damit zu verbessern bzw. kostengünstiger zu gestalten. Die gestiegene Unabhängigkeit der Hochschulen, die Konkurrenzsituation sowie der fiskale Druck erzeugen hierbei natürlich ein erhebliches Spannungsfeld.

3 Der Rechnerverbund NRW (RV-NRW)

Die RZ in NRW haben 1999 in einer Eigeninitiative den Rechnerverbund NRW (RV-NRW) gegründet und eingerichtet. Das primäre Ziel des Verbundes war es, einen innovativen, flexiblen und nutzernahen Ansatz für den Bereich des Hochleistungsrechnens zu schaffen und dabei auf die Problematik der schnellen Innovationszyklen vor dem Hintergrund langer Beschaffungszyklen besser reagieren zu können. Durch die unterschiedlichen Investitionszeitpunkte stehen in einzelnen Universitäten in schnellerer Reihenfolge die neuesten Systeme im Verbund zur Verfügung. Somit wird eine Innovationsspirale bezüglich der IT-Ausstattung realisiert. Der Verbund versteht sich nicht als Ersatz für die Bundes-Höchstleistungsrechner sondern als eine sinnvolle Ergänzung.

Ausgangspunkt der Überlegungen für den Rechnerverbund ist die Abdeckung des Rechenbedarfs der Hochschulen entsprechend der „Rechnerpyramide“ (Abb. 1). Diese Darstellung entspricht dem von der DFG seit Jahren empfohlenen Konzept der kooperativen Rechnerversorgung.

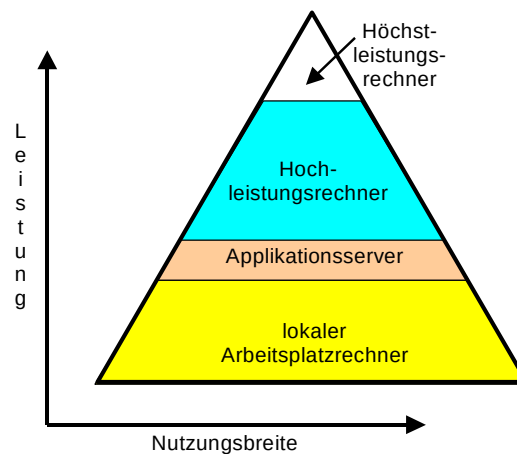


Abb. 1 : Rechnerpyramide

Der Verbundgedanke bezieht sich dabei sowohl auf Ressourcen als auch Kompetenzen [2].

3.1 Aufbau eines Kompetenznetzwerkes

Aufgaben, die in einzelnen Hochschulrechenzentren gleichermaßen anstehen oder nicht zur Kernkompetenz gehören, können arbeitsteilig gelöst werden, um so Potential für die dringende Weiterentwicklung der IT-Infrastrukturen und IT-Anwendungen zu schaffen. Im Gegensatz zu der häufig diskutierten Möglichkeit des Outsourcing von Diensten verbleiben allerdings das Know-how und die Kontrolle über den Einsatz von neuen Technologien bei den Hochschulrechenzentren. Dies spart Kosten, verstärkt die Motivation der Mitarbeiter und unterstützt den Innovationsanspruch der Hochschulen.

Als ein Beispiel sei hier die Arbeit im Rahmen der Tivoli-Projektes angegeben. Nach dem Abschluss einer Landeslizenz mit der Fima IBM/Tivoli wurden mehrere überuniversitäre Arbeitsgruppen gebildet, die in Pilotinstallationen die Software in die Produktionsumgebung integrieren. Das RZ der RWTH setzt in einer Pilotinstallation das Produkt „Tivoli Identity Manager“ zur Verwaltung von Benutzern und der automatisierten Provisionierung von Diensten ein.

3.2 Gegenseitige Verfügbarmachung von DV-Ressourcen

Da durch die langen Reinvestitionszyklen bei der Beschaffung von Hardware oder wegen der personellen bzw. fiskalischen Restriktionen vor dem Hintergrund der jeweiligen programmatischen Schwerpunktsetzung nicht alle Dienste nachhaltig an jeder Hochschule unterstützt werden können, ist es sinnvoll, solche Ressourcen von mehreren Hochschulen gemeinsam und im gegenseitigen Austausch zu nutzen. Teure Systeme, die aus finanziellen Gründen nicht an vielen Hochschulen installiert werden können, können so auch anderen Forschern einfach zugänglich gemacht werden.

Hierbei wird der Ansatz verfolgt, dass ein RV-Nutzer sich die für sein HPC-Problem, von der interaktiven Entwicklung über die Optimierung bis zur Produktion, die für ihn in Summe effizienteste Umgebung auswählt. Der RV ist also kein Verbund im „Grid“-Sinne, bei dem Aufträge transparent verteilt werden. Durch eine gemeinsame, verbundweite Nutzerverwaltung haben die Nutzer allerdings Zugriff auf die angeschlossenen Ressourcen.

4 RV-Dienste des Rechen- und Kommunikationszentrums der RWTH

Das im HIPEC-Papier des MWF angeführten kooperative Konzept, sowie die Idee des RV-NRW sind Grundlage mehrerer Aktivitäten des RZ. Im folgenden werden einige Dienste, die auch im Verbund zur Verfügung stehen, näher beschrieben.

Von grundlegender Bedeutung ist dabei das Faktum, dass bei intensiver Zusammenarbeit in überregionalen Verbünden die WAN-Komponente zu einem „Landes-LAN“ mutiert und als Konsequenz von dem wide-area-Netzprovider eine gesicherte Dienstqualität auch für zeitkritische und interaktive Dienste erwartet wird.

4.1 HPC-Server (SMP-Cluster)

Das Rechen- und Kommunikationszentrum der RWTH leistet seit 1989 Landesdienste. Bis 2001 betrieb es den „Landesvektorrechner“, seitdem ist ein leistungsfähiger SunFire SMP-Cluster in Betrieb. Die zeitliche Entwicklung der HPC-Ressourcen ist in Abb. 2 veranschaulicht.

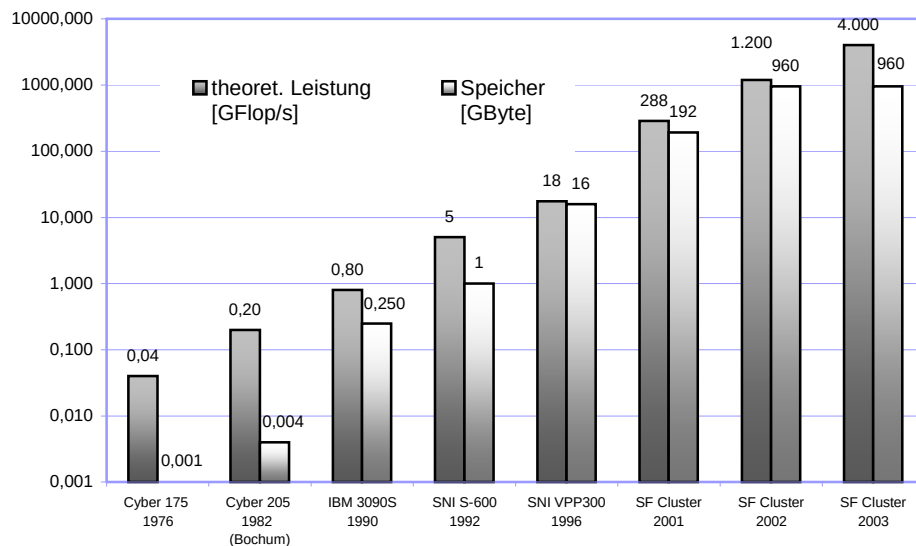


Abb. 2: Historie des Hochleistungsrechnens an der RWTH

Der RWTH steht ein leistungsstarker SunFire SMP-Cluster zur Verfügung. Die Bewilligung des Rechners durch das MWF erfolgte unter der Maßgabe, dass das System zu 20% im RV-NRW den übrigen Hochschulen zur Verfügung steht. Dieses Kontingent wird regelmäßig erreicht bzw. übertroffen.

Derzeit leistet das System 1,2 Tflops bei einem Hauptspeicher von ca. 1 Tbyte. Nach dem Austausch aller Prozessoren durch dual-core UltraSPARK IV CPUs

zu Beginn des nächsten Jahres wird das System eine akkumulierte Leistung vom 4 Tflops besitzen.

Die Verfügbarkeit von drei durch SunFire Links intern breitbandig vernetzten Clustern mit unterschiedlich starken SMP-Knoten erlaubt es, flexibel auf den Anforderungs-Mix der Nutzer mit seriellen, SMP-parallelen bzw. autoparallelisierten und MPI-parallelen Jobs unterschiedlicher Größe zu reagieren. Zur Verfügung stehen den Anwendern die in Abb. 3 angegebenen Ressourcen.

Programmiermodell / Ressourcen	theoret. Leistung	Speicher
SMP Programmierung		
• Sun Fire 6800	43 (144) Gflops	24 Gbyte
• Sun Fire 15K	130 (432) Gflops	144 Gbyte
MPI, Hybrid		
• Sun Fire 6800 Cluster (2 Cluster)	345 (1.152) Gflops	192 Gbyte
• Sun Fire 15K Cluster	518 (1.728) Gflops	576 Gbyte
• Sun Fire 6800 & Sun Fire 15K	1.200 (4.000) Gflops	960 Gbyte

Abb. 3: Verfügbare Ressourcen für die verschiedenen Programmiermodelle
(zweite Installationsstufe in Klammern)

Das RZ betreibt gegenwärtig keinen Linux-basierten Parallelrechner, da größere Systeme von anderen Hochschulen in den Verbund eingebracht werden. Das RZ hat sich entschlossen, solche Systeme wegen des nicht unerheblichen Arbeitsanfalls nicht selbst zu betreiben, sondern auf die Verfügbarkeit solcher Systeme im Verbund zu warten.

4.2 Archivserver

Notwendig für den effizienten Betrieb von großen HPC-Systemen sind hochvolumige Archivserver zur längerfristigen Datenhaltung. Da von den Nutzern enorme Datenmengen erzeugt werden, die u.U. von den lokalen Rechenzentren in NRW nicht gehandhabt werden können, wird die RWTH einen solchen Archivdienst auch als Landesdienst anbieten.

In der RWTH wird für Backup- und Archivdienste vom RZ die Software Tivoli Storage Manager (TSM) seit längerer Zeit erfolgreich eingesetzt. Hiermit ist es möglich, mittels einer Client-Komponente (GUI) von beliebigen Systemen Dateien auch zu archivieren. Durch den landesweiten Zugriff auf den Archivdienst wird so ergänzend zu den an den Hochschulen vorhandenen Archivdiensten auch ein Kooperationsmedium für verteilte Benutzergruppen geschaffen.

Installiert wird gegenwärtig ein Bandarchiv der Firma IBM. Das Gerät hat ein Bandvolumen von über einem Pbyte (komprimiert), 16 Laufwerke der neuesten Generation und einen Server mit sehr hoher interner Bandbreite und einem extrem großen Plattencache von 30 Tbyte. Der TSM-Server ist über Gbit-Ethernet direkt an das G-WiN angebunden.

In Zusammenarbeit mit den Universitäten in Essen und Münster wurden mittels TSM auch verteilte Sicherungsstrategien für bandgesicherte Daten erprobt. Im Dreieck Aachen-Essen-Münster wurden inkrementell die gesicherten Daten und die TSM-Datenbank an unterschiedlichen Orten gespeichert. Dies schützt die Vertraulichkeit der Daten, da die Restauration der Benutzerdaten nur mit allen Teilen möglich ist.

Hier zeigte sich, dass die ursprüngliche verkehrsoptimierte Routingstrategie des G-WiN über Knoten in München oder Hamburg zu Latenzen führte, welche den Dienst erheblich beeinflussten bzw. sogar unmöglich machten. Der DFN-Verein hat sich diesen Landeskooperationsanstrengungen gegenüber sehr aufgeschlossen gezeigt und in der Zwischenzeit das Routing geändert. Gleichwohl zeigt diese Erfahrung, dass kooperative IT-Anstrengungen zu Netzwerksqualitätsanforderungen führen, die über "best effort" Ansätze weit hinausgehen.

4.3 Beratungs- und Schulungsleistung

Verbundweit wird Beratungsleistung über einen Web-basierten Supportserver („CallManager“) erbracht. Auch Schulungen im HPC-Bereich sowie Unterstützung bei HPC-Projekten (z.T. auch Kooperationen) stehen den Nutzern NRW-weit zur Verfügung.

4.4 Collaborative Product Commerce (CPC-System)

Neuland betritt die RWTH mit dem Betrieb eines landesweit zur Verfügung stehenden Dienstes aus dem Bereich des konstruktiven Maschinenbaus. In einem DFG-geförderten Gemeinschaftsprojekt des RWTH-Institutes für konstruktiven Maschinenbau (IKT), des Softwareherstellers PTC und des RZ steht ein Server mit dem Softwareprodukt Windchill aus dem Bereich Collaborative Product Commerce zur Verfügung. An dem Projekt nehmen Lehrstühle aus mehreren Hochschulen und Fachhochschulen des Landes als Nutzer teil. Windchill wird von den teilnehmenden Lehrstühlen mit Nachdruck in der Lehre eingesetzt werden, um den Studenten innovative verteilte Arbeitsformen im Bereich Maschinenbau zu vermitteln.

Das Windchill-System ermöglicht ein Web-basiertes, verteiltes Konstruieren von komplexen technischen Systemen und stellt eine Umgebung zur Durchführung des dazu notwendige Projekt- und Produktdatenmanagements zur Verfügung. Die Nutzung der Software erfolgt über einen lokal installierten Web-Clients der Zugriff auf den Windchill-Server im RZ hat. Alle Daten werden vom Server in einem hochverfügbar ausgelegten Oracle-Server abgelegt.

Zur Verfügung steht ein Windchill-System mit den Komponenten (Solution Links):

- Projektmanagement
- elektronischer Bauteilkatalog
- Produktkonfigurator
- Produktdatenmanagement

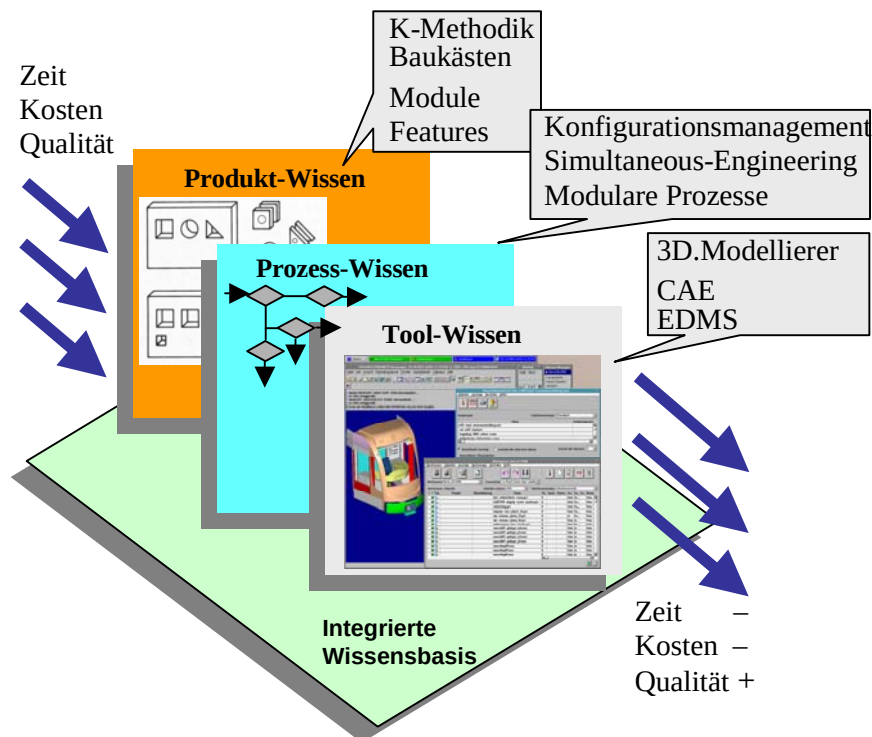


Abb. 4: Komponenten der CPC-Software Windchill

Da die interaktiven Zugriffe der Klienten auf den Server und die darunter liegende Oracle-Datenbank über das G-WiN erfolgen, sind die Latenzzeiten des WAN und der lokalen Netzwerke in den Instituten kritisch. Unabhängig von der Lage der Hochschule im Lande sind gravierende Sprünge in der Latenz unakzeptabel. Wie schon bei dem Datendienst im Dreieck Aachen-Essen-Münster wird die durchgängig erforderliche Netzqualität zu einer vertieften Zusammenarbeit der beteiligten Hochschulrechenzentren mit dem DFN führen müssen.

4.5 Virtual Reality

Für viele Bereiche des Hochleistungsrechnens ist die geeignete Darstellung der Ergebnisse nicht trivial. Im Jahr 2000 wurde daher das Virtual Reality Center Aachen (VRCA) als Forschungsverbund und Informationsnetzwerk gegründet. Derzeit sind 34 Lehrstühle und Institute aus unterschiedlichen Fachbereichen sowie 6 Unternehmen im VRCA organisiert. Insbesondere das koordinierte Vorgehen in der Lehre über Fakultätsgrenzen hinweg sowie die gemeinsame, vom RZ koordinierte Entwicklung von für den wissenschaftlich-technischen Anwendungsbereich zugeschnittener VR-Software durch Institute mehrerer Fachrichtungen zeichnet das VRCA aus und scheint an einer deutschen Hochschule bislang einzigartig zu sein.

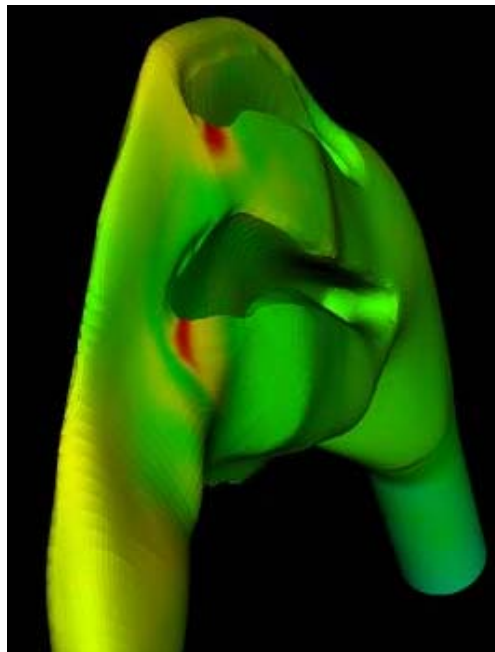


Abb. 5: VR-gestützte Analyse der Strömung in einer menschlichen Nase

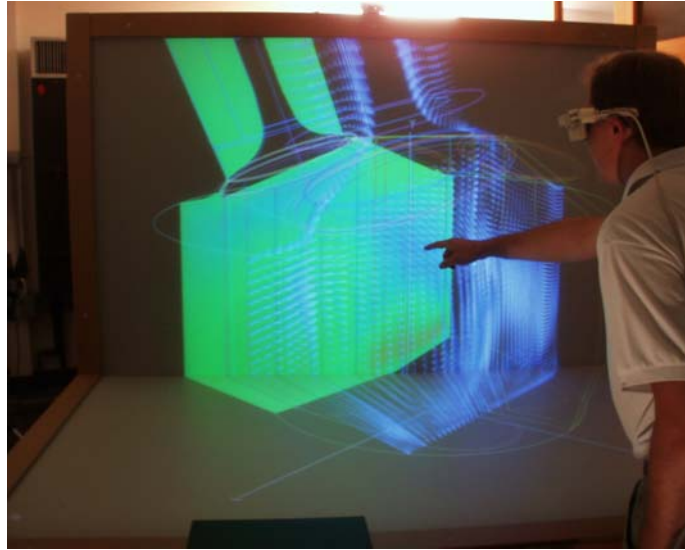


Abb. 6: VR-gestützte Analyse einer Motorinnenströmung auf einer HoloBench
(Aerodynamisches Institut und RZ der RWTH Aachen)

Das RZ führt eine Vielzahl von interdisziplinären Forschungs- und Entwicklungsprojekten mit den Instituten der RWTH, der Industrie, aber auch Projektteilnehmern aus den anderen Hochschulen in NRW durch (siehe auch unter: www.rwth-aachen.de/vrca).

Dazu steht eine Vielzahl von unterschiedlichen Projektionssystemen zur Verfügung. Diese „Display-Pyramide“ reicht vom Monitor über Großbildprojektionen und Tischprojektionen („WorkBench“, „HoloBench“) bis hin zu in der Beschaffung befindlichen CAVE- und Großbild-Rückprojektionssystemen mit sehr hoher Bildauflösung zur Darstellung größerer Objekte und Prozesse im Maßstab 1:1. Die Installation der Projektionssysteme im Rechen- und Kommunikationszentrum ermöglicht eine enge Kopplung zwischen High Performance Computing und Virtual Reality. Ebenfalls von besonderem Interesse sind die Erhöhung des immersiven Charakters durch die Nutzung von Interaktionssystemen.

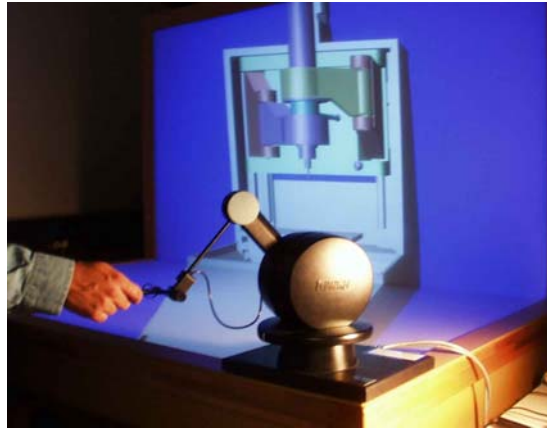


Abb. 7: Kraftrückkopplung bei der Interaktion mit virtuellen Prototypen von Werkzeugmaschinen (Lehrstuhl für Werkzeugmaschinen und RZ der RWTH Aachen)

5. Fazit

Die Bedeutung der IT-Infrastruktur für eine Hochschule erfordert performante IT-Dienste wobei die fiskalischen Randbedingungen allerdings weitere Kostenreduktionen erzwingen. Vor diesem Hintergrund sind kooperative Anstrengungen eine Möglichkeit, qualitativ hochwertige Dienste auch weiterhin anzubieten und know-how und Flexibilität in den Hochschulen zu halten. Die Aktivitäten der RWTH Aachen in den Bereichen HPC, Archiv usw. im Konzert mit den anderen Hochschulen sind erste, wichtige Schritte in diese Richtung.

Referenzen

- [1] Ministerium für Wissenschaft und Forschung NRW (MWF), *High Performance Computing in Nordrhein-Westfalen – Kooperatives Versorgungskonzept für das Hoch- und Höchstleistungsrechnen in den Hochschulen des Landes (HIPEC NRW)*, Jan. 2002, www.mwf.nrw.de/hochschulen_in_nrw/verwaltung/dv_netze_hochschulen/HIPEC-NRW_Konzept2002.pdf
- [2] Arbeitskreis der Leiter der wiss. Rechenzentren in NRW, *Rechnerverbund NRW* Dez. 2000, www.rv-nrw.de/RV-Doku/Abschlussbericht.pdf

Auswirkungen der Virtualisierung auf Transparenz und Fehlerdiagnose in lokalen Netzen

Andreas Binzenhöfer, Kurt Tutschku

Lehrstuhl für Informatik III, Universität Würzburg
Am Hubland
97074 Würzburg
Germany

E-mail: {binzenhoefer,tutschku}@informatik.uni-wuerzburg.de

Die Virtualisierung im IT-Betrieb zielt darauf ab, vorhandene IT-Ressourcen zu einem gemeinsamen Pool zusammenzufassen und dabei den mit der Verwaltung der Ressourcen verbundenen Aufwand, sowohl administrativer als auch technischer Art, zu reduzieren und so insgesamt Kosten einzusparen. In unserem Beitrag zeigen wir die hiermit verbundenen Probleme und Gefahren auf und stellen einen ersten Lösungsansatz vor. Anschließend geben wir einen Überblick über bereits vorhandene Möglichkeiten, verdeutlicht am Beispiel des Freeware Pakets Nagios. Die Ausarbeitung wird schließlich mit einem kurzen Blick in die Zukunft abgerundet.

1 Einleitung

In den letzten zwei Jahrzehnten hat sich die IT-Landschaft rasant verändert. Verteilte Computersysteme werden immer komplizierter und komplexer. Der Trend geht eindeutig in Richtung immer leistungsfähigerer Komponenten. Große Teile der Forschung konzentrieren sich zudem darauf, bestehende Komponenten bis an ihre Grenzen hin optimal auszunutzen. Als Folge verlieren schon heute zahlreiche Firmen die Übersicht über ihre internen Firmennetze. Die Komplexität der hochgradig vernetzten IT-Komponenten ist nur noch schwer überschaubar und von menschlichen Administratoren beinahe nicht mehr in den Griff zu bekommen. Das Zusammenspiel von Netzwerkkomponenten verschiedener Hersteller sowie fehlende Standards verstärken diese Effekte zusätzlich. Als direkte Folge sind Ursachen für Systemausfälle und sonstige Fehler im Betrieb nur schwer zu lokalisieren und erinnern an die berühmte Nadel im Heuhaufen.

Im Folgenden wollen wir die Virtualisierung von IT-Ressourcen als eine der zurzeit wohl meist diskutierten Gegenmaßnahmen kurz aus unserer Sicht definieren und vorstellen. Der hier vorliegende Beitrag versteht sich als Konzeptstudie und möchte die offenen Fragen in Zusammenhang mit einer

verstärkten Virtualisierung in Rechnersystemen und Rechnernetzen diskutieren. Neben den bekannten Vorteilen einer Virtualisierung möchten wir vor allem aber die Nachteile und Gefahren aufzeigen. Um diese Gesichtspunkte zu verdeutlichen, skizzieren wir dazu die Problemsituation in aktuellen Netzen am Beispiel einer Direktion eines großen deutschen Versicherungsunternehmens (ca. 14000 Anwender unternehmensweit und ca. 150 Einzelplatz- und Serverrechner pro Direktion).

Als ersten möglichen Lösungsansatz umreißen wir kurz das neue Forschungsgebiet „Autonomic Networks“ [5] am Lehrstuhl III für Informatik der Universität Würzburg. Abschließend geben wir einen Überblick über bereits vorhandene Möglichkeiten, die Transparenz von abstrakten IT-Systemen aufrecht zu erhalten, verdeutlicht am Beispiel des Freeware Pakets Nagios [6].

2 Auswirkungen der Virtualisierung

Der Begriff Virtualisierung von IT-Ressourcen scheint derzeit allgegenwärtig zu sein. Er dient gleichermaßen als Thema für zahlreiche wissenschaftliche Veröffentlichungen und als Zugpferd großer IT-Firmen wie IBM, HP und SUN. Bisher gibt es jedoch keine zufrieden stellende einheitliche und allgemein akzeptierte Definition des Begriffes der Virtualisierung. Wir bemühen uns daher im nächsten Abschnitt um eine Definition der Virtualisierung von IT-Ressourcen. Weiterhin wird erläutert, was sich hinter dem Begriff der Netzwerk-Transparenz verbirgt und wie diese mit der Virtualisierung in Beziehung gebracht werden kann.

2.1 Definition Virtualisierung und Transparenz

Virtualisierung ist dabei die Reduktion des mit der Verwaltung der Ressourcen verbundenen Aufwands sowie die Senkung der globalen Betriebskosten. Im Zuge der Virtualisierung wird eine weitere Abstraktionsebene in das bestehende IT-System eingefügt. Die neue virtuelle Komponente erscheint dem Benutzer als eigenständige, leistungsfähigere Einheit (siehe Abbildung 1).



Abb. 1: IT-Ressourcen werden zu einem virtuellen Massenspeicher zusammengefasst

Die Virtualisierung der Ressourcen ist eng verbunden mit dem so genannten Grid Computing, bei dem die Ressourcen mehrerer vernetzter Computer zur gleichen Zeit zur Lösung eines bestimmten Problems eingesetzt werden. Diese Technik findet ihre Anwendung meist in der Forschung, wie z.B. bei der verteilten Suche nach Außerirdischen im Projekt *seti@home* der Universität Berkeley. Auch zahlreiche kommerzielle Firmen wie HP, IBM und SUN bieten bereits erste Lösungen für auf mehrere Computer verteilte Massenspeicher an. Die virtuellen Speichersysteme erscheinen dem Administrator dabei als einheitlicher Massenspeicher, der in etwa mit einer virtuellen Festplatte vergleichbar ist. Die Speicherkapazität lässt sich von zentraler Stelle aus verwalten und z.B. verschiedenen Servern zuweisen. Schon durch diese gemeinsame Nutzung der Speicherressourcen lässt sich der physikalische Speicher wesentlich effizienter nutzen. Alles in allem beschreibt die Virtualisierung also die Emulation einer physikalischen Komponente mittels verschiedener Architekturen, Produkte und Geräte.

Im Gegensatz zur Virtualisierung zielt die Transparenz darauf ab, zu jeder Zeit die direkte Sicht auf die einzelnen Komponenten und deren Zustand zu gewährleisten. Das Ziel der transparenten Sicht auf das lokale Netzwerk besteht darin, die wichtigen Leistungsmerkmale mit angemessenem Aufwand überprüfen zu können und im Bedarfsfall schnell an der richtigen Stelle reagieren zu können. Ein transparentes Netzwerk kennt die Antworten auf Fragen der Art: „Welche Ressourcen sind zurzeit überlastet und stellen somit den Engpass dar?“, „Welche Komponenten müssen aufgerüstet bzw. erneuert werden?“ und „Wo entstehen die eigentlichen Kosten in meinem Netzwerk?“. Um weiterhin die Übersicht in komplexen IT-Systemen zu gewährleisten, werden meist verschiedene Sichten auf die gleichen Daten zur Verfügung gestellt. So benötigt ein Vorstandsmitglied z.B. andere Merkmale zur Entscheidungsfindung, als etwa ein Systemadministrator. Mit der zusätzlichen Abstraktionsebene der Virtualisierung läuft man nun aber Gefahr, die Transparenz des eigenen Netzwerkes zu riskieren.

Aufgrund der oben angeführten Argumente kommen wir zu dem Schluss, dass ein Gleichgewicht zwischen Virtualisierung und Transparenz in lokalen Netzwerken erreicht werden muss, sodass eine effiziente Fehlerdiagnose ermöglicht wird. Das bedeutet, die wichtigen Leistungsmerkmale müssen vom Anwender beobachtbar bleiben, während die Virtualisierung weiterhin die einfache Nutzung der Ressourcen gewährleistet. Somit erscheint es unabdingbar, dass für komplexe IT-Systeme im Zuge der Virtualisierung parallel ein unabhängiges Überwachungs- und Diagnosesystem entwickelt wird. Zunächst erscheint es aber sinnvoll, die Vor- und Nachteile einer Virtualisierung im IT-Bereich gegeneinander abzuwägen.

2.2 Positive und negative Auswirkungen der Virtualisierung

Die positiven Auswirkungen der Virtualisierung auf den IT-Bereich liegen auf der Hand. Gerade in kommerziellen Systemen geht es zunächst primär darum, Kosten einzusparen und den Return on Invest zu verbessern. Dies wird hauptsächlich durch die effizientere Ausnutzung vorhandener Ressourcen erreicht. Während bei einer herkömmlichen Server-Architektur für gewöhnlich mindestens 15% des Festplattenspeichers als Sicherheitsreserve einkalkuliert werden, lassen sich zusätzliche Festplatten und Speichersysteme ohne Betriebsunterbrechung in einen virtuellen Speicherpool integrieren. Durch die Möglichkeit, im laufenden Betrieb unbemerkt in das System einzugreifen, werden bisher diffizile Datenspiegelungen, Datenduplizierungen und vor allem Datensicherungen deutlich vereinfacht.

Diese Ressourcenaustauschbarkeit garantiert zudem eine erhöhte Flexibilität und Verfügbarkeit. Weiterhin ermöglicht die Virtualisierung den Zusammenschluss einer heterogenen Serverlandschaft und verhindert somit die Abhängigkeit von einzelnen Herstellern. Unternehmen erhalten hierdurch einen größeren Spielraum und mehr Freiheit in der Gestaltung ihrer Hardware-Ausstattung.

Auf der anderen Seite darf man allerdings nicht die möglichen Gefahren und Nachteile, die mit der Einführung einer weiteren Abstraktionsebene unweigerlich einhergehen, vernachlässigen. Der Grund, der bisher am stärksten gegen eine Virtualisierung von IT-Ressourcen spricht, ist das Fehlen allgemein akzeptierter Standards in diesem Bereich. Die Open Grid Service Architecture (OGSA) stellt einen ersten Schritt in die richtige Richtung dar, von einem marktreifen Konzept ist man allerdings noch weit entfernt. Erste Erfahrungen in lokalen Netzwerken zeigen zudem, dass die Virtualisierung ab einer gewissen Größenordnung nicht unbedingt die angestrebte Skalierbarkeit mit sich bringt. Alleine die Organisation und Einteilung der einzelnen Ressourcen birgt einen nicht zu vernachlässigenden Overhead in Form eines zusätzlichen Verwaltungsaufwands in sich. Im Falle kritischer Anwendungen und sensibler Daten kommen zusätzlich sicherheitsrelevante Aspekte mit ins Spiel. So stellt etwa die gemeinsame Nutzung einer gemeinsamen Ressource durch zwei konkurrierende Prozesse ein weiteres Sicherheitsrisiko dar.

Die aus unserer Sicht größte Gefahr einer weiteren Abstraktion durch Virtualisierung liegt jedoch in dem Verlust der Transparenz in lokalen Netzwerken. Eine virtualisierte Komponente verhält sich zunächst für den Benutzer wie eine Art Black Box, über die er nur beschränkte Kontrolle besitzt. Fehlerdiagnose und Lokalisierung von Schwachstellen und Ursachen werden hierdurch zusätzlich erschwert. Die Virtualisierung der Ressourcen reduziert nun also als Nebeneffekt die direkte Sicht auf die einzelnen Komponenten und deren Zustand. Aus diesem Grund ist es notwendig, einen Kompromiss zwischen Virtualisierung, im Sinn der Vereinfachung der Ressourcenaustauschbarkeit und

Ressourcenverwaltung, und der Transparenz der Systemzustände zu finden. Das bedeutet, die wichtigen Leistungsmerkmale bleiben vom Anwender beobachtbar, während die Virtualisierung weiterhin die einfache Nutzung der Ressourcen gewährleistet. Im nächsten Abschnitt umreißen wir daher kurz die Problemsituation in aktuellen Netzen, um dieses Argument am konkreten Beispiel zu verdeutlichen.

2.3 Problemsituation in aktuellen Netzen

Bestehende Netzwerke enthalten, selbst wenn sie eine ausgeprägte Client-Server Architektur aufweisen, zahlreiche Fehlerquellen und sind von zentraler Stelle aus nur schwer zu überschauen. Die Einführung einer weiteren Abstraktionsebene durch die Virtualisierung einer gewissen Anzahl von Einzelressourcen maskiert nun diese Fehlerquellen und fügt damit eine weitere Sichtbarriere bei der Fehlerdiagnose in lokalen Netzwerken ein.

Um die aktuelle Problemsituation zu dokumentieren und in die Entwicklung neuer Administrationsverfahren mit einbeziehen zu können, wurden in einer Direktion eines großen deutschen Versicherungsunternehmens (ca. 14000 Anwender mit ca. 150 Einzelplatz- und Serverrechnern pro Direktion) Messungen durchgeführt und Logfiles ausgewertet. Dabei wurde festgestellt, dass sich typische Fehler und Leistungsengpässe in heutigen lokalen Netzen durch unpräzise Fehlersymptome auszeichnen. Diese werden weiterhin vom Benutzer meist nur vage zum Beispiel als ein „nicht reagierendes Netzlaufwerk“, eine „langsame Web-Site“, oder eine „geringe Dateiübertragungsleistung“ beschrieben. Die ungenaue Beschreibung erschwert die Diagnose und Behebung der Fehler. Bei genauer Betrachtung stellen sich diese Fehler häufig jedoch als ein Konglomerat verschiedener einzelner Ursachen dar. Alleine das Zusammenspiel von Netzwerkkomponenten verschiedener Hersteller kann unvorhersehbare Folgen auf das Leistungsniveau eines IT-Systems haben.

Eine der Hauptursachen für Probleme in lokalen Netzwerken findet sich in den zahlreichen Einstellungsmöglichkeiten und Konfigurationsdateien, die über das komplette Netzwerk verteilt liegen und nicht automatisch verwaltet werden. Überlegungen und Ideen hinter der aktuellen Konfiguration sind dem System meist nicht bekannt und spiegeln sich ausschließlich in den Köpfen weniger Mitarbeiter wider. Darüber hinaus führt das System keine eigenständigen Plausibilitätschecks der aktuellen Konfiguration durch. Ein DNS-Server in einem nicht erreichbaren Subnetz fällt somit erst im laufenden Betrieb auf. Die bisherige Fehlerdiagnose findet häufig redundant und dezentral statt. Mehrere Administratoren suchen parallel die Ursache für ein und dasselbe Problem. In unserem Forschungsgebiet „Autonomic Networks“ arbeiten wir daher an Lösungen, die die Fehlerdiagnose und -behebung automatisieren, Netzwerke an

den richtigen Stellen transparent halten und sich mit der Virtualisierung von IT-Ressourcen in Einklang bringen lassen.

3 Unsere Vision: Autonomic Networks

Das Forschungsgebiet Autonomic Networks [5] beschäftigt sich mit der Entwicklung selbstorganisierender Algorithmen. Das erklärte Ziel sind Computernetzwerke, die sich selbstständig konfigurieren, verwalten und reparieren.

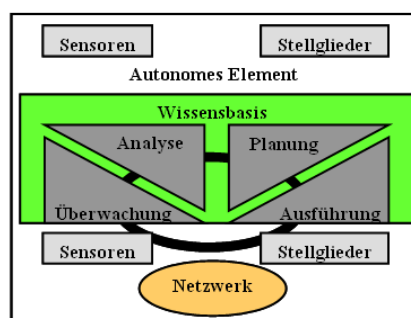


Abb. 2: Ein Autonomes Netzwerk-Element nach dem Vorbild von IBM

Abbildung 2 zeigt den Regelkreis einer autonomen Netzwerkkomponente nach dem Vorbild der Firma IBM [4]. Jede Netzwerkkomponente ist mit Sensoren ausgestattet, über die von einer Überwachungseinheit aktuelle Systemzustände erfragt werden können. Die Analyseeinheit interpretiert die gewonnenen Daten und vergleicht die Ergebnisse mit einer Wissensbasis. Im gegebenen Fall wird die Planungseinheit verständigt, die eine entsprechende Lösung erarbeitet. Die nötigen Gegenmaßnahmen werden letztendlich von der Ausführungseinheit über die Stellglieder der Netzwerkkomponente eingeleitet. Dabei ist zu beachten, dass es sich nicht um einen strengen Regelkreis handelt, d.h. die einzelnen Komponenten können bei Bedarf z.B. neue Messdaten von der Überwachungseinheit anfordern.

Durch diese autonome Verwaltung der Netzwerkkomponenten wird die Ausfallsicherheit des gesamten Systems verbessert. Administratoren erhalten im Optimalfall ausschließlich Fehlermeldungen, die einen manuellen Eingriff zwingend erfordern, wie etwa im Falle einer defekten Hardwarekomponente. Im Hinblick auf die Virtualisierung lassen sich die gefundenen Ergebnisse zur Selbstverwaltung des virtuellen Ressourcenpools einsetzen. Da sich Ausfälle von Komponenten und Softwarefehler nicht gänzlich vermeiden lassen, kommen zusätzlich Algorithmen wie etwa das Tool Pinpoint der Universität Berkely [3]

zum Einsatz, um Fehlerursachen genauer lokalisieren zu können. Im nächsten Abschnitt geben wir am Beispiel Nagios [6] einen kurzen Überblick über die aktuellen Möglichkeiten zur transparenten Fehlerdiagnose.

4 Der erste Schritt in Richtung Autonomic Networks

Die Fehlerdiagnose in aktuellen Netzwerken geschieht im Normalfall mit Hilfe von Softwarepaketen wie z.B. Tivoli oder HP Openview. All diesen Tools ist gemeinsam, dass sie sich auf messbare Daten stützen müssen. Angefangen mit einfachen pings, über snmp-Anfragen bis hin zu port-Tests wird alles ausgenutzt, was sich aktiv überprüfen lässt. In diesem Abschnitt möchten wir am Beispiel des Freeware-Pakets Nagios aufzeigen, was mit den herkömmlichen Mitteln möglich ist und wie sich die Fehlerdiagnose in Richtung Selbstverwaltung ausdehnen lässt.

Wie in Abbildung 3 zu sehen, stellt Nagios das lokale Netzwerk übersichtlich auf einen Blick dar. Problemzonen werden dabei rot gekennzeichnet. Neben den üblichen ping-Anfragen, die überprüfen, ob ein Computer noch ansprechbar ist, besteht die Möglichkeit, aktiv Services auf externen Rechnern zu überprüfen. Nagios sendet hierzu z.B. eine http-Anfrage an den dedizierten Webserver und bereitet die Antwort grafisch für den Administrator auf. In Fehlersituationen (z.B. Papierstau im lokalen Drucker) benachrichtigt der Nagios-Prozess die zuständige Kontakt-Gruppe per eMail, ICQ oder SMS. Externe Messgrößen, wie z.B. der belegte Plattenplatz, können über einen lokal installierten Daemon, der periodisch aktiviert wird, abgefragt und an den Nagios-Server übermittelt werden. Dieses Beispiel beschreibt, wie Nagios das so genannte „Push“-Konzept umsetzt. Beim „Push“-Konzept werden die Daten nicht erst auf Anfrage des Managementsystems übermittelt. Das überwachte System führt vielmehr eine Selbstdiagnose durch, sammelt die Daten, bereitet sie auf und übermittelt sie selbstständig an die Managementstation. Das „Push“-Konzept steht somit im Kontrast zum „Pull“-Konzept, das in den heute üblichen Management-Architekturen, wie zum Beispiel SNMP, verwendet wird. Da beim „Push“-Modus keine Anfragen versendet werden, wird der Kommunikationsaufwand im Netz zusätzlich reduziert. Die Vorverarbeitung der Daten vermindert die zusätzliche Belastung der Managementstation und erlaubt eine genauere Analyse der Daten. Das „Push“-Konzept reduziert somit die Fehleranfälligkeit der Managementstation und erhöht ihre Leistungsfähigkeit.

Der Hauptvorteil des Open Source Projektes Nagios liegt nun in seiner Konfigurierbarkeit und Erweiterbarkeit. Über externe Skripten lässt sich die Fehlerdiagnose an individuelle Bedürfnisse anpassen. Auf diese Art lassen sich gezielt die für das jeweilige Netzwerk entscheidenden Größen überwachen.

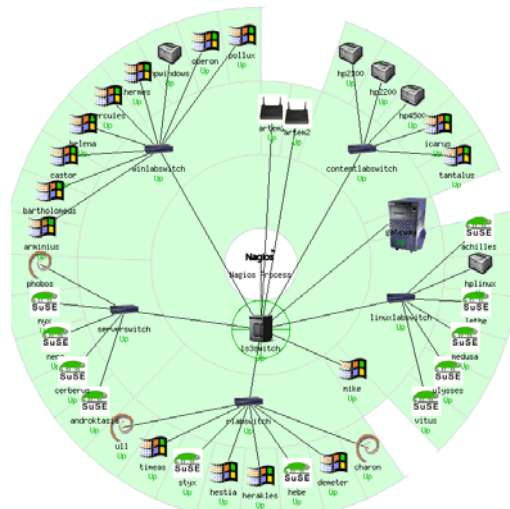


Abb. 3: Das Lehrstuhl-Netz in der übersichtlichen Nagios Statusmap

Weiterhin erlaubt eine einfache Benutzerverwaltung verschiedene Sichten auf das Netzwerk. Administratoren können sich somit gezielt auf den für sie interessanten Bereich konzentrieren, wodurch eine effizientere Fehlerdiagnose ermöglicht wird.

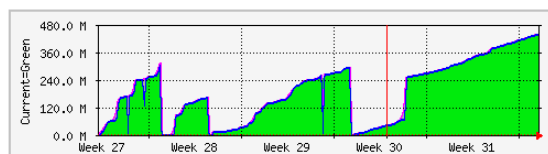


Abb. 4: Zahlreiche Statistiken erzeugen die gewünschte Transparenz

Durch systematische Messungen lassen sich außerdem umfangreiche Statistiken (siehe z.B. Abbildung 4) erzeugen, die das lokale Netzwerk an den gewünschten Stellen transparent gestalten.

Die so genannten Service-Escalations stellen den ersten Schritt in Richtung „Autonomic Networks“ dar. Hier lassen sich automatische Gegenmaßnahmen für mögliche Fehlerszenarien im Vorhinein festlegen. Eine mögliche Anwendung wäre etwa, den Webserver im Fehlerfall ohne menschlichen Eingriff automatisch neu zu starten. Eine Meldung an den zuständigen Kontakt findet erst statt, falls der Neustart das ursprüngliche Problem nicht beheben konnte. Wie aber lässt sich dieses Überwachungs- und Diagnosesystem nun zur Lösung des Konfliktes zwischen der Virtualisierung und der Transparenz in lokalen Netzwerken

einsetzen? Mit Hilfe der oben erwähnten Verfahren besteht die Möglichkeit einzelne Teil-Ressourcen eines virtualisierten Ressourcenpools individuell zu überwachen. Je nach Abstraktionsebene und gewünschter Transparenz, kann ein Fehler in einer Teilkomponente als Ausfall der gesamten virtuellen Black Box oder transparent mit detaillierter Beschreibung gemeldet werden. Mit Hilfe verschiedener Sichten auf das Netzwerk, lassen sich also die Vorteile einer Virtualisierung ausnutzen, ohne dabei die Kenntnis über die internen Systemzustände zu verlieren.

5 Ein Blick in die Zukunft

Abschließend lässt sich zusammenfassen, dass die Virtualisierung von IT-Ressourcen ein sinnvoller Schritt ist, um Benutzern unnötig wiederkehrende Aufgaben abzunehmen, Systeme effizienter ausnutzen und Kosten einzusparen. Im Gegenzug droht die Gefahr, die Transparenz, in dem Sinn, dass die wichtigen Leistungsmerkmale beobachtbar bleiben, zu verlieren.

Aktuelle Fehlerdiagnosesysteme verstehen es bereits sehr gut, die zur Verfügung stehenden Messdaten effizient auszuwerten. Um jedoch auch in Zeiten der Virtualisierung eine transparente Selbstverwaltung zu ermöglichen, sind neue Messtechniken erforderlich. Zum einen werden Informationen benötigt, die mit bisherigen Mitteln nicht erfasst werden können, angefangen mit der Temperatur einzelner kritischer Hardwarekomponenten bis hin zu quantitativen Aussagen über die Qualität verlegter Kabel. Zum anderen müssen überflüssige manuelle Prozesse, wie etwa eine redundante Konfiguration ähnlicher Komponenten, automatisiert werden. Mit den entsprechenden Grundvoraussetzungen und dem nötigen Weitblick kann eine transparente Virtualisierung also Realität werden.

Referenzen

- [1] D. Ruzicka, *Network Management with Nagios, Netsaint's Successor: What's Going On?*, Linux Magazine, Ausgabe 29, Seite 62 (April 2003)
- [2] R. C. Harlan, *Aus der Praxis: Netzwerkmanagement, Wer sucht, der findet*, Linux Administrator, Ausgabe Juli 2003, Seite 1 (Juli 2003)
- [3] D. A. Patterson et al, *Recovery-Oriented Computing (ROC): Motivation, Definition, Techniques, and Case Studies*, UC Berkeley Computer Science Technical Report UCB//CSD-02-1175 (März 2002)

- [4] IBM, *An architectural blueprint for autonomic computing*,
<http://www-3.ibm.com/autonomic/r/downloads/blueprint/index.html> (April 2003)
- [5] A. Binzenhöfer, K. Tutschku, *Autonomic Networks*,
<http://www3.informatik.uni-wuerzburg.de/research/autonomic.shtml>
- [6] Nagios, <http://www.nagios.org>

Anwendung von NAS-Systemen im Deutschen Wetterdienst

Alexander Harth

Deutscher Wetterdienst
Kaiserleistr. 42
63067 Offenbach
E-mail : Alexander.Harth@dwd.de

Der Deutsche Wetterdienst konsolidiert Daten und Backup auf wenige Systeme in einer Network Attached Storage (NAS) Infrastruktur. Diese über ganz Deutschland verteilte Struktur wird als vom jeweiligen Rechnerbetriebssystem unabhängige Datenablage sowohl für Büro- und Fachanwendungen, als auch für spezielle, hoch verfügbare Applikationen genutzt. Das Gesamtsystem wird zentral administriert und überwacht. IT-Personal vor Ort ist in der Regel nicht notwendig. Damit hat der DWD eine einheitliche, flexible, erweiterbare und sichere Infrastruktur geschaffen, die eine zukunftsichere Basis für die Computer gestützten Arbeiten bietet.

1 Aufgabe und Infrastruktur des Deutschen Wetterdienstes

Der Deutsche Wetterdienst (DWD) ist eine Oberbehörde des Bundesverkehrsministeriums. Er erbringt verschiedenste meteorologische Dienstleistungen, insbesondere Wettervorhersagen mit Beratung und den sich daraus ergebenden Warnungen. Weitere Kernaufgaben sind der Betrieb eines meteorologischen Messnetzes und die meteorologische Sicherung der Straßen-, Luft- und Seewege. Darüber hinaus stellt der DWD meteorologische Daten und Archive für Forschung und Allgemeinheit bereit. Der DWD hat den gesetzlichen Auftrag, die Atmosphäre auf Radioaktivität zu überwachen und Ausbreitungsprognosen zu erstellen. Als interner Dienstleister betreibt der DWD das Weitverkehrsnetz der Bundesverkehrsverwaltung. Diese Dienstleistungen erbringen ca. 2700 Mitarbeiterinnen und Mitarbeiter an etwa 150 Standorten, die über ganz Deutschland verteilt sind.

Dazu nutzt der DWD eine Vielzahl von Rechnersystemen in einem Netzwerk, das über ganz Deutschland ausgedehnt ist (siehe Abb. 1). Diese Rechner sind entsprechend ihren Aufgaben in drei Ebenen kategorisiert.

Die Ebene 1 stellt die zentrale Versorgung mit Daten und Rechenleistung bereit. Hier betreibt der DWD eine IBM SP mit 120 Knoten und 1920 Prozessoren als zentralen Hochleistungsrechner (HLR), dieses System in seiner vorherigen

Ausbaustufe mit 1280 Prozessoren belegt auf der Top500-Liste der Supercomputer [1] derzeit Platz 35 mit 1404 GFlops. Zur Archivierung und weiteren Verarbeitung der von diesem HLR bereitgestellten Daten sowie zum Preprocessing der Rohdaten setzt der DWD mehrere Datenbank- und Routineserver [2] mit derzeit 12TB-online- und 400TB-offline-Kapazität ein.

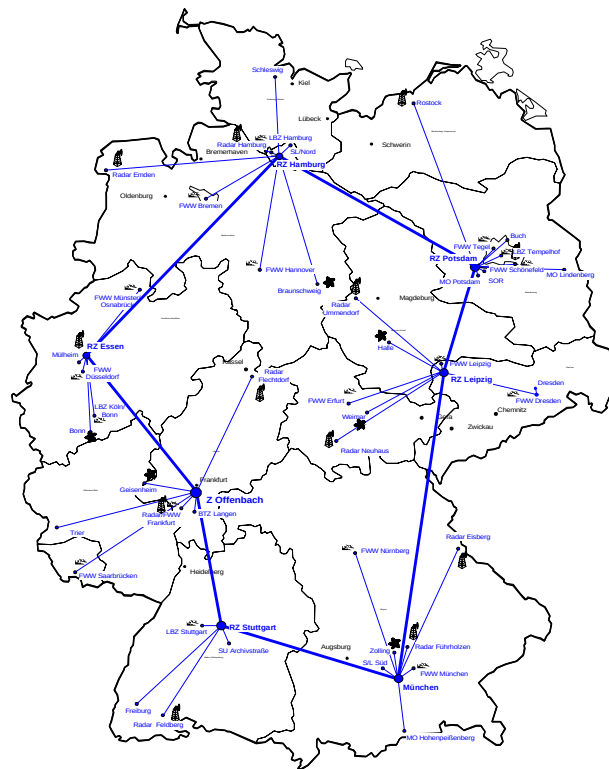


Abb. 1 Das Weitverkehrsnetz des DWD

Die dezentrale Versorgung [3] mit IT leisten die Rechner der Ebenen 2 (220 Server) und 3 (3100 Workstations, PC und Notebooks). Diese Systeme versorgen sowohl die Nutzer im DWD als auch die externen Kunden mit Daten und lokaler Grafik- und Rechenleistung. Auf diesen Rechnern werden die Betriebssysteme IRIX, Linux, HP-UX und MS-Windows NT eingesetzt. In den folgenden Kapiteln möchte ich die verschiedenen Anwendungen von NAS-Systemen darstellen, die in den Ebenen 2 und 3 Verwendung finden.

2 Die Datenversorgung dezentraler Systeme in der Vergangenheit

Der DWD baute die bestehende Infrastruktur der Ebenen 2 und 3 im Wesentlichen in den Jahren 1999 und 2000 auf. Er wählte einen für diese Zeit typischen Ansatz. Jeder größere Standort erhielt ein Serverensemble aus zwei Servern unter dem Betriebssystem IRIX und drei Servern unter MS-Windows NT. Server mit RAID-Systemen hielten die Daten und exportierten sie an die Clients. Zur Erhöhung der Verfügbarkeit war eines der RAID-Systeme an beiden IRIX-Servern angeschlossen, um bei Systemausfällen und Wartungsarbeiten den Betrieb aufrechterhalten zu können. Die Datensicherung erfolgte jeweils auf lokale Tapelibraries, die sowohl lokal als auch zentral angesteuert wurden. Die kleineren Standorte wurden jeweils mit einem Server unter MS-Windows NT ausgestattet, der sowohl die Datenablage als auch die Datensicherung durchführte.

Im langjährigen Betrieb zeigten sich jedoch Schwächen der damals erworbenen Systeme. So erwiesen sich die externen RAID-Systeme wegen einer ungünstigen Paarung aus relativ hoher Wärmeentwicklung einer bestimmten Plattenserie und einer dichten Packung im Chassis als sehr anfällig. In Spitzenzeiten fielen bis zu drei von ca. 140 eingesetzten Festplatten pro Woche aus. Erschwerend wirkte vor allem, dass diese Systeme nicht automatisch überwacht werden konnten und somit jeweils einzeln durch einen Administrator begutachtet werden mussten. Die an jedem Standort vorhandenen Tapelibraries erwiesen sich als zusätzlicher Hemmschuh in der zentralisierten Administration. Ihre Laufwerke waren dem Start-Stop-Betrieb auf Dauer nicht gewachsen und reagierten mit Ausfällen und Aufforderungen zu manuellen Eingriffen. Dies behinderte insbesondere den Betrieb, da in der Regel kein IT-Personal vor Ort im Einsatz ist.

Sowohl die Daten als auch deren jeweils am Standort befindliche Sicherung waren weitreichend verteilt. Daraus resultierte eine ungleiche Ausnutzung der Kapazitäten. Während einige der Systeme überlastet waren, bestanden auf anderen akute Kapazitätsmängel.

3 NAS-Systeme als dezentraler Datenspeicher

Im vorigen Kapitel wurden die inhärenten Probleme aufgezeigt, die sich im Betrieb der bisherigen Speicherstruktur zeigten und nur mit hohem Personalaufwand zu bewältigen waren. Um die Lage strategisch zu verbessern, wurde mit der Planung einer neuen Infrastruktur für die Datenablage und -sicherung begonnen.

Bei dieser Planung legten wir die folgenden Kernziele fest:

- UNIX/Linux- und Windows-Systeme können gleichzeitig auf den Datenbestand zugreifen
- Speicher und Datensicherung konsolidieren
- Daten und deren Sicherung nicht am gleichen Ort halten
- Management und Überwachung zentralisieren und insgesamt minimieren
- Hard- und Software vereinheitlichen
- Migration der Betriebssysteme der angeschlossenen Rechner erleichtern

Nach einer eingehenden Evaluierung der Marktlage wurde für den DWD eine neue Infrastruktur vorgesehen. Zunächst wird an den großen und mittleren Standorten jeweils ein Network Attached Storage (NAS) installiert. Der DWD entschied sich für NAS und gegen ein Storage Area Network (SAN) am jeweiligen Standort aufgrund folgender Fakten:

- Auf NAS exportierte Filesysteme bzw. Freigaben sind im ganzen WAN verfügbar.
- Eine Dateiablage auf NAS und teilweise auf Servern resultiert in nur einem Zugriffsparadigma über das Netzwerk.
- Eine gleiche Speicherarchitektur ist für alle Server und Clients möglich.
- Der Anschluss eines SAN an jede Workstation bzw. PC ist weder adäquat noch finanzierbar.

Auf die NAS greifen sowohl PC unter MS-Windows, als auch UNIX/Linux-Systeme zu. Als Netzwerkprotokolle werden CIFS und NFS verwendet. Die "Home"-Verzeichnisse der Nutzer sind im ganzen WAN erreichbar. Diese Speicherinfrastruktur, verbunden mit einer Windows-Domäne, ermöglicht es unseren Mitarbeiterinnen und Mitarbeitern, sich an jedem Arbeitsplatz in Deutschland anzumelden und dort ihre gewohnte Dateiumgebung vorzufinden. Zusätzlich wird an jedem UNIX/Linux-Rechner, an dem ein Nutzer Zugangsrechte hat, auf Wunsch sein "Home"-Verzeichnis zur Verfügung gestellt. Dies gibt den Nutzern die Möglichkeit, an den gleichen Daten unter verschiedenen Betriebssystemen zu arbeiten.

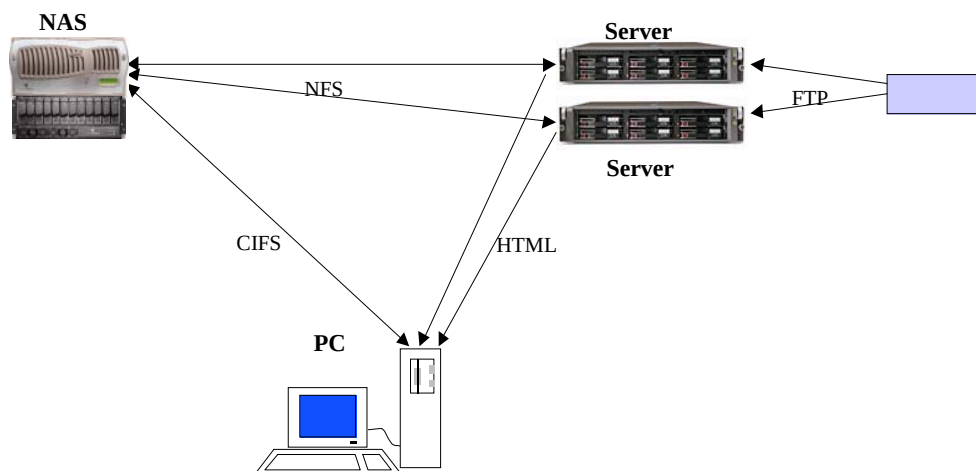


Abb. 2: Beispielkonfiguration Informationspräsentationssystem

In Abbildung 2 ist eine typische Mischkonfiguration mit Servern unter einem UNIX/Linux-Betriebssystem und einem bzw. vielen PC als Clients aufgezeigt. Bei dieser Konfiguration werden alle gemeinsam genutzten bzw. sicherungswürdigen Daten auf dem NAS abgelegt. Auf dieses System kann sowohl von den verarbeitenden Servern (via NFS) als auch von jedem PC (via CIFS) zugegriffen werden. Damit ist eine einfache Pflege und Einbindung der Daten in die Office-Umgebung ermöglicht. Die automatisierte Verarbeitung und Bereitstellung der Daten erfolgt durch Skripte auf den Servern. Im operationellen Arbeitsbereich kann jeder Arbeitsplatz via HTML auf die zur Verfügung gestellten Ansichten zugreifen. Die Zulieferung der erforderlichen Rohdaten erfolgt via FTP durch die DWD-Datenversorgung.

4 Übergreifende Datensicherung auf NAS-Systeme

Wesentliches Ziel der Restrukturierung der Dateninfrastruktur war die Verbesserung und Konsolidierung der Datensicherung innerhalb des DWD. Hierzu wurde unter Einbeziehung des Herstellers Network Appliance ein Konzept erarbeitet, das es ermöglicht, die regelmäßigen Datensicherungen aller NAS in Deutschland auf ein oder mehrere Backup-NAS abzulegen.

Zunächst werden auf den dezentralen NAS viermal täglich sog. Snapshots erzeugt. Diese Snapshots werden zweimal täglich in die Zentrale übertragen. Die hier eingesetzte Software SnapVault [4] erlaubt eine blockweise inkrementelle Datensicherung über das Netzwerk. Wie in allen Prognosen vorhergesagt wurde,

Wie in Kapitel 3 schon erwähnt, hat der DWD bis auf weiteres nur ausgewählte Standorte mit einem NAS ausgestattet. In der Fläche ist weiterhin die Dateiablage auf vorhandenen Windows-Servern vorgesehen. Um die Datensicherung dieser Geräte ebenfalls zentral zu konsolidieren, wurde ein Server mit der Software Tivoli Storage Manager (TSM) [6] aufgebaut. Diese Software ist ebenfalls in der Lage, ihren Datenbestand mit ausschließlich inkrementellen Sicherungen zu pflegen. Diese Fähigkeit ist unabdingbar, da bei nutzbaren Leitungskapazitäten von maximal 64 kb/s eine Komplettsicherung der Server mehrere Wochen in Anspruch nimmt. Eine Erhöhung der Leitungskapazitäten ist wegen der hohen Kosten für Standleitungen derzeit nicht vorgesehen. TSM komprimiert die Daten direkt auf den beteiligten Rechnern vor dem Transport und ist somit unabhängig von speziellen Fähigkeiten der Netzwerkinfrastruktur. Die Prozessorbelastung ist wegen der kleinen Netzwerkbandbreite vernachlässigbar. Als Massenspeichermedium wird ebenfalls das Backup-NAS eingesetzt, da es sich bei einer Kostenanalyse als preiswertes System für diesen Einsatzzweck zeigte. Zusätzlicher Vorteil ist natürlich die kurze Latenzzeit beim Dateizugriff. Auch bei dieser Struktur können einzelne Dateien über das Netz zurückgespielt werden. Bei einem kapitalen Hardwareschaden wird ein Reservesystem in Offenbach mit den Daten versehen und anschließend vom Techniker an den Einsatzort verbracht.

5 NAS-Systeme als Datenspeicher für hoch verfügbare Systeme

Derzeit installiert der DWD mehrere hoch verfügbare Systeme, deren Datenablage auf NAS erfolgt. Die wesentlichen Verfahren, die derzeit auf eine Hochverfügbarkeitslösung unter Linux mit Datenablage auf NAS umgestellt werden, sind der Mailhub und die interne Verteilung von Fertigprodukten zur meteorologischen Beratung. In Abbildung 4 ist als typisches Beispiel die WWW-Server-Farm aufgezeigt. Dieses Serversystem ist seit ca. 6 Monaten im Betrieb und erfüllt problemlos die gestellten Anforderungen.

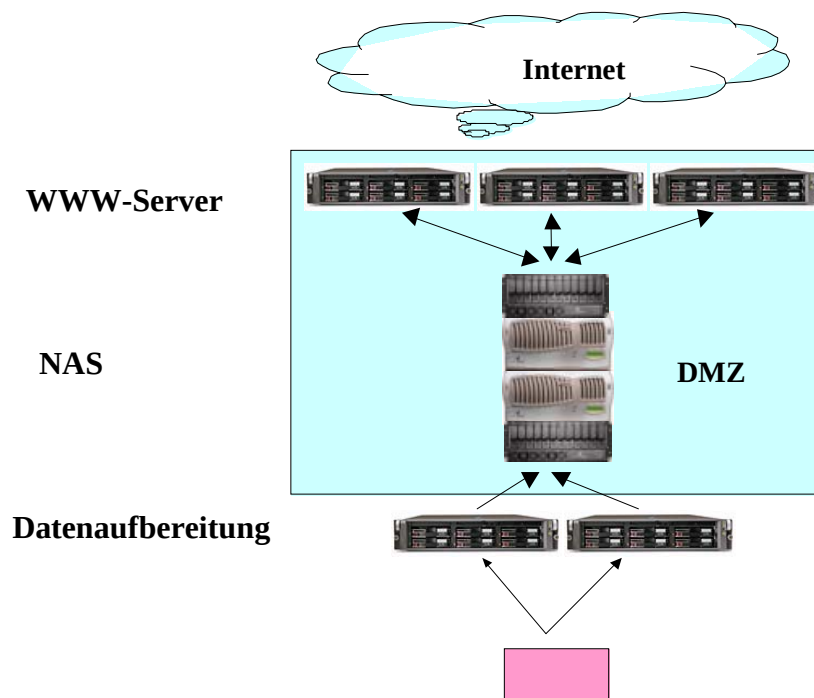


Abb. 4: Struktur der WWW-Server

Die Webpräsenz des DWD wird von mehreren Servern geleistet, die im Loadbalancing-Verfahren betrieben werden. Diese Rechner stellen eine Vielzahl von Warnungen und sonstigen Produkten für die verschiedenen Kundengruppen des DWD bereit. Wegen der speziellen Anforderungen an die Daten ist es erforderlich, dass jeder Server zusätzlich zum lesenden auch schreibenden Zugriff auf den gemeinsam verfügbaren Datenbestand hat. Dies ist am einfachsten mittels eines gemeinsamen Speichers zu realisieren. Dieser wiederum wird von der Datenaufbereitung im Intranet kontinuierlich mit den neuesten Vorhersagedaten, Bilddaten und Warnungen beschickt. Als Ablagesystem wird hier ein NAS verwendet, dessen Hardware keinen Single Point of Failure hat und somit als ausfallsicher gilt. Das komplette System aus Servern und NAS ist in der Demilitarisierten Zone (DMZ) aufgestellt.

Durch diesen Einsatz von NAS ist es möglich, Hochverfügbarkeitsanwendungen vom Speicher zu entkoppeln. Das Management dieser Systeme wird durch diese Maßnahme vereinfacht und das Personal entlastet.

6 Fazit

Die ersten NAS im DWD sind inzwischen seit über einem Jahr im Betrieb. Die Verfügbarkeit dieser Geräte ist exzellent. Ungeplante Ausfälle waren bislang auf den Produktionssystemen nicht zu verzeichnen. Bei einem notwendigen Upgrade des Betriebssystems (Data ONTAP) [7] wurde die extrem kurze Ausfallzeit von ca. 50 Sekunden gemessen. Dies zeigt unseres Erachtens deutlich, wie sinnvoll das Konzept eines speziellen, besonders schlanken Betriebssystems für NAS ist. Der Betreuungsaufwand konnte bei verbesserter Leistung deutlich gesenkt werden. Insgesamt wurde durch die Einführung von NAS im DWD eine gesteigerte Nutzerzufriedenheit sowohl durch geringere Ausfallzeiten von insgesamt weniger als fünf Minuten im Jahr als auch durch gute Performancewerte erreicht. Da Überwachung und Management an einem Ort konzentriert sind, ist der Verbund aus mehreren NAS unter Verwendung von SnapVault ein stabiles, robustes und handliches System. Unter zusätzlicher Verwendung von TSM konnte dieses Konzept auf die kleineren Standorte erweitert werden.

Mit dem Einsatz von NAS-Systemen hat der DWD eine von jedem Client und Server nutzbare, einheitliche, flexible, erweiterbare und sichere Infrastruktur geschaffen, die eine zukunftssichere Basis für die Computer gestützten Arbeiten bietet. Durch diese Lösung ist insbesondere sichergestellt, dass keinerlei Abhängigkeiten von Server-Betriebssystemen einen Einsatz von Software behindern. Somit ist der DWD auch für den Einsatz von Open System Software bestens gerüstet.

Referenzen

- [1] TOP500.Org, *TOP500 Supercomputer sites: TOP500 List 06/2003*, <http://www.top500.org/list/2003/06/>, 2003
- [2] Deutscher Wetterdienst, Technische Infrastruktur, *Ausstattung des Rechenzentrums*, <http://www.dwd.de/de/Technik/Datenverarbeitung/zentral/Ausstattung.htm>, 2003
- [3] Deutscher Wetterdienst, Technische Infrastruktur, *Dezentrale Systeme*, <http://www.dwd.de/de/Technik/Datenverarbeitung/dezentral.htm>, 2003
- [4] Network Appliance GmbH, *Network Appliance - Filer - SnapVault-Software*, <http://www-de.netapp.com/products/filer/snapvault.html>, 2003
- [5] Network Appliance GmbH, *Network Appliance - NearStore - R100*, http://www-de.netapp.com/products/nearstore/r100_ds.html, 2003

- [6] Rolan Tretau et al., *IBM Tivoli Storage Manager Implementation Guide*,
<http://publib-b.boulder.ibm.com/Redbooks.nsf/RedbookAbstracts/sg245416.html>, 2003
- [7] Network Appliance GmbH, *Network Appliance - Filer Data ONTAP-Software*, <http://www-de.netapp.com/products/filer/ontap.html>, 2003

Hochverfügbarkeit mit Oracle-Datenbanken

Torsten Rosenwald
Ingo Frost

Trivadis GmbH
Ferdinand-Stuttman-Strasse 13
65428 Rüsselsheim
Germany
E-mail: torsten.rosenwald@trivadis.com

Mit der Verfügbarkeit immer preiswerterer Hardware werden Hochverfügbarkeitslösungen immer erschwinglicher. Welche Möglichkeiten durch Oracle-Software geboten werden, um Hochverfügbarkeit auf derartiger Hardware zu erreichen, soll dieser Artikel beschreiben.

1 Einleitung

Grundsätzlich stellt sich die Frage, was durch das Einrichten eines hochverfügbaren Systems erreicht werden soll. Ein maßgebliches Kriterium für eine solche Entscheidung ist die Antwort auf die Frage: was kostet die Nichtverfügbarkeit dieses Systems?

1.1 Wieso Hochverfügbarkeit?

Stellt man sich die Frage nach den Kosten von Nichtverfügbarkeit eines Systems, so erhält man je nach Geschäftsbereich sehr unterschiedliche Antworten. So erhält man Antworten, die in der Größenordnung von einigen Millionen Euro liegen können. Gerade für solche Systeme muss also eine Ausfallzeit gegen Null gehen.

1.2 Worin bestehen die Gefahren für die Verfügbarkeit eines Systems?

Interessanterweise liegen die Probleme zu fast der Hälfte im Bereich von Hardwarefehlern in den Festplatten-Subsystemen. Dicht heran an diese Fehlerrate kommt das menschliche Versagen. Viren, Korruptionen und Naturkatastrophen ergeben zusammen gerade einmal 14% der Ursachen für Datenverluste. Somit zeigt sich, dass sich allein durch das Verwenden geeigneter Hardware ein Großteil an Verfügbarkeit herstellen lässt.

1.3 Wie erreicht man nun Hochverfügbarkeit?

Zum einen erweist es sich als sehr vorteilhaft, Redundanz an so vielen Stellen wie möglich zu erzeugen, zum anderen sollte man zu viele menschliche Eingriffe zu verhindern suchen.

Redundanz bezieht sich in diesem Kontext auf Hardware-Redundanz in Form von Prozessoren, Festplattenspeichern, Netzwerken und Stromversorgungen, auf Software-Redundanz in Bezug auf das Betriebssystem und der Anwendung als solcher und schließlich in Form von Daten-Redundanz in Form von Campus-Cluster oder Geomirroring.

Um den Faktor Mensch als Einflussgröße für Nichtverfügbarkeit niedrig zu halten, sollte man versuchen, menschliche Eingriffe in das System möglichst zu vermeiden. Dies gilt sowohl für reguläre Operatoren, als auch für so genannte Anwendungs-Administratoren, die aus Aspekten der Sicherheit heraus eigentlich nicht für einen direkten Zugriff auf ein Produktionssystem geeignet sind. Es hat sich als sehr vorteilhaft erwiesen, regelmäßige Crash-Übungen durchzuführen, so dass die Administratoren Routine in der Bewältigung von Problemfällen erlangen können.

1.4 Wie kann Oracle9i Release 2 dabei von Nutzen sein?

Oracle stellt verschiedene Optionen und Werkzeuge zur Verfügung, die dazu dienen können, hochverfügbare Systeme zu erhalten.

Meistens werden die Werkzeuge in Kombination eingesetzt, um fallspezifisch die bestmögliche, d.h. unter anderem auch kürzeste Ausfallzeit erreichen zu können.

Als Werkzeuge wären hier folgende zu nennen:

- RMAN (Recover Manager)
- Log Miner

Optionen in diesem Kontext sind:

- Data Guard
- Real Application Clusters

Weiterhin gibt es noch Eigenschaften der Oracle-Software, die im Lieferumfang enthalten sind und bei der Herstellung von Hochverfügbarkeit von Nutzen sind. Diese sind:

- Fast Restart
- Dynamic Reconfiguration
- Online Redefinition
- Partitioning

Datenverfügbarkeit erreicht man durch Verwendung von RAID-Systemen und Data Guard.

Dienstverfügbarkeit erreicht man durch Real Application Clusters, Failover Cluster-Systeme (SUN Cluster, Veritas Cluster, Microsoft Cluster, etc) und Data Guard.

Wie bereits erwähnt, sind häufig Kombinationen der vorgenannten Technologien im Einsatz, um den Hochverfügbarkeitsanforderungen wirklich gerecht werden zu können.

1.5 Fazit

Man sollte sauber zwischen den Begriffen Daten- und Diensthochverfügbarkeit trennen. Hochverfügbarkeit kann über verschiedene Wege erreicht werden. Dabei werden häufig Kombinationen der vorgestellten Methoden Verwendung finden. Oracle unterstützt Hochverfügbarkeit vom Kern bis hin zu spezialisierten Optionen.

2 Oracle Data Guard

Eine Oracle-Data-Guard-Konfiguration besteht aus zwei Oracle-Datenbanken, die beide im Archivelog-Modus laufen. Eine Oracle-Datenbank, die im Archivelog-Mode läuft, erstellt eine Kopie der Online-Redo-Logdateien als so genanntes Archived-Redo-Log jedes Mal, wenn ein Online-Redo-Log vollständig beschrieben wurde und beschreibt dann das nächste Redo-Log.

Eine dieser Datenbanken ist die so genannte ‚Primary Site‘, während die andere Datenbank ‚Standby Site‘ genannt wird.

2.1 Physical Standby

Die physische Standby-Datenbank ist eine exakte Kopie der Primärdatenbank. Sogar die Strukturen auf der Festplatte sind absolut identisch. Datenbankschemata

und Indizes sind identisch. Die Synchronisation einer physischen Standby-Datenbank erfolgt durch das Recovery der archivierten Redo-Logs, die von der Primary Site übertragen werden.

2.2 Logical Standby

Die logische Standby-Datenbank enthält aus logischer Sicht, d.h. struktureller Sicht, dieselben Informationen wie die Primary Site. Die physikalische Organisation und Struktur der Daten können aber durchaus unterschiedlich sein. Die Synchronisation der Daten erfolgt über die Transformation der Daten der Redo-Log-Dateien in SQL-Ausdrücke und anschließendes Anwenden dieser SQL-Ausdrücke auf die Standby-Datenbank. Damit dies geschehen kann, muss die Datenbank geöffnet sein.

Eine logische Standby-Datenbank kann also auch jederzeit für Abfragen und Zwecke des Berichtswesens genutzt werden.

2.3 Data Protection Modes

Data Guard stellt drei so genannte Protection Modes zur Verfügung:

Im ‚Maximum Protection‘-Modus ist die höchste Datensicherheit gewährleistet. Die Daten werden synchron von der Primär- zur Standby-Datenbank gesendet. Allerdings kann die Primärdatenbank durch einen Fehler in der Verbindung zur Standby-Datenbank zum Stillstand kommen, da die Datenübertragung synchron vorgenommen wird.

Der ‚Maximum Availability‘-Modus bietet sehr hohe Datensicherheit, leider sind hierbei aber dennoch Datenverluste möglich, da die Daten divergieren können.

Der ‚Maximum Performance‘-Modus erzeugt keine Performanz-Einbußen auf der Primärdatenbank. Die Datensicherheit ist hoch, die Daten werden asynchron transportiert und, wenn gewünscht, verzögert auf die Datenbank angewendet.

2.4 Fazit

Data Guard eignet sich hervorragend, um Disaster-Szenarien abzufangen und sich vor durch Menschen verursachten Fehlern zu schützen. Data Guard ist leicht einzurichten und zu pflegen. Reporting beeinträchtigt das ‚Tagesgeschäft‘ einer OLTP-Anwendung nicht, da es komplett auf einer Logical-Standby-Datenbank ausgeführt werden kann. Damit können die Ressourcen sowohl des Primär-Datenbank-Servers als auch des Standby-Servers effizient genutzt werden.

3 Oracle Real Application Clusters

Oracle Real Application Clusters ist eine neue Oracle-Datenbank-Technologie, die mit der Version Oracle9i Release 1 eingeführt wurde.

Ein Oracle Real Application Clusters Rechnersystem besteht aus mehreren gleichwertigen Computern, die mit einem privatem Netzwerk, dem Interconnect, und einem gemeinsamen Plattenbereich ausgestattet sind, auf den beide Computer konkurrierend zugreifen können. Die Client-Konnektivität entsteht durch ein weiteres Netzwerk.

Signifikant für diese neue Oracle-Technologie ist die Verwendung des Speichers beider Computer als virtuellen gemeinsamen Speicherbereich über die so genannte Cache Fusion.

3.1 Cache Fusion

Mit Hilfe der Cache Fusion erreicht man schnellere Instanz-übergreifende konkurrierende Zugriffe.

In den Oracle-Datenbank-Releases vor der Version Oracle9i wurden Datenbankblöcke, die in einer Instanz verändert wurden und durch eine andere Instanz angefordert wurden, von der ersten Instanz auf die Festplatte geschrieben und dann von der anfordernden Instanz von der Festplatte gelesen. Gegebenenfalls konnte die erste Instanz diesen Block erneut anfordern, womit der vorher beschriebene Vorgang, diesmal allerdings in umgekehrter Richtung, erneut begann. Dieses Ping-Pong-Spiel wird Block Pinging genannt. Für sehr viele derartige konkurrierende Zugriffe war immer mit eklatanten Performanz-Einbußen zu rechnen.

Ab Oracle9i kann das Oracle RDBMS diesen im Speicher befindlichen Block direkt über den Interconnect an die anfordernde Instanz schicken. Da der Speicher und die Netzwerkverbindung geringere Latenz besitzen als die zugehörigen Festplatten, erreicht man dadurch erhebliche Performanz-Vorteile.

3.2 Fazit

In Real-Application-Clusters-Umgebungen können alle aktiven Instanzen konkurrierend auf eine zwischen allen Instanzen geteilte Datenbank zugreifen. Real-Application-Clusters koordiniert den Zugriff jeder Instanz auf die geteilten Daten um Datenkonsistenz und Datenintegrität zu gewährleisten.

Stürzt eine dieser Instanzen ab, können die anderen Instanzen deren Aufgaben mit übernehmen. Dadurch können sehr hohe Verfügbarkeitsanforderungen realisiert werden.

4 Oracle Grid Computing

Grid Computing ruht auf fünf Fundamenten:

- Virtualisierung
- Dynamische Versorgung mit Ressourcen
- Bildung von Ressourcen-Pools
- Selbstadaptive Systeme
- Vereinheitlichtes Management.

4.1 Virtualisierung

Virtualisierung ist die Abstraktion jedes physikalischen und logischen Bestandteils in einen Dienst. Virtualisierung ist deswegen wichtig, da sie die Grid-Komponenten (z.B. Storage, Prozessoren, Datenbanken, Applikationsserver und Applikationen) eng integriert.

Die Virtualisierung ermöglicht es jeder Grid-Komponente schneller auf geänderte Umstände zu reagieren und sich Fehlern von Komponenten anzupassen, ohne die Performanz des Gesamtsystems zu beeinträchtigen.

4.2 Dynamische Versorgung mit Ressourcen

Ressourcen-Anfragen sollen dorthin verteilt werden, wo sie am sinnvollsten ausgeführt werden können. Im Kontext des Grid können diese Ressourcen-Anfragen Server-Anfragen, Datenzugriffe oder Berechnungen sein. Im Grid gibt es den Ressource Broker, dem der Ressourcen-Bedarf eines Elements bekannt ist und der weiß, welche Ressource diesen Bedarf decken kann. Dieser Broker bringt diese beiden Elemente dann zusammen, so dass die Ressourcen effizient genutzt werden können. Sollten sich die Gegebenheiten des Systems ändern, so ändern sich auch die Verknüpfungen. Richtlinien, wie z.B. Antwortzeit-Schwellwerte oder vorausgeahnte Spitzenlasten, können genutzt werden, die Verknüpfungen zwischen Ressourcen-Anforderern und Ressourcen-Providern zu optimieren.

4.3 Bildung von Ressourcen-Pools

Konsolidierung und Vereinigung von Ressourcen ist notwendig, damit das Grid die Ressourcen besser ausnutzen kann. Durch das Pooling erzeugt man mehr Flexibilität, um Verknüpfungen der Ressourcen zwischen den Systemen zu optimieren.

4.4 Selbstadaptive Systeme

Eine kritische Anforderung an ein Grid-System ist die Automation von Wartungs- und Tuning-Aufgaben, die für gewöhnlich von den Administratoren der Systeme durchgeführt wurden. Mehr und mehr dieser Aufgaben müssen in Zukunft von den Systemen selbst erledigt werden.

4.5 Vereinheitlichtes Management

Selbst mit selbstwartenden Systemen müssen jedoch immer noch Menschen in das Management des Grid verwickelt sein. Aber diese Aufgaben sollten durch ein einziges Werkzeug, das jedes Element des Grid überwachen und verwalten kann, vereinfacht werden. Ein solches Tool sollte Verfügbarkeit und Performanz aus der Sicht des Anwenders beurteilen, so dass jeder Flaschenhals oder jede nicht verfügbare Komponente einen Alarm erzeugt. Weiterhin muss es möglich sein, Gruppen von Systemen als logische Einheit zu verwalten, so dass Aufgaben ein einziges Mal durchgeführt, aber gleichzeitig auf mehreren Maschinen wirken können.

Referenzen

- [1] Oracle Corporation, *Oracle Data Guard – Concepts and Administration – Release 2 (9.2)*
http://download-uk.oracle.com/docs/cd/B10501_01/server.920/a96653.pdf,
October 2002, Part No. A96653-02
- [2] Oracle Corporation, *Oracle9i Real Application Clusters – Concepts – Release 9.2*
http://download-uk.oracle.com/docs/cd/B10501_01/rac.920/a96597.pdf,
March 2002, Part No. A96597-01
- [3] Oracle Corporation, *Oracle 10g: Infrastructure for Grid Computing - An Oracle White Paper*,
http://otn.oracle.com/tech/grid/collateral/GridTechWhitePaper_final.pdf,
September 2003

DBShadow

Uwe Stocker

Libelle Informatik GmbH
70565 Stuttgart
Germany
E-mail: ustocker@libelle.com

1 Ausfallursachen und Vorkehrungen

Konzepte wie RAID- und Clusterlösungen helfen nicht bei logischen Fehlern. Klassische Abhilfe ist in diesen Fällen ein zeitaufwendiges Recovery.

Libelle DBShadow ermöglicht Ihnen ein Weiterarbeiten innerhalb weniger Minuten, unabhängig von der Fehlerursache.

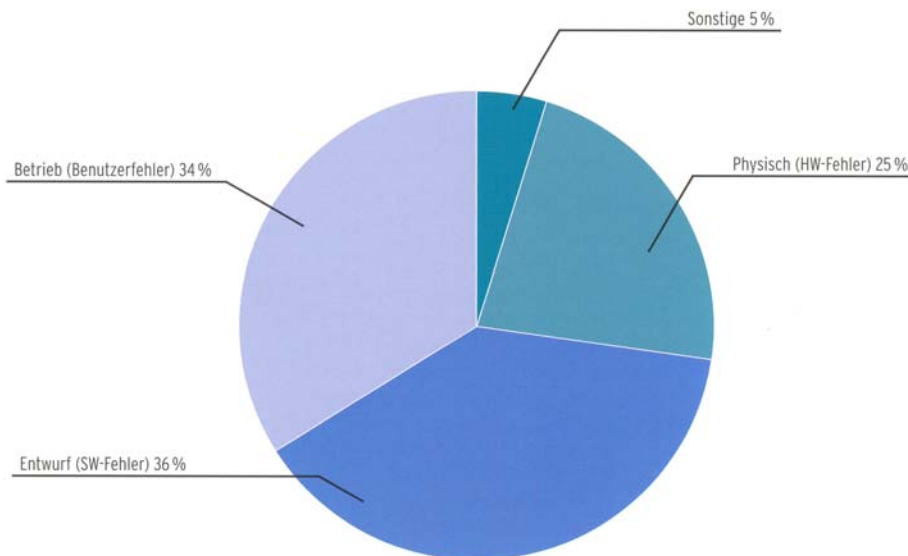


Abb. 1: Oracle

2 Ausfallursachen und Vorkehrungen

Libelle DBShadow schützt Sie vor logischen Fehlern in der Datenbank. Typische logische Fehler sind zum Beispiel Softwarefehler, Benutzerfehler, fehlerhafte Batchjobs, korrupte Datenblöcke und viele weitere.

Das Konzept von Libelle DBShadow schützt zusätzlich vor Hardware-, Betriebssystem- und Middlewareproblemen durch Umschalten auf das Spiegelsystem.

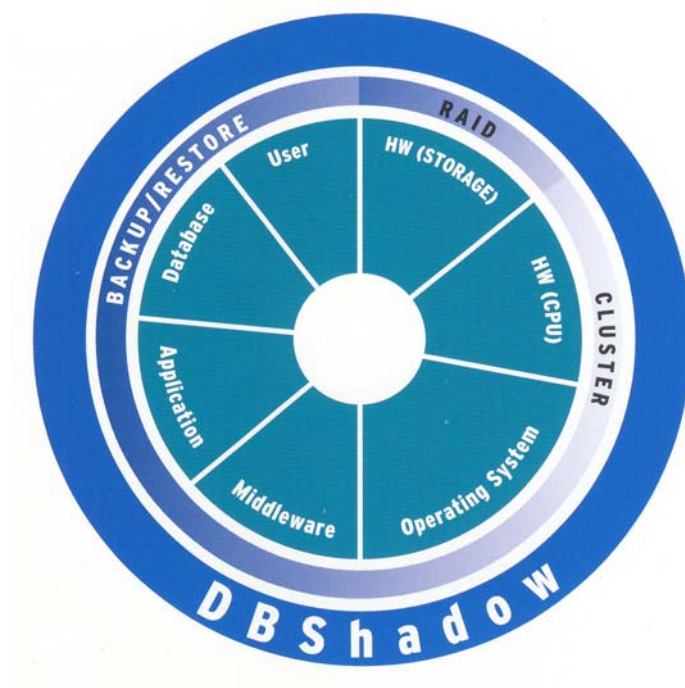


Abb. 2: Libelle

3 Funktionsprinzip

Schritt 1: Die gesamte Datenbank wird im laufenden Betrieb auf das Zielsystem kopiert. Die Kopie wird durch eine TCP/IP-Socketkommunikation realisiert. Hohe Performance ergibt sich durch paralleles und komprimiertes Kopieren (DBShadow InitialCopy).

Schritt 2: Alle Veränderungen in der Echtdatenbank werden sofort auf das Spiegelsystem kopiert (DBShadow Archiver-Prozess).

Schritt 3: Veränderungen werden zeitversetzt nachgeführt. Hierfür wird das Libelle-Trichterprinzip verwendet (DBShadow Recover-Prozess).



Abb. 3: Libelle

4 Funktionsprinzip Time-Delay

In der Beispielgrafik benötigt eine Transaktion 6 Stunden, bevor diese den Trichter durchlaufen hat. Erst jetzt wird die Transaktion der Spiegeldatenbank zugeführt. Die Spiegeldatenbank befindet sich hierbei in einem permanenten Recovery-Zustand.

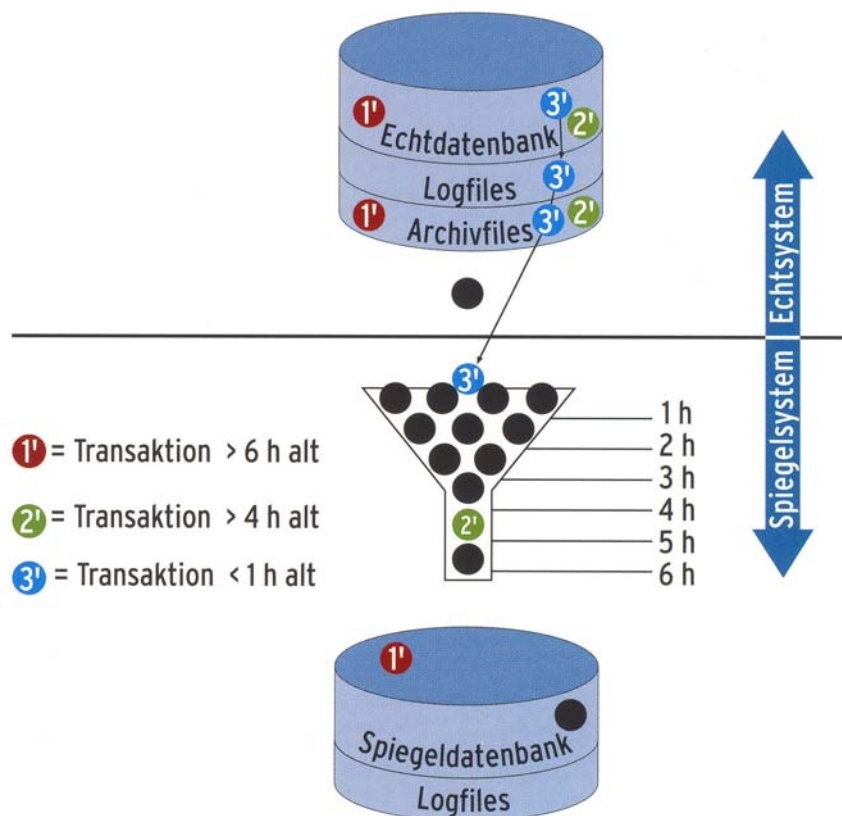


Abb. 4: Libelle

5 Normalbetrieb der Datenbank

Vor der InitialCopy finden zahlreiche Prüfungen statt (Plattenplatz, Verzeichnisse, Berechtigungen, etc.). Erst dann wird die Datenbank vom Echt- auf das Spiegelsystem kopiert.

Die InitialCopy erfolgt während des laufenden Betriebs. Je nach Systemumgebung (Betriebssystem-, Netzwerk- und Hardware-Performance) sind Kopierleistungen von über 200 GByte pro Stunde möglich.

Während des Normalbetriebes stehen die DBShadow-Prozesse ständig in Kontakt mit dem Echt- und Spiegelsystem. Es erfolgt ein kontinuierlicher Informationsabgleich.

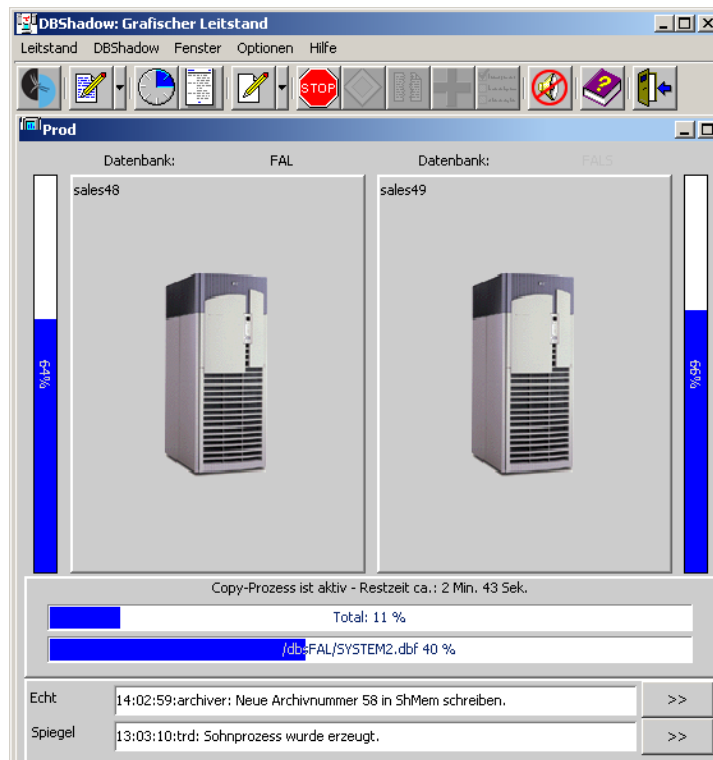


Abb. 5: Libelle

6 Fehler in der Datenbank

Beim Auftreten einer Störung wird umgehend ein akustisches und grafisches Signal ausgelöst.

In Verbindung mit dem Libelle SMSCommander können Störungen sofort per SMS an ein definiertes Mobiltelefon versendet werden. Auch eine Fehleranalyse und Umschaltung ist per Mobiltelefon möglich.

DBShadow lässt sich in alle gängigen Systemmanagement-Tools integrieren.

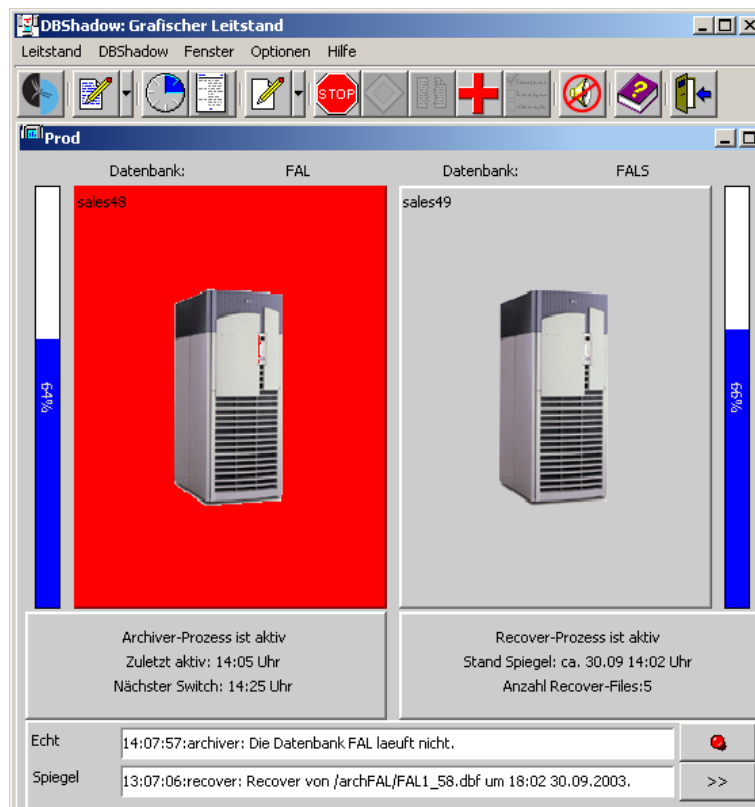


Abb. 6: Libelle

7 Umschaltkonzept

Nach Auftreten eines Störfalls kann die Datenbank umgeschaltet werden. Durch Aktivieren des "roten Kreuzes" haben Sie die Auswahl zwischen verschiedenen Umschaltvarianten.

Sofern die aktiven Logfiles verfügbar sind, bietet sich bei Hardwarefehlern die Möglichkeit, bis zur letzten Transaktion "verlustfrei" umzuschalten. Bei einem Anwender- oder Softwarefehler bestimmen Sie exakt, auf welche Uhrzeit der Spiegel hochgefahren wird.

Das Spiegelsystem steht dann als produktives System voll zur Verfügung.

The screenshot shows a Windows-style application window titled 'Prod'. At the top, there is a button 'Umschalten auf den Spiegelrechner' and two smaller buttons 'Abbrechen' and 'Hilfe'. Below this, there are four main sections, each with a checkbox and a corresponding 'Umschalten' button.

- Umschalten bis <Zeitpunkt>** (checked): This section includes a label 'Daten Spiegelrechner nachziehen bis:' followed by a date/time picker set to 01/10/2003 23:59:59. A 'Umschalten' button is to the right.
- Lossless-Switch** (unchecked): This section includes a text field for 'Redo-Log-Verzeichnis' and a checkbox for 'Stop Switch' with the note 'wenn kein Online-Log vorhanden ist'. A 'Umschalten' button is to the right.
- Defined-Switch** (unchecked): This section is mostly empty with a 'Umschalten' button to the right.
- Notfall-SID: Umschalten bis <Zeitpunkt>** (unchecked): This section includes a text field for 'Notfall-SID', a label 'Daten Spiegeldatenbank nachziehen bis:', and a date/time picker set to 01/10/2003 23:59:59. It also has a 'Lossless-Switch auf den Spiegelrechner' checkbox, a 'Redo-Log-Verzeichnis' text field, and a 'Stop Switch' checkbox with the note 'wenn kein Online-Log vorhanden ist'. A 'Umschalten' button is to the right.

Abb. 7: Libelle

8 Vorteile durch DBShadow

Optimaler Schutz vor:

- logischen Fehlern
- Hardwarefehlern
- Corrupt-Blocks
- Katastrophen und Sabotage

Unabhängig von:

- Entfernung Echt- und Spiegelsystem
- Hardware und Applikation
- Datenbankgröße

Auch einsetzbar für:

- das Erstellen von Testdatenbanken
- Auswertungsläufe vom Spiegel
- Zeitgewinn bei Softwareupdates

Weitere Vorteile:

- Schneller Wiederanlauf nach einer Störung
- Verringerung der potentiellen Downtime
- Grafische Überwachung aller Prozesse
- Geringe Belastung des Spiegelsystems
- Sehr einfache Bedienung
- Backup vom Spiegel, dadurch Entlastung des Echtsystems

Virtuelles SAN für einen Supercomputer

Lothar Wollschläger

Forschungszentrum Jülich GmbH
Zentralinstitut für Angewandte Mathematik
52425 Jülich
E-Mail: l.wollschlaeger@fz-juelich.de

Es wird beschrieben, wie der Einsatz eines virtuellen SAN für einen Supercomputer, ein Cluster aus 37 Systemen, dazu genutzt wird, die große Menge der Daten zu speichern und auf jedem System innerhalb des Clusters verfügbar zu machen. Ausgehend von einem kleineren System mit 6 Systemen wird der Übergang von einem realen SAN auf ein virtuelles SAN beschrieben.

1 Einführung

Beim Aufbau eines Supercomputers aus mehreren Systemen stellt sich die Frage, wie man dieses System effizient mit der Plattenperipherie verbindet, um einerseits ein Single System Image zu erhalten, andererseits aber genügend Performance erzielt um die enormen Anforderungen an die I/O-Leistungen der Anwendungen zu erfüllen. Dabei bietet sich die Technik des Storage Area Networks (SAN) an, die es ermöglicht, die Plattensysteme mit mehreren Rechnern zu verbinden. Wenn jedoch die Anzahl der Systeme zu groß wird, wird auch eine SAN-Lösung zu komplex, so dass dann auch das SAN virtualisiert werden muss.

2 SAN-Lösungen bei kleineren Clustern

Bei kleineren Clustern bis zu 16 Knoten bietet sich der Einsatz eines SAN an, um allen Knoten einen gemeinsamen Zugriff auf die Plattenperipherie zu ermöglichen. Bei deutlich größeren Clustern wird das SAN zu komplex und die notwendige Switch-Technologie zu aufwändig. Dann muss nach anderen Lösungen gesucht werden.

Für das zurzeit am Forschungszentrum Jülich installierte Cluster bestehend aus 6 Knoten ist eine SAN-Lösung die einzige Möglichkeit, da das System über keine schnelle interne Switch-Verbindung verfügt. Abb. 1 zeigt die aktuelle SAN-Konfiguration. Das interne Gigabit-Ethernet-Netzwerk ist nicht schnell genug um

darüber einen Datenzugriff für alle Knoten zu realisieren, der den Geschwindigkeitsanforderungen der Anwendungen entspricht.

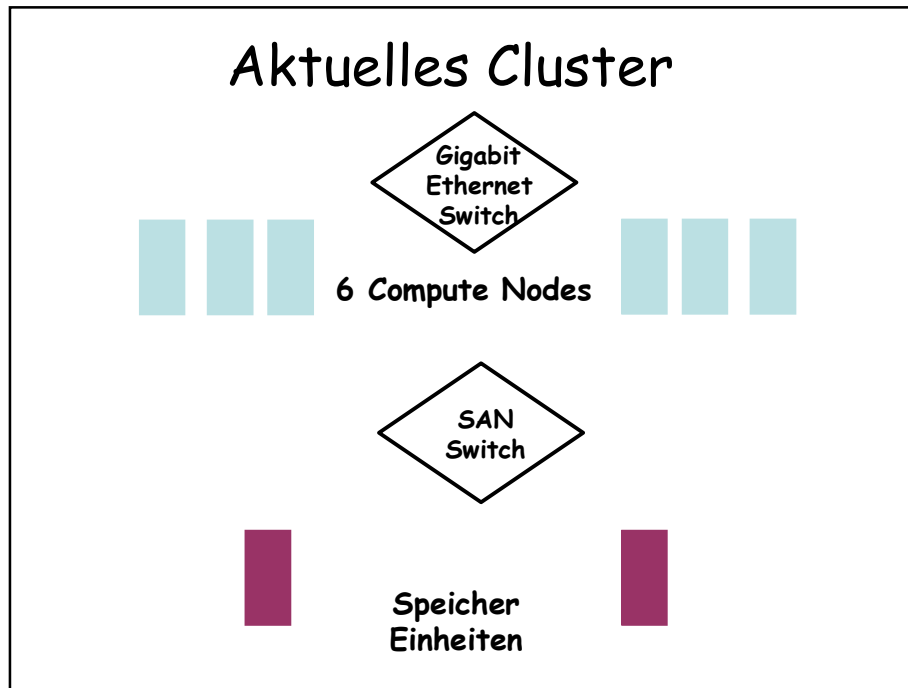


Abb. 1 SAN-Konfiguration aktuelles Cluster

Die aktuelle SAN an Konfiguration besteht aus zwei FASTT Storage Servern mit je 70 Festplatten zu je 72 GB. Zusätzlich ist ein StorageTek-Bandroboter mit 6 Bandlaufwerken vom Type 9940B angeschlossen. Um Ausfallsicherheit zu gewährleisten, werden zwei 16 Port FibreChannel Switches eingesetzt. Da jedoch entgegen aller schönen SAN-Theorie Plattenlaufwerke und Bandlaufwerke nicht am gleichen Fibre-Channel-Netzwerk betrieben werden können, müssen die Switches jedes in zwei virtuelle Switches eingeteilt werden, in so genannte Zonen, je eine Zone für die Plattenlaufwerke und eine Zone für die Bandlaufwerke. Die Bandlaufwerke sind nicht mit allen Knoten des Clusters verbunden, sie werden nur von einer LPAR eines Knoten benutzt, auf dem die Speichermanagement-Software Tivoli Storage Manager (TSM) eingesetzt wird. Die Bandstationen werden nur von dieser Komponente benutzt, die Benutzer haben keinen direkten Zugriff zu den Bandstationen. Abbildung 2 zeigt die detaillierte SAN-Konfiguration. Schon für ein relativ kleines Cluster ist die SAN-Konfiguration recht komplex. Für das geplante Cluster mit 35 Compute-Knoten wäre eine solche Konfiguration, bei der jeder Knoten über das SAN auf alle

Storage-Systeme zugreift, kaum noch zu durchschauen und zu verwalten. Deshalb wird für das große Cluster auch das SAN nochmals virtualisiert.

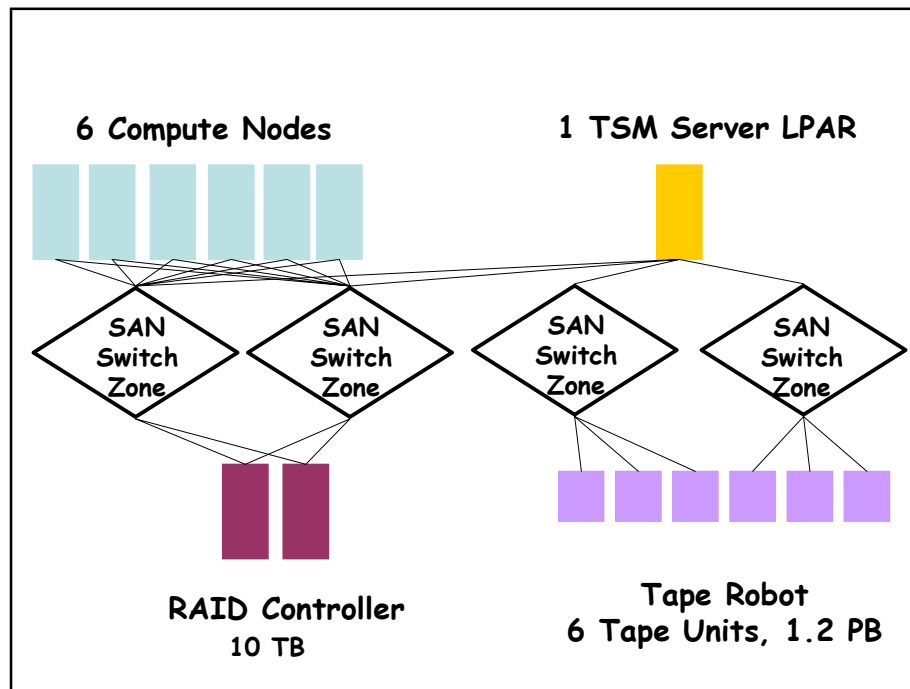


Abb. 2 Detaillierte SAN-Konfiguration

3 RAID-Kontrolller

Aktuell werden zur Speicherung der Benutzerdaten 140 Festplatten mit einer Kapazität von je 72 GB benutzt, im nächsten Schritt werden über 800 Festplatten verwendet. Bei einer so großen Anzahl muss man sich Gedanken über die Ausfallwahrscheinlichkeit einer Festplatte machen. Um einen ausfallsicheren Betrieb gewährleisten zu können, müssen die Daten auf den Festplatten mit Hilfe von RAID-Kontrollern abgesichert werden. Wir verwenden deshalb zur Speicherung der Benutzerdaten nur RAID5 Volumes mit 4+1 Festplatten. Das heißt, aus je 5 Festplatten wird ein virtuelles Volume gebildet, indem zu jeweils 4 Festplatten mit Daten eine fünfte Festplatte mit Parity-Informationen gebildet wird. Damit kann der Ausfall einer Festplatte ohne Datenverlust und ohne Betriebsunterbrechung überstanden werden. Die defekte Platte wird dann im

laufenden Betrieb durch eine Hot-Spare-Platte ersetzt. Bei den RAID-Controllern werden also aus den physikalischen Festplatten jeweils 18 „virtuelle“ Platten gebildet, die Aufteilung zeigt Abb. 3.

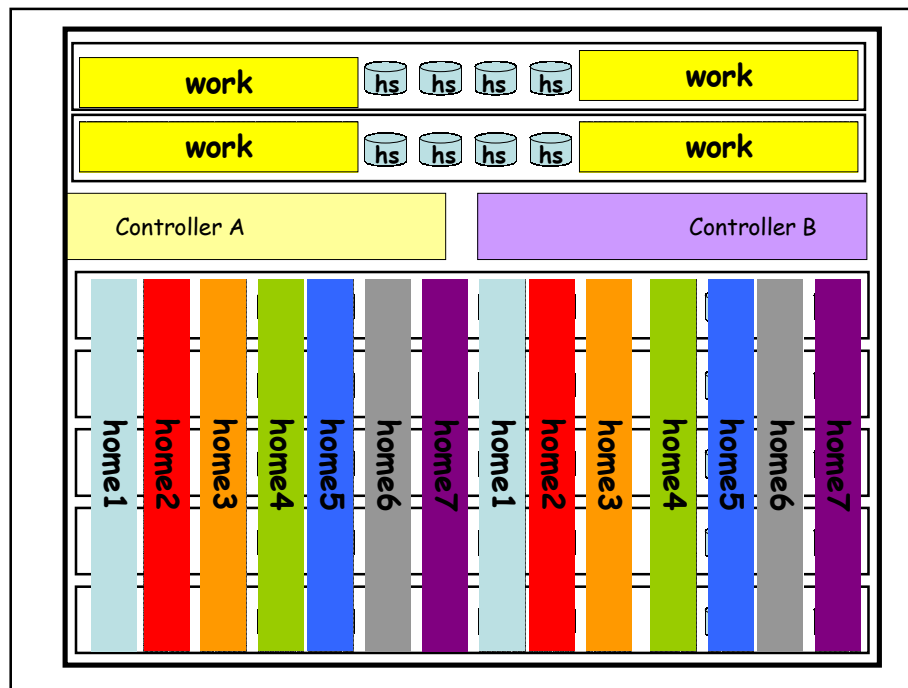


Abb. 3 RAID Volumes

4 Filesystem

Von entscheidender Bedeutung ist natürlich das zugrunde liegende Filesystem. Bei einem Cluster müssen auf allen Knoten die gleichen Daten mit der gleichen Datensicht zur Verfügung stehen, natürlich mit der entsprechenden Performance. Außerdem muss es möglich sein, von mehreren Knoten aus gleichzeitig auf die selbe Datei zu schreiben. Der Einsatz von NFS oder ähnlichen Techniken wird nicht die gewünschte Geschwindigkeit bringen und der parallele Schreibzugriff ist nicht möglich. Auf unserem System verwenden wir das General Parallel File System (GPFS) von IBM. Das GPFS ist ein hochperformantes paralleles Filesystem für AIX oder LINUX-Cluster, welches mehreren Anwendungsprozessen ermöglicht, von verschiedenen Knoten aus konkurrierend auf die gleiche Datei zuzugreifen. Das GPFS sieht für den Benutzer aus wie ein lokales Filesystem,

welches gleichzeitig auf allen Knoten zur Verfügung steht. Dabei sorgen Locking-Mechanismen dafür, dass die Konsistenz der Dateien und des Filesystems erhalten bleibt. Diese werden auf die am GPFS teilnehmenden Knoten verteilt, so dass nicht ein einzelner Knoten einen Engpass bildet. Auch können bis zur Hälfte der Knoten ausfallen, ohne dass die Funktionalität des GPFS beeinträchtigt wäre. Zugriff auf die physikalischen Festplatten kann entweder über ein SAN realisiert werden, oder wenn ein genügend schneller interner Switch vorhanden ist, über ein virtuelles SAN, welches unter Mitwirkung des internen Verbindungsnetzwerkes realisiert wird.

5 Große Cluster

Wenn größere Cluster zum Einsatz kommen wird die Verwendung der SAN-Technologie sehr komplex. In unserem Fall werden insgesamt 37 Knoten zu einem Cluster zusammengeschaltet. Für die Plattenperipherie werden 8 FAS*StT* Storage Server verwendet. Will man diese über ein SAN mit allen Knoten verbinden, sind pro Knoten mindestens 16 Fibre-Channel-Anschlüsse notwendig, um die benötigte Performance zu erhalten. Das heißt es sind fast 600 Anschlüsse für die einzelnen Knoten notwendig. Damit wird die Switch-Konfiguration sehr unübersichtlich und nur sehr schwer verwaltbar. Normalerweise verfügen große Cluster schon über ein schnelles Netzwerk, welches zur Kommunikation zwischen den einzelnen Prozessen dient.

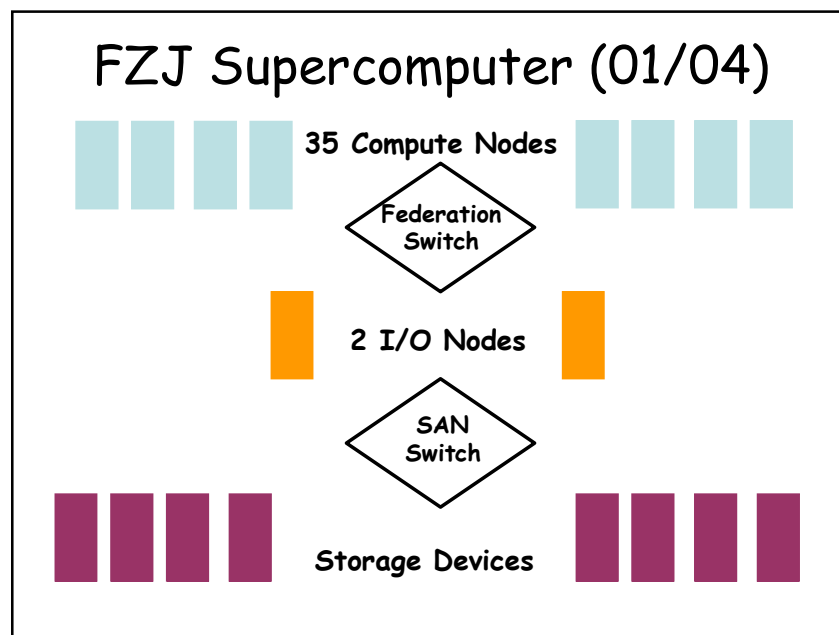


Abb. 4 Supercomputer im Forschungszentrum Jülich

Deshalb bietet es sich an, mit Hilfe dieses Netzwerkes ein virtuelles SAN zu realisieren und den Zugriff auf die Festplatten über dieses Netzwerk zu ermöglichen. Der physikalische Zugriff auf die Festplatten muss dann nur von wenigen Knoten aus möglich sein. Damit vereinfacht sich die Gesamtkonfiguration erheblich, wie die Abb. 4 zeigt. Sie stellt die Konfiguration für den Supercomputer dar, der im Forschungszentrum Jülich aufgebaut wird.

6 Virtual Shared Disk

Damit ein virtuelles SAN über einen internen Switch aufgebaut werden kann, muss eine entsprechende Softwareschicht eingezogen werden. Für die AIX-Systeme wird dafür von IBM die Virtual Shared Disk (VSD) Umgebung angeboten. Die VSD-Umgebung benutzt ein Client/Server-Protokoll welches über den Switch arbeitet. Das GPFS initiiert Lese- und Schreib-Anforderungen an den VSD-Client. Diese werden zum VSD-Server weitergeleitet, der den physikalischen Zugriff auf die Festplatte hat. Damit ist es einem Knoten möglich, einen lokalen I/O-Request zu einer Platte abzusetzen, die physikalisch an einen anderen Knoten angeschlossen ist. Als Protokoll kann entweder das IP-Protokoll oder ein proprietäres Protokoll (KLAPI) verwendet werden. Abb. 5 zeigt den prinzipiellen Aufbau einer VSD-Umgebung.

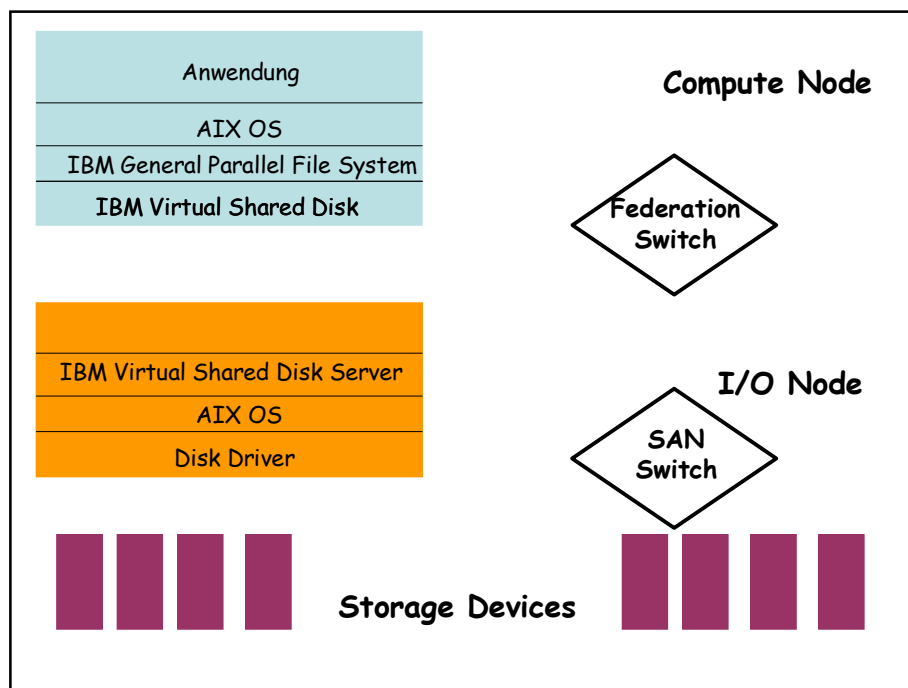


Abb. 5 Virtual Shared Disk Umgebung

Mobile Hotspots – Status Quo auf dem Weg zum mobilen Computing

Andy Schweiger

Lufthansa Systems Infratec GmbH
65451 Kelsterbach
E-Mail: Andy.Schweiger@Lhsystems.com

Vor dem Hintergrund eines breitbandigen Zugangs zur globalen Datenautobahn sind vor allem Hotspots von Interesse, da sie in der Regel auf standardisierten Zugangsverfahren zu IP-Netzen aufbauen und potenzielle Nutzer schneller mit Informationen versorgen als etwa analoge oder digitale Zugangsverfahren. Welche planerischen Aspekte sind bei der Netzeinrichtung zu beachten? Was kostet die breitbandige Verbindung am Hotspot? Welche Zugangstechnologien werden angeboten? Wo sind diese Services in Zukunft zu erwarten? Welche weiteren Entwicklungen sind in den nächsten Jahren zu erwarten? Der folgende Beitrag soll diese Fragen beantworten.

Hotspots lernen fliegen

Ein aktuelles Beispiel für einen Hotspot stammt von der Lufthansa. Mit einem breitbandigen Zugang von ca. 3 Mbit/s (ausbaufähig bis zu 20 Mbit/s) für den Downlink und derzeit 128 kbit/s (ausbaufähig bis 1,5 Mbit/s) für den Uplink ist es der Airline mit dem Kranich gelungen, bei ca. 900 km/h pro Stunde alle Passagiere eines Jumbos in 10.000 m Höhe mit Web Access zu versorgen. Planung, Konzeption und Implementierung des fliegenden Hotspots haben rund 18 Monate in Anspruch genommen. Vor dem Hintergrund der spezifischen Umgebung des Flugzeugs musste sich das beauftragte IT-Projektteam der Lufthansa Systems bei der Implementierung an offenen Standards wie dem Ethernet-Standard IEEE 802.3 oder Wireless-LAN-Protokollen IEEE 802.11b (Skalierungen auf 802.11a oder g sind in weiteren Planungen bereits berücksichtigt) orientieren. Nur so konnte sichergestellt werden, dass die proprietäre IT-Landschaft eines zivilen Flugzeugs durch ein weitgehend offenes und transparentes Kabinennetzwerk für knapp 400 Fluggäste ergänzt werden konnte. Der Transparenz im mobilen Netzwerk lag eine Kundenbefragung bezüglich der Technologie an Bord zu Grunde. Die Analyse gab auch Aufschluss über das zu erwartende Nutzungsverhalten: Kunden haben das vorrangige Ziel ortsunabhängig, sicher und schnell ins Internet zu gelangen, um E-Mails zu lesen und um arbeitsrelevante Informationen zu erhalten. Erst in

zweiter Linie sind für die Fluggäste weiterführende Informationen in Abhängigkeit von der Verweildauer am Hotspot relevant, wie sie beispielsweise ein Service-Portal mit Informationen rund um Flug und Zielgebiete liefert.

2 Wer fliegen kann, kann auch fahren

Die Ergebnisse des erfolgreichen Prototyp-Tests im Flugzeug lassen sich auch auf die Planung und Konzeption in andere Mobilitätsbereiche übertragen. Wichtig ist dabei die Bereitschaft, bei der Implementierung auf offene Systeme zu setzen. Die Ausrichtung an den bereits in Gebrauch befindlichen Systemen und Geräten der Enduser ist entscheidend für den Nutzwert. Wichtig für einen idealen Hotspot ist die Berücksichtigung von Sicherheitsaspekten. Der Muster-Hotspot sollte in der Lage sein, alle IP-fähigen Geräte entsprechend mit Datenverkehr zu versorgen, sämtliche Standardprotokolldienste wie http, ftp, smtp zu unterstützen und den Aufbau von gesicherten Verbindungen wie etwa IPsec- oder PPTP-Tunnel zur sicheren Kommunikation mit Drittnetzen ermöglichen.

Welche Gesichtspunkte sind beim Einsatz mobiler Technologien in bewegten Objekten noch zu beachten? Über die Struktur des Netzes am mobilen Hotspot hinaus ist insbesondere die Anbindung an den Rest der Internetwelt von zentraler Bedeutung. Art und erschlossene Bandbreite der Zugangstechnologie entscheiden hier im wesentlichen über das Surferlebnis der Nutzer. Beim Aufbau der weiterführenden Zugangstechnologien vom Access Point ins Internet unterscheiden sich Hotspots aufgrund divergierenden Anforderungen. So ist zum Beispiel bei der Versorgung eines Automobils mit wenigen Insassen eine geringere Bandbreite pro Fahrzeug notwendig als in Flugzeugen oder Eisenbahn-Wagons, die eine relativ hohe Nutzerzahl transportieren. Bei Bahn und Automobil wäre dabei besonders auf die Wetterresistenz der Funksignale zum Daten-Highway zu achten, da diese zwingende Voraussetzung für den unterbrechungsfreien Datenfluss zwischen Web und mobiler Einheit ist. Passagierflugzeuge wiederum verbringen einen Großteil der Systemeinsatzzeit über den Wolken. Damit können Satellitensignale aus dem geostationären Orbit weitgehend frei von Umwelteinflüssen empfangen werden. Für das Automobil und andere terrestrische Fortbewegungsmittel bieten sich daher entweder erdnähere Satellitendienste oder GSM/UMTS Multiplexdienste an. Ein Standard für breitbandige Datendienste hat sich derzeit – mit Ausnahme der Navigationssysteme – noch nicht durchgesetzt. Hohe Übertragungsgeschwindigkeiten stellen vor allem die Mobilfunkdienste noch vor Herausforderungen.

3 Kostenaspekte: Automatisierung und Abrechnung

Die Implementierung mobiler Zugangstechnologien erfordern einen erheblich höheren Kostenaufwand gegenüber verkabelten Access Points. Umso mehr ist bei den Überlegungen zur Schaffung mobiler Hotspots auf ein ausgewogenes Verhältnis zwischen Nutzungspotenzial, Implementierungskosten, Administration und den zu erzielenden Umsätzen zu achten. In diesem Spannungsfeld lohnt es sich, die Systeme zu optimieren, einen hohen Grad an Modularität anzustreben und vor allem die Administration der Geräte so weit wie möglich zu automatisieren. Gerade Letzteres ist im Netzwerkbereich der IT wenig verbreitet. Meist eignet sich die angebotene Hardware nicht für den vollautomatisierten Betrieb. Netzwerkdesigner und Spezialisten mit Know-how zur Schaffung von interventionsfreien Systemen haben hier die Nase vorn: Nicht jedes Auto, jeder Zug oder jedes Flugzeug kann schließlich mit einer eigenen Plattform zur Administration ausgestattet werden. Unrealistisch ist auch das Mitfahren oder Mitfliegen von IT-Administratoren an Bord mobiler Hotspots. Bestehende Sicherheitsrichtlinien bei den jeweiligen Transportmitteln erschweren zudem externe Zugriffe. Möglichst autonom arbeitende Systeme sind daher ein wichtiger Aspekt, um mobile Hotspots auch wirtschaftlich betreiben zu können.

Unbeantwortet ist derzeit die Frage nach der Abrechnung der genutzten Leistungen. Von Nischen-Accounts exotischer Provider oder der Abrechnung einer Flatfee via Kreditkarte abgesehen, existieren derzeit keine durchgängigen Abrechnungsverfahren. Globale Roaming-Verfahren zur Abrechnung über die weltweit heterogene Landschaft der ISPs sind vor allem vor dem Hintergrund unterschiedlicher rechtlicher Grundlagen der einzelnen Länder eine gewaltige Herausforderung. Geradezu kontraproduktiv scheint sich das Abstecken der „Claims“ der größten ISPs auszuwirken: Diese Unternehmen tun sich schwer damit, dass spezialisierte Carrier aufgrund der Besonderheiten des Zugangsverfahrens Schlüsseltechnologien besetzen und einen Marktvorsprung gewinnen. Hier wird allerdings der zunehmende Marktdruck durch die Endkunden entgegenwirken. Darüber hinaus sind in den nächsten Jahren großangelegte Roaming-Abkommen unter den führenden ISPs zu erwarten.

4 Ausblick

Die Implementierung mobiler IP-basierter Dienste ist spannend und hat Zukunft.

Unabhängig davon, welche Soft- und Hardwareprodukte oder Dienstleister die Aufgaben zur teil- oder vollautomatisierten Administration, der Weiterleitung von Authentifizierungsdaten oder der Abrechnung erledigen werden. Nachdem der erste mobile Hotspot in einem Passagierflugzeug der Lufthansa am 15. Januar gestartet ist, kann das Rad der Zeit nicht mehr zurückgedreht werden. Passagiere genießen jetzt das Surfen über den Wolken - und was sollte diese Passagiere zukünftig davon abhalten, ähnliche Datendienste in Bahn, Schiff, Bus und Auto zu nutzen?

Sponsoren

Wir danken den Firmen und Institutionen, die uns durch ihre finanzielle und materielle Hilfe bei der Ausrichtung der Tagung unterstützen.



Bisher sind erschienen:

**Modern Methods and Algorithms of Quantum Chemistry –
Proceedings**

Johannes Grotendorst (Hrsg.)

NIC-Serie Band 1

Winterschule, 21. - 25. Februar 2000, Forschungszentrum Jülich

ISBN 3-00-005618-1, Februar 2000, 562 Seiten

nicht mehr lieferbar

**Modern Methods and Algorithms of Quantum Chemistry –
Poster Presentations**

Johannes Grotendorst (Hrsg.)

NIC-Serie Band 2

Winterschule, 21. - 25. Februar 2000, Forschungszentrum Jülich

ISBN 3-00-005746-3, Februar 2000, 77 Seiten

nicht mehr lieferbar

**Modern Methods and Algorithms of Quantum Chemistry -
Proceedings, Second Edition**

Johannes Grotendorst (Hrsg.)

NIC-Serie Band 3

Winterschule, 21. - 25. Februar 2000, Forschungszentrum Jülich

ISBN 3-00-005834-6, Dezember 2000, 638 Seiten

**Nichtlineare Analyse raum-zeitlicher Aspekte der
hirnelektrischen Aktivität von Epilepsiepatienten**

Jochen Arnold

NIC-Serie Band 4

ISBN 3-00-006221-1, September 2000, 120 Seiten

**Elektron-Elektron-Wechselwirkung in Halbleitern:
Von hochkorrelierten kohärenten Anfangszuständen
zu inkohärentem Transport**

Reinhold Löwenich

NIC-Serie Band 5

ISBN 3-00-006329-3, August 2000, 145 Seiten

**Erkennung von Nichtlinearitäten und
wechselseitigen Abhängigkeiten in Zeitreihen**

Andreas Schmitz

NIC-Serie Band 6

ISBN 3-00-007871-1, Mai 2001, 142 Seiten

**Multiparadigm Programming with Object-Oriented Languages -
Proceedings**

Kei Davis, Yannis Smaragdakis, Jörg Striegnitz (Hrsg.)

NIC-Serie Band 7

Workshop MPOOL, 18. Mai 2001, Budapest

ISBN 3-00-007968-8, Juni 2001, 160 Seiten

**Europhysics Conference on Computational Physics -
Book of Abstracts**

Friedel Hossfeld, Kurt Binder (Hrsg.)

NIC-Serie Band 8

Konferenz, 5. - 8. September 2001, Aachen

ISBN 3-00-008236-0, September 2001, 500 Seiten

NIC Symposium 2001 - Proceedings

Horst Rollnik, Dietrich Wolf (Hrsg.)

NIC-Serie Band 9

Symposium, 5. - 6. Dezember 2001, Forschungszentrum Jülich

ISBN 3-00-009055-X, Mai 2002, 514 Seiten

**Quantum Simulations of Complex Many-Body Systems:
From Theory to Algorithms - Lecture Notes**

Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Hrsg.)

NIC-Serie Band 10

Winterschule, 25. Februar - 1. März 2002, Rolduc Conference Centre,
Kerkrade, Niederlande

ISBN 3-00-009057-6, Februar 2002, 548 Seiten

**Quantum Simulations of Complex Many-Body Systems:
From Theory to Algorithms - Poster Presentations**

Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Hrsg.)

NIC-Serie Band 11

Winterschule, 25. Februar - 1. März 2002, Rolduc Conference Centre,
Kerkrade, Niederlande

ISBN 3-00-009058-4, Februar 2002, 88 Seiten

**Strongly Disordered Quantum Spin Systems in Low Dimensions:
Numerical Study of Spin Chains, Spin Ladders and
Two-Dimensional Systems**

Yu-cheng Lin

NIC-Serie Band 12

ISBN 3-00-009056-8, Mai 2002, 131 Seiten

**Multiparadigm Programming with Object-Oriented Languages -
Proceedings**

Jörg Striegnitz, Kei Davis, Yannis Smaragdakis (Hrsg.)

Workshop MPOOL 2002, 11 Juni 2002, Malaga

NIC-Serie Band 13

ISBN 3-00-009099-1, Juni 2002, 133 Seiten

**Quantum Simulations of Complex Many-Body Systems:
From Theory to Algorithms - Audio-Visual Lecture Notes**

Johannes Grotendorst, Dominik Marx, Alejandro Muramatsu (Hrsg.)

NIC-Serie Band 14

Winterschule, 25. Februar - 1. März 2002, Rolduc Conference Centre,
Kerkrade, Niederlande

ISBN 3-00-010000-8, November 2002, DVD

Numerical Methods for Limit and Shakedown Analysis

Manfred Staat, Michael Heitzer (Hrsg.)

NIC-Serie Band 15

ISBN 3-00-010001-6, Februar 2003, 282 Seiten

**Design and Evaluation of a Bandwidth Broker that Provides
Network Quality of Service for Grid Applications**

Volker Sander

NIC-Serie Band 16

ISBN 3-00-010002-4, Februar 2003, 185 Seiten

**Automatic Performance Analysis on Parallel Computers with
SMP Nodes**

Felix Wolf

NIC-Serie Band 17

ISBN 3-00-010003-2, Februar 2003, 143 Seiten

**Haptisches Rendern zum Einpassen von hochaufgelösten
Molekülstrukturdaten in niedrigaufgelöste
Elektronenmikroskopie-Dichteverteilungen**

Stefan Birmanns

NIC-Serie Band 18

ISBN 3-00-010004-0, September 2003, 153 Seiten

Alle Bände stehen online zur Verfügung unter
<http://www.fz-juelich.de/nic-series/>.