

Rechnerverbund NRW

Arbeitskreis der Leiter Wissenschaftlicher Rechenzentren in NRW (ARNW)

Autoren:

Christian Bischof, Rechenzentrum, Aachen

Klaus Brühl, Rechenzentrum, Aachen

Thomas Eifert, Rechenzentrum, Aachen

Wilhelm Held, Zentrum für Informationsverarbeitung, Münster

Jürgen Hölters, Zentrum für Informationsverarbeitung, Münster

Bruno Lix, Hochschulrechenzentrum, Essen

Klaus-Bolko Mertz, Zentrum für Informationsverarbeitung, Münster

Stefan Ost, Zentrum für Informationsverarbeitung, Münster

Günter Schwichtenberg, Hochschulrechenzentrum, Dortmund

Benno Süselbeck, Zentrum für Informationsverarbeitung, Münster

Harald Ziegler, Hochschulrechenzentrum, Dortmund

Dezember 2000

<http://www.arnw.de/docs/rvnrw/> und <http://www.wissenschaft.nrw.de/rvnrw/>

Zusammenfassung

Im Rechnerverbund NRW wurde ein einfach nutzbarer, leistungsfähiger Verbund von IT-Ressourcen eingerichtet, der derzeit konkurrenzlos ist. Die Benutzerverwaltung ist landeseinheitlich, die Daten sind automatisch dort, wo sie benötigt werden und die Datenstrukturen sind wohl definiert. Die Ressourcen können landesweit in Anspruch genommen werden. Die weitgehende Ortsunabhängigkeit der Server, die bisher schon innerhalb einer Hochschule gegeben war, ist damit auf das Land NRW ausgedehnt worden.

Der Rechnerverbund NRW ist zu sehen vor dem Hintergrund der zunehmenden Ansprüche der Benutzer in den Hochschulen. Diese verlangen zuverlässige Mail-, Web-, Rechen-, und Datendienste (z.B. Fileserver, Archiv- und Backupserver) sowie Beratung bei der Konfiguration der i.a. dezentral mit Webschnittstellen zu verwaltenden Dienste. Darüber hinaus führt der ansteigende Einsatz von Simulationswerkzeugen zu einer stetig wachsenden Nachfrage an Anwenderberatung, Schulung und methodischer Unterstützung. Dies erfordert aufgrund der personellen und finanziellen Restriktionen eine andere Organisationsform, welche über die traditionellen Universitätsgrenzen hinausgeht. Mit dieser Organisation der Ressourcen ergänzen sich die Dienstleistungen der einzelnen Hochschulrechenzentren zu einem neuen Ganzen und verbessern das Niveau der IT-Versorgung damit deutlich.

Aufgaben, die in einzelnen Hochschulrechenzentren gleichermaßen anstehen, können arbeitsteilig gelöst werden, um so die dringende Weiterentwicklung der IT-Infrastrukturen und IT-Anwendungen voran zu bringen. 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.

Die mittlerweile verfügbare Softwaretechnologie erlaubt es, einen solchen landesweiten Verbund in die schon vorhandenen IT-Infrastrukturen der beteiligten Hochschulrechenzentren mit der für ein solches Projekt notwendigen Zuverlässigkeit einzubinden und mit der notwendigen Dienstgüte unter Produktionsbedingungen zu betreiben.

Inhalt

1	ANLASS UND NEUE ZIELE.....	5
1.1	BESTANDSERHEBUNG HOCH- UND HÖCHSTLEISTUNGSRECHNER	5
1.2	INNOVATIVE, FLEXIBLE UND NUTZERNAHE RECHNERPYRAMIDE.....	5
1.3	HOCHLEISTUNGSRECHNER BEI HETEROGENER AUSSTATTUNG	6
1.4	LÖSUNGSANSATZ „VERTEILTES SYSTEM“	6
1.5	NEUE QUALITÄT DER KOOPERATION	7
2	TECHNIK, SYSTEMPFLEGE UND ORGANISATION	8
2.1	ÜBERBLICK.....	8
2.2	RANDBEDINGUNGEN.....	9
2.3	DCE/DFS ALS TECHNISCHE GRUNDLAGE	9
2.4	VERTEILTES LANDESRECHENZENTRUM.....	10
2.5	DIE BENUTZERVERWALTUNG	11
2.6	SOFTWAREINSTALLATION.....	12
3	ANWENDUNGEN.....	14
3.1	HOCHLEISTUNGSRECHNEN	14
3.2	DATENSICHERUNG.....	14
3.3	MILESS	15
3.4	SMART-KARTEN	16
3.5	PUBLIC-DOMAIN-SOFTWARE.....	17
3.6	KOMPETENZZENTRUM FÜR STATISTIK-SOFTWARE	17
3.7	VISUALISIERUNG	18
3.8	BERATUNG, VOIP, UMS, CALL CENTER UND TROUBLE TICKET.....	18
3.9	AUS- UND WEITERBILDUNG	19
3.10	ZUKÜNFTIGE MÖGLICHKEITEN	19
4	KOOPERATION UND KONKURRENZ	19
5	ANHANG	20
5.1	LEISTUNGSFÄHIGE RECHENSYSTEME IN DEUTSCHLAND	20
5.2	ERLÄUTERUNGEN ZU DCE/DFS	21
5.3	SOFTWAREINSTALLATION.....	22

1 Anlass und neue Ziele

1.1 Bestandserhebung Hoch- und Höchstleistungsrechner

Der Arbeitskreis der Leiter Wissenschaftlicher Rechenzentren in NRW (ARNW) hat durch besondere Mitwirkung der Hochschulrechenzentren (HRZ) in Aachen, Dortmund und Münster vor einiger Zeit die Hoch- und Höchstleistungsrechnerausstattung in den Universitäten in NRW beschrieben. Die Ergebnisse lassen sich als dringender Handlungsbedarf wie folgt zusammenfassen:

Im wissenschaftlich-technischen Bereich gewinnt die Informationsverarbeitung gegenwärtig immer mehr an Bedeutung. Neben den klassischen Bereichen Theorie und Experiment entwickelt sich die numerische Simulation zum dritten Standbein der Forschung. Die hierfür benötigte maschinelle Ausstattung besteht zum einen aus PCs oder Workstations am Arbeitsplatz der Forscherinnen und Forscher, aus Servern in den Instituten, zentralen Hochleistungsrechnern in der Hochschule und bundesweit zugänglichen Höchstleistungsrechnern.

Für die meist in den HRZ installierten Hochleistungsrechner erfolgt eine Bedarfsanpassung in verhältnismäßig großen zeitlichen Abständen, so dass regelmäßig Mangelsituationen bezüglich der verfügbaren Rechenleistung auftreten. Diese Situation wird zudem noch durch die extrem schnellen Entwicklungszyklen im IT-Bereich verschärft. Im internationalen Maßstab wächst die installierte Rechenleistung alle 18 Monate um den Faktor 2,5. Daraus resultiert in den Forschungsdisziplinen, die auf leistungsstarke Rechnersysteme angewiesen sind, eine beträchtliche Verschiebung der Wettbewerbsfähigkeit, auch der Hochschulen untereinander, wenn sie diesem Wachstum nicht folgen können.

Die Tabelle im Anhang 5.1 zeigt die Leistungsfähigkeit der an deutschen Rechenzentren aus Industrie und Forschung installierten Rechnersysteme im nationalen und internationalen Vergleich. Bemerkenswert ist, dass sich nur ein einziges an einer Hochschule in NRW installiertes System unter den 500 leistungsstärksten Systemen befindet, dagegen die Hochschulen anderer Bundesländer eine massivere Präsenz zeigen.

Die Anzahl der Hochleistungsrechner im industriellen Umfeld ist in den letzten Jahren ebenfalls kontinuierlich gewachsen; die Akzeptanz solcher Systeme steigt hier ständig. Waren die ersten Supercomputer oft im universitären Umfeld installiert, hat sich dieses inzwischen erheblich geändert. Die apparativen Ausstattungen der Hochschulen haben hier mit der allgemeinen Entwicklung nicht Schritt gehalten.

Für die Hochschulen in NRW besteht daher dringender Handlungsbedarf, um im nationalen und internationalen Wettbewerb nicht ins Hintertreffen zu geraten und somit auch in der Einwerbung von Drittmitteln nicht mehr mit der Konkurrenz mithalten zu können.

1.2 Innovative, flexible und nutzernahe Rechnerpyramide

Die HRZ in NRW haben als Antwort auf die Mangelsituation 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 (Innovationsspirale bezüglich der IT-Ausstattung). Der Verbund versteht sich allerdings nicht als Ersatz für die Bundes-

Höchstleistungsrechner sondern als eine sinnvolle Ergänzung. Diese Höchstleistungssysteme sollen weiterhin durch Projekte aus den Hochschulen in NRW mit extrem hohen Rechenzeitbedarf genutzt werden.

Ausgangspunkt der Überlegungen für den Rechnerverbund ist die Abdeckung des Rechenbedarfs der Hochschulen entsprechend der „Rechnerpyramide“ (siehe Abb. 1) in den Bereichen Grund-, Hochleistungs- und Höchstleistungsversorgung. Im Bereich der hohen Rechenleistung ist angedacht, dass in jeder Hochschule im gegenseitigen Austausch ca. 20% der Rechenleistung für den Verbund zur Verfügung gestellt werden.

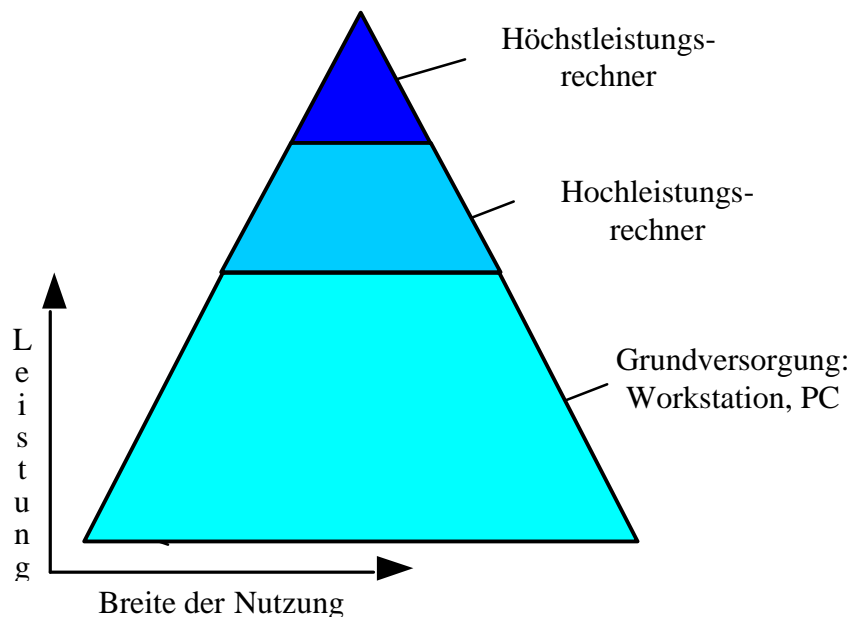


Abb. 1: Abdeckung des Rechnerbedarfs (Rechnerpyramide)

1.3 Hochleistungsrechner bei heterogener Ausstattung

Die Nutzung der verschiedenen Systeme soll für die Anwender vollständig transparent erfolgen und die Nutzungserlaubnis für die verfügbaren Ressourcen soll vom jeweiligen Heimat-Rechenzentrum des Nutzers geregelt werden. Jedem Nutzer des RV-NRW steht über ein Computing-Portal potentiell der gesamte RV-NRW zur Verfügung, so dass er die günstigste Plattform entsprechend seiner Erfordernisse wählen kann. Dies gewinnt noch dadurch an Bedeutung, dass durch die heterogene und an den lokalen Erfordernissen der jeweiligen Hochschule ausgerichtete Infrastruktur der teilnehmenden HRZ die geeignete Hard- und Software mit höherer Wahrscheinlichkeit im Verbund als an der eigenen Universität verfügbar ist.

Wichtig in diesem Zusammenhang ist, dass die HRZ ihre lokale maschinelle und softwaremäßige Ausstattung entsprechend der lokalen Notwendigkeiten eigenverantwortlich durchführen können und sollen. Durch die unterschiedlichen Schwerpunktsetzungen der Themen in den Hochschulen ist ein breites Spektrum zu erwarten bzw. kann ein gegenseitiger Austausch von Schwerpunktaktivitäten erwogen werden.

1.4 Lösungsansatz „Verteiltes System“

Das Projekt RV-NRW ist als Lösungsansatz ein „Verteiltes System“, das von vornherein praxisnah unter Produktionsgesichtspunkten und nicht als reines Forschungsprojekt konzipiert wurde. Das Konzept entstand durch Nachdenken vor allem in den HRZ in Aachen und Müns-

ter, aber auch durch kompetente Mitwirkung der HRZ in Dortmund und Essen. Eine der Hauptforderungen des Projektes war, dass in den am Verbund beteiligten HRZ bis auf die Anbindung eines zusätzlichen Gatewayrechners keine Änderungen der bestehenden Produktionsumgebung vorausgesetzt wird. Hierdurch unterscheidet sich dieser Ansatz von solchen, die gegenwärtig an anderen Stellen durchgeführt werden. Der praxisnahe Ansatz - für die Nutzer und die betreibenden HRZ - hat dazu geführt, dass das Konzept in mehreren Gremien, wie zum Beispiel dem DV-Infrastrukturausschuss für NRW¹ und dem WAL² sehr positiv bewertet wurde. Bei den übrigen HRZ des Landes ist die eindeutige Tendenz zu erkennen, dem initiierten Verbund beizutreten.

1.5 Neue Qualität der Kooperation

Das flexible und moderne Konzept des RV-NRW geht über das ursprünglich ins Auge gefasste Konzept des Hochleistungsrechnens deutlich hinaus. Der gewählte Ansatz ermöglicht in vielen anderen Themenfeldern eine viel weitergehende Zusammenarbeit der HRZ des Landes untereinander. Derartige Ideen zur Kooperation werden schon seit mehreren Jahren vom HRZ in Dortmund verfolgt. Nunmehr besteht eine neue technische Grundlage, um diese Zusammenarbeit weiter zu forcieren. Vor dem Hintergrund einer permanenten Personalnot bei steigender Vielfalt der Anforderungen können neue Möglichkeiten erschlossen werden. Nicht essentielle Aufgaben, für die nicht zwingend in einem HRZ Kompetenz vorhanden sein muss, können im Rahmen eines „Verteilten Rechenzentrums NRW“ delegiert werden. Durch die große Vielfalt der in einem HRZ anfallenden Aufgaben werden Mitarbeiter meist mit mehreren Aufgaben betraut. Viele Tätigkeiten werden zzt. nicht mehr agierend, sondern nur auf aktuelle Anforderungen oder Probleme reagierend durchgeführt. Dies führt oft dazu, dass viele Tätigkeiten nur mehr oder weniger oberflächlich ausgeführt werden können. Im Rahmen eines kooperativen Verbundes können die überall gleichermaßen anfallenden Aufgaben auf einen größeren Personenkreis verteilt werden. Dies bewirkt unter anderem eine höhere Verfügbarkeit von Diensten, da z.B. während einer Ausfallzeit eines Mitarbeiters in einem Zentrum die Betreuung der Software bzw. die Beratung von Anwendern durch einen Mitarbeiter eines anderen Rechenzentrums durchgeführt werden könnte. Durch die Vermeidung der Replikation von Arbeiten der HRZ in einem kooperierenden Verbund ist eine verstärkte Kompetenzbündelung bei den Mitarbeitern möglich, die sich neben einer erhöhten Dienstqualität auch positiv auf die Motivation der Mitarbeiter auswirkt. Endlich finden sie Partner, mit denen sie ihre Probleme diskutieren und gemeinsam lösen können. Der Erfahrungsaustausch wird damit auf eine neue Stufe gestellt. Der Mehraufwand, der sich hochschulübergreifend durch die Betreuung größerer Anzahlen ergibt, ist verglichen mit der grundsätzlichen Beschäftigung mit einem Thema an sich eher kleiner. Vor diesem Ziel ist zuerst allerdings eine Übergangszeit zu überbrücken, die zusätzliche Belastungen für die HRZ bringen wird.

¹ Eine Arbeitsgruppe, in der je 2 Rektoren/Rektorinnen und Kanzler/Kanzlerinnen der Hochschulen, je 1 Rektor/Rektorin und Kanzler/Kanzlerin der Fachhochschulen, der/die Vorsitzende des WAL, die Sprecher des ARNW und der DV-Zentralen der Fachhochschulen, der Leiter/die Leiterin der Netzagentur und das MSWF vertreten sind.

² WAL = Wissenschaftlicher Ausschuss zur Kapazitätsverteilung der Landeshöchstleistungsrechner; später hat sich der WAL anders entwickelt

2 Technik, Systempflege und Organisation

Der Rechnerverbund wurde aus der Sicht der Hochleistungsrechner konzipiert. Das bedingt eine gewisse Unix-Lastigkeit des Konzepts. Windows-Nutzer werden allerdings nicht ausgeschlossen, sondern ausdrücklich eingebunden.

2.1 Überblick

Die Dienste, welche die HRZ erbringen müssen, kann man in der Grobübersicht (Abb. 3) erkennen. Wenn nun jedes HRZ im RV-NRW einen Teil seiner Dienste in diesen Verbund einbringt, ergibt sich ein neues verteiltes HRZ (Abb. 2). Ein derartiger Verbund lässt sich natürlich auch auf die IT-Dienste der Fachbereiche übertragen. Dies soll aus pragmatischen Gründen bei Bedarf in einem zweiten Schritt geschehen. Die jeweiligen HRZ werden dabei Hilfestellung leisten.

HRZ-Dienste
Kommunikationssysteme
Rechner u. Betriebssysteme
Anwendungen u. Tools
Daten u. Informationen
Aus- u. Weiterbildung
Organisatorisches
IT-Weiterentwicklungen
Personen u. Know-how

Abb. 3: HRZ-Dienste

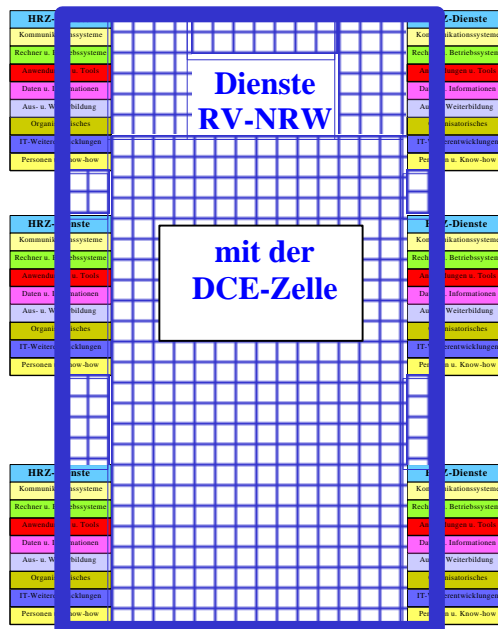


Abb. 2: Dienste des RV-NRW als Summe der HRZ-Anteile

Der Kern des RV-NRW wird durch eine unten beschriebene DCE-Zelle (DCE = Distributed Computing Environment) gebildet, an die mittels Gateways IT-Ressourcen der einzelnen HRZ angekoppelt werden können. Der RV-NRW umfasst aber mehr als DCE-Zelle, Gateways und angekoppelte Ressourcen. Es handelt sich vor allem auch um einen Know-how-, Daten- und Software-Verbund. Mit DCE können Ressourcen sicher zwischen mehreren Zellen, d.h. in diesem Falle den Rechenzentren verteilt werden. Die für den Verbund wichtigste Anwendung ist DFS (Distributed File System), das eine verteilte Datenhaltung erlaubt, bei der die Daten entsprechend der Erfordernisse automatisch von einer Stelle zur anderen verschoben werden.

2.2 Randbedingungen

Verteiltes Dateisystem

Voraussetzung für den RV-NRW ist ein verteiltes Dateisystem. Dabei musste in dem landesweiten Nutzer- und Datenverbund die historisch gewachsene IT-Landschaft der HRZ respektiert und in das Konzept eingebunden werden. Traditionell benutzen Unix-Systeme NFS³ als verteiltes Dateisystem. Leider kommt NFS wegen eklatanter Sicherheitsprobleme für ein landesweites Dateisystem nicht in Frage. Viele Universitäten blockieren wegen dieser Probleme z.B. schon in ihren Eingangs-Routern die für NFS benötigten RPC-Ports⁴. Gesucht war also ein landesweites Dateisystem mit akzeptabler Sicherheit, das zusätzlich einen problemlosen Übergang in die jeweiligen lokalen Dateisysteme ermöglicht.

Gemeinsame Benutzer- und Gruppenkennungen

Häufig benutzen Unix-Systeme NIS⁵ zur Benutzerverwaltung und Authentifizierung. Dabei ist NIS sicherheitstechnisch noch problematischer als NFS. NIS schied damit ebenfalls für eine landesweite Lösung aus.

Nur wenige Universitäten haben ein einheitliches System von Benutzerkennungen. Um so unwahrscheinlicher erschien die landesweite Vereinheitlichung aller Benutzerkennungen. Im RV-NRW wurden deshalb neue Benutzerkennungen eingeführt, die einen eigenen, vollständig neuen Namensraum bilden, mit der Konsequenz, dass jeder Benutzer des RV-NRW zwei Kennungen besitzt: Seine lokale und die des Rechnerverbundes. Entsprechend wurden auch eigene Gruppennamen sowie eigene Bereiche für die numerischen Kennungen der Benutzer und Gruppen vereinbart, über die in UNIX-Systemen die Zugriffsrechte auf Dateien gesteuert werden.

2.3 DCE/DFS als technische Grundlage

Basierend auf den guten Erfahrungen, die sowohl die Universität Münster als auch das NRW-Metacomputing-Projekt⁶ gemacht haben, wurde DCE/DFS⁷ als technische Grundlage des RV-NRW ausgewählt. DCE ist die Grundlage für das verteilte Dateisystem DFS. DCE ist eine Entwicklung der OSF⁸, eines Zusammenschlusses namhafter Firmen wie IBM, DEC, Siemens und anderer, und steht als Technik seit etwa 10 Jahren zur Verfügung. Es besteht aus den Komponenten DCE-RPCs, Cell Directory Service, Security Service, Distributed Time Service und Distributed File System, die im Anhang (5.2) näher erläutert werden.

³ NFS = Network File System

⁴ RPC = remote procedure call. Ein von SUN für NFS entwickeltes Kommunikations-Protokoll für entfernte Funktionsaufrufe von Server-Diensten.

⁵ NIS = Network Information Service, a service that provides information, that has to be known throughout the network, to all machines on the network.

⁶ NRW-Metacomputing-Projekt = Im Februar 1995 wurde der Forschungsverbund NRW-Metacomputing „Verteiltes Höchstleistungsrechnen“ gegründet. In diesem Verbund hatten sich die Universitäten Aachen, Bochum, Dortmund, Köln und Paderborn mit der GMD (St. Augustin), der DLR (Köln) und der KFA (Jülich) zu einer offenen Forschungs Kooperation zusammengeschlossen.

⁷ DCE = Distributed Computing Environment, DFS = Distributed File System. DFS ist sehr viel älter als das zufällig namensgleiche verteilte Dateisystem für Microsoft-Betriebssystem und hat mit diesem leider nichts zu tun.

⁸ OSF = Open System Foundation

2.4 Verteiltes Landesrechenzentrum

Die landesweite DCE-Zelle **rv-nrw.de** (Abb. 4) erfordert kein zentrales Landesrechenzentrum, das real existiert. Man könnte diese Zelle aber als „verteiltes Landesrechenzentrum“ bezeichnen. In den Verbund können einzelne HRZ weitere Ressourcen einbringen, die dann zusammen mit der DCE-Zelle den RV-NRW bilden. Alle relevanten DCE/DFS-Server-Prozesse der Zelle können prinzipiell in jedem Rechenzentrum repliziert werden. Die Summe aller Server-Prozesse realisiert die Zelle in Form der Kopfstationen und somit wird die Zelle von allen beteiligten Rechenzentren verteilt realisiert. Die über das Gateway der DCE-Zelle eingebrachten Ressourcen können vollständig dem RV-NRW zugewiesen werden. Sie können aber auch eine Zwitterstellung einnehmen; sie werden in diesem Fall von lokalen Benutzern mit der lokalen Benutzererkennung und aber eben auch von Benutzern des RV-NRW mit der Verbunderkennung ansprechbar sein.

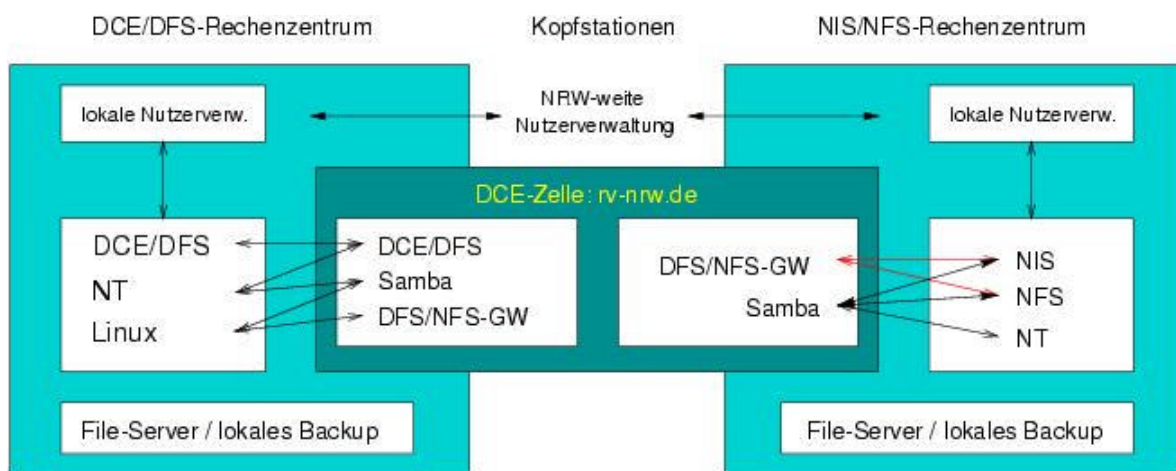


Abb. 4: Struktur der DCE-Zelle *rv-nrw.de*

Die Pfeile zeigen die möglichen bzw. erforderlichen Übergänge zwischen Zelle und lokalen Gegebenheiten an. So muss eine Abbildung der Benutzer des RV-NRW in die lokale Benutzerverwaltung erfolgen, denn ohne lokalen Benutzereintrag wäre die „ferne“ Rechnernutzung unmöglich (s.u.). Die zweite Pfeilgruppe bezeichnet die Möglichkeiten, von verschiedenen Betriebssystemen aus auf die Daten des Rechnerverbundes zuzugreifen.

Bemerkenswert ist, dass auf keinem lokalen Rechner eine besondere Software installiert werden muss. Alle Dateizugriffe auf die Landeszelle sind mit den „natürlichen“ Methoden der Betriebssysteme möglich.

Der Übergang vom DFS in andere Netzwerkdateisysteme wird – wie oben schon gesagt – durch Gateways realisiert. Das Gateway zum NFS gehört zum Standard-Lieferumfang der eingesetzten DCE-Software. Für das Gateway zum SMB⁹, über das Windows-Systeme in all ihren Varianten und auch Linux (alternativ zum auch möglichen NFS) erreicht werden, wurde das wohlbekannte Programm Samba⁹ von der Universität Münster um die DCE-Autorisierung erweitert. Auf diese Weise sind als Randprodukt selbst noch die in die Jahre

⁹ SMB (Server Message Block) ist ein Netzwerk-Dateiprotokoll, zu dem viele Firmen beigetragen haben, unter anderem Xerox, 3Com und vor allem Microsoft. Es gibt zu SMB ein X/Open Dokument. Microsoft hat das Protokoll allerdings nach seinen eigenen Bedürfnissen erweitert. Es gibt einen SMB-Server (heute als Samba bekannt), der von Andrew Tridgell begonnen wurde. Samba wurde immer populärer und hat heute einen durchaus signifikanten Anteil am SMB-Server Markt.

gekommenen Systeme LAN-Manager, OS/2 und DOS einbindbar. Die dazu notwendigen Zusatzinformationen werden in einer Security-Datenbank abgelegt und stehen damit landesweit zur Verfügung. Damit können prinzipiell beliebig viele Samba-Gateways an beliebig vielen Standorten betrieben werden.

Einheitliche Kopf-Stationen

Kopfstationen stellen für die Benutzer des RV-NRW die Verbindung zu den Plattenkapazitäten der eigenen Universität bereit. Hinzu kommt ein Anteil, der für die Software-Bereitstellung benötigt wird. Schon sehr früh zeigte sich bei der Realisierung der DCE-Zelle die Notwendigkeit, einen einheitlichen Wartungsstand des Betriebssystems und der DCE-Software zu erhalten. Deshalb wurde beschlossen, die Kopfstationen von einer Stelle aus einheitlich zu warten. Alle Kopfstationen sollen Unix-Rechner vom Typ IBM RS/6000 mit dem Betriebssystem AIX und der DCE-Software der Firma IBM sein¹⁰. HRZ mit weniger Erfahrung beim Einsatz von IBM-Rechnern wird eine Starthilfe und Beratung angeboten.

Betrieb

Betrieben wird die DCE-Zelle von allen beteiligten Rechenzentren gemeinsam im Sinne gegenseitigen Vertrauens. Ein an der Universität Münster entwickeltes grafisches Betriebsüberwachungsprogramm „dcemon“ erlaubt, sich jederzeit einen Überblick über den Betriebszustand der Zelle zu verschaffen. Wir können auf diese Weise einen stabilen Betrieb gewährleisten, der außerdem ebenso sicher ist oder sicher zu machen ist wie ein Betrieb innerhalb einer Universität.

Software-Wartung oder andere gravierende Eingriffe in die Rechner-Konfiguration benötigen eine gegenseitige Abstimmung.

2.5 Die Benutzerverwaltung

Anmeldung und Benutzer-Administration

Die Zulassung von Benutzern zum RV-NRW erfolgt auf der Basis des gegenseitigen Vertrauens der beteiligten HRZ. Dazu wurde festgelegt, dass jedes HRZ für die Zulassung seiner Benutzer zu allen dem RV-NRW zur Verfügung stehenden Ressourcen verantwortlich und berechtigt ist. Zu diesem Zweck muss jede der beteiligten Einrichtungen mindestens eine Person als lokalen Benutzer-Administrator benennen, die das Recht hat, jeden Angehörigen seiner Hochschule als Benutzer des RV-NRW anzumelden und ihn in die gewünschten Benutzergruppen einzutragen. Damit jedem Benutzer mit seiner RV-Kennung alle Ressourcen des RV-NRW zur Verfügung stehen, müssen die Kennungen und Nummern für Benutzer und Gruppen sowohl in der gemeinsamen DCE-Zelle als auch lokal eingetragen sein. Für den Eintrag in die lokalen IT-Systeme ist ebenfalls der lokale Benutzer-Administrator zuständig.

Benutzerdatenbank

Die Zuordnung von Benutzerkennungen zu Benutzergruppen und die damit beschriebene Erteilung von Zugangsberechtigungen ist natürlich auf jeden Fall befristet. Verwaltung und Kontrolle dieser Zuordnungen erfolgen über eine Datenbank. Zur Zeit wird dafür die zentrale Benutzerdatenbank des Zentrums für Informationsverarbeitung der Universität Münster mitbenutzt, mit deren Verfahren ähnliche Verwaltungsaufgaben seit mehreren Jahren bearbeitet werden. Eine eigenständige Benutzerdatenbank für den RV-NRW könnte mit der in Münster

¹⁰ Der Preis einer Kopfstation liegt derzeit je nach Ausstattung zwischen 15.000,- und 20.000,- DM.

vorhandenen Software ohne große Probleme eingerichtet werden, wenn die Anzahl der Benutzer dies später erfordern sollte.

Übergabe und Auswertung der Informationen

Die lokalen Benutzer-Administratoren können die von ihnen zu vertretenden Eintragungen und Änderungen in der Datenbank entweder direkt vornehmen oder in einer vereinbarten Form in besonders gesicherten DFS-Dateien des gemeinsamen Systems ablegen, die regelmäßig durch den Datenbank-Administrator oder eine automatisch gestartete Prozedur kontrolliert und in die entsprechenden Änderungen der Datenbank umgesetzt werden.

Auf der anderen Seite haben Änderungen in der Datenbank in aller Regel Konsequenzen für die mit ihrer Hilfe verwalteten IT-Systeme. Diese Konsequenzen werden durch die Datenbank-Programme sofort berechnet und in einer standardisierten Form in der Datenbank niedergelegt. Sie in den betroffenen IT-Systemen umzusetzen und den Erfolg der Umsetzung wiederum in der Datenbank zu dokumentieren, ist Aufgabe der lokalen Benutzer-Administratoren bzw. des Administrators des gemeinsamen IT-Systems. Diese Aufgabe wird in der Regel durch automatisch gestartete Prozeduren erfüllt, die dazu entweder direkt auf den entsprechenden Teil der Datenbank zugreifen können oder, soweit dies vereinbart ist, Dateien im gemeinsamen DFS-System benutzen, die hier regelmäßig durch Auswertung der Datenbank-Einträge erzeugt und selbstverständlich ebenfalls besonders gesichert werden.

2.6 Softwareinstallation

Die Softwareverteilung im RV-NRW basiert auf Erfahrungen, die innerhalb der Universitäten Aachen und Münster beim Aufbau und Betrieb eines heterogenen Unix-Clusters seit 1991 gemacht wurden.

Voraussetzungen und Ziele

- Ausgangspunkt der Überlegungen war ein RV-NRW mit einer großen Anzahl von Rechnern mit einer unterschiedlichen Architektur und einer uneinheitlichen Verwaltung.

Ziele eines Konzepts zur Versorgung eines solchen Rechnerverbundes mit Anwendungssoftware waren:

- Es sollte hohe Akzeptanz durch den Endbenutzer erreicht werden. Der Anwender wird eine zentrale Bereitstellung von Anwendungssoftware nur dann akzeptieren, wenn er das Gefühl hat, es selbst nicht besser machen zu können. Im Einzelnen bedeutet dies, dass die Software jederzeit effizient verfügbar sein muss und dass sie bequem aufzurufen ist.
- Der Zugang zur Software ist gut zu dokumentieren. Neue Versionen von Produkten sollten nach Erscheinen rasch zur Verfügung stehen. Im Idealfall merkt der Anwender nicht, dass die Software nicht lokal installiert ist.
- Der Aufwand für lokale Systemverwalter muss gering sein. Für den Systemverwalter eines lokalen Systems muss es leicht möglich sein, seine Maschinen in das System RV-NRW der Softwareverteilung einzuhängen. Die dazu notwendigen Mechanismen sollten einen einmaligen Vorgang darstellen, der sich automatisieren lässt.
- Der Aufwand für den Softwareverwalter muss gering sein. Die Software soll im Netz genau einmal installiert werden. Alle Maßnahmen zur Wartung werden an zentraler Stelle vorgenommen.
- Die Kosten sollen niedrig sein und möglichst noch reduziert werden. Durch die zentrale Installation von Software lassen sich Lizenzen und Personalkosten sparen.

Auf den ersten Blick scheinen die Ziele unter den gegebenen Voraussetzungen nicht erfüllbar zu sein. Sie wurden im RV-NRW jedoch in erfolgreicher Kooperation alle erreicht. Mit dem im Anhang (5.3) ausführlicher vorgestellten Konzept werden möglich:

- Eine Kostenersparnis durch eine landesweit zentrale Installation der Anwendungssoftware ist erreichbar, da in der Regel insgesamt weniger Lizenzen benötigt werden.
- Eine Erweiterung des Produktspektrums, das den Benutzern zur Verfügung steht, wird gegeben sein.
- Die arbeitsintensive Bereitstellung von Public-Domain-Software erfordert weniger Personalaufwand.
- Die Arbeit für den Softwareverwalter wird reduziert. Der Softwareverwalter kann mit nahezu gleichem Aufwand ein Produkt nicht nur universitätsweit, sondern auch landesweit zur Verfügung stellen.
- Durch Arbeitsteilung sind eine größere Spezialisierung und höhere Qualität zu erwarten.
- Der Aufwand für die lokalen Systemverwalter wird verringert. Der lokale Systemverwalter muss lediglich in den Startup-Skripts seiner Maschine ein Dateisystem montieren. Dies ist eine automatisierbare Aufgabe. Zusätzlich sind (einmalig!) einige symbolische Links und Pfadeinstellungen vorzunehmen.
- Für den Anwender wird der Zugang zur Software bequem. Der Anwender muss die Software nur noch aufrufen, egal auf welcher Maschine er arbeitet.

Für jedes Produkt oder jede Produktgruppe wird ein Softwareverwalter (Paketverantwortlicher) landesweit bestimmt, der die ihm übertragene Softwarekomponente vollständig verwaltet. Dazu gehören u.a. folgende Aufgaben:

- Beschaffung und Weiterführung des Produkts durch Abschluss von Lizenz- und Wartungsverträgen. Pflege der kommerziellen Kontakte zum Hersteller. Ermittlung des Bedarfs im Rechnerverbund.
- Installation und Wartung der Software für alle unterstützten Architekturen. Kontakt zum technischen Support der Hersteller.
- Entgegennahme von Fehlermeldungen seitens der Anwender.
- Dokumentation des Zugriffs auf die Software, einschließlich der technischen Voraussetzungen für die unterstützten Architekturen.

Um einen reibungslosen Ablauf der Aufgaben zu ermöglichen, muss jeder Paket-Verantwortliche einen Vertreter (möglichst aus dem gleichen Umfeld) besitzen.

Software-Verträge

Das Konzept eines unter Produktionsbedingungen operierenden landesweiten „Verteilten Rechenzentrums“ stellt zzt. auch für die meisten Software-Firmen noch Neuland dar und führt zu sehr unterschiedliche Reaktionen. Rein technisch ist eine Steuerung der Nutzungsrechte durch die Verwendung eines Lizenzservers oder durch die unter DCE möglichen strikten Zugriffsregelungen in Verbindung mit örtlichen Lizenzregelungen vergleichsweise problemlos möglich. Dennoch bestehen hier bei einigen Firmen noch prinzipielle Bedenken bezüglich einer Redistribution der Software. Andere haben das Problem einfach noch nicht bedacht. Die am RV-NRW beteiligten HRZ werden das Thema bei ihren Gesprächen mit Software-Herstellern und -Vertreibern problematisieren, um zu akzeptablen Lösungen zu kommen. Zu klären ist dabei, inwiefern unter dem Dach des RV-NRW in Zukunft landesweite Lizenzen

von Software verstärkt eine Rolle spielen werden und wie sie möglichst ohne großen Abstimmungs- und Organisationsaufwand zu beschaffen und zu finanzieren sind.

3 Anwendungen

Der RV-NRW ist soweit vorbereitet, dass er nun mit Anwendungen gefüllt werden muss. Im Folgenden werden dazu einige Ideen aufgezeigt. Willkommen sind weitere Hinweise und Vorschläge der potenziellen Nutzer, um Wichtiges mit Vorrang anzugehen. Aus verschiedenen HRZ (z.B. Dortmund, Düsseldorf) liegen bereits Vorschläge vor, konkrete Aufgaben zu übernehmen. Dies soll im 1.Quartal 2001 erweitert und koordiniert werden.

3.1 Hochleistungsrechnen

Im technisch-wissenschaftlichen Bereich gewinnt die Informationsverarbeitung immer mehr an Bedeutung. Neben den klassischen Bereichen Theorie und Experiment entwickelt sich die numerische Simulation immer mehr zum dritten Standbein der Forschung. Es entsteht dadurch ein stark steigender Bedarf an Rechenleistung, dass auf der einen Seite immer mehr Disziplinen die Rechner für ihre Belange einsetzen und dadurch dass auf der anderen Seite die Problemgrößen aufgrund besserer Modelle und genauerer Simulationen stetig ansteigen und mit derart großen Systemen berechnet werden können.

War das Hochleistungsrechnen traditionell eine Domäne der Fluidmechanik (CFD) und Teilchenphysik, werden mittlerweile in den meisten Forschungsdisziplinen leistungsstarke Rechnersysteme zur Lösung aktueller Probleme intensiv genutzt. Beispielhaft seien hier nur Crash-Analysen zur Berechnung der Verformung von großen Strukturen, z.B. Autos oder ganzen Zügen (Gesetzgebung zum Schutz der Verbraucher), die molekulare Modellierung zur Analyse komplexer Molekülstrukturen (Designer Drugs) sowie im Bereich der Klimavorhersage die Lösung des gekoppelten Problems Ozean-Atmosphäre über Jahrzehnte hinweg (Greenhouse Effekt) genannt. Generell bleibt anzumerken, dass erst die Leitungsfähigkeit der verfügbaren Systeme und die Verfügbarkeit geeigneter Simulationsmodelle es erlaubt, praxisrelevante Probleme zu lösen.

Es sind also leistungsfähige Systeme der einzelnen Universitäten in den RV-NRW einzubringen. Dazu gehören vor allem auch Parallelrechner. Die effiziente Nutzung dieser parallelen Systeme bringt allerdings zusätzliche Anforderungen an die verwendete Software bzw. an den Programmierer, der diese Software schreibt. Auch bleibt festzustellen, dass nicht alle Systeme für jedes Problem gleich gut geeignet sind. Der Verbund bietet hier die Möglichkeit der Vielfalt der Systeme und der Vielfalt des Bündels an Know-how aus den verschiedenen HRZ.

Die RWTH Aachen beschafft gegenwärtig einen neuen Hochschulrechner. Das System wird im ersten Halbjahr 2000 installiert und in mehreren Stufen bis Ende 2002 auf eine Rechenleistung von 2,3 Teraflops mit einer Hauptspeicherkapazität von einem Terabyte ausgebaut. Es ist vorgesehen, diesen Parallelrechner auch den anderen am Verbund beteiligten Hochschulen in NRW zur Abdeckung des Bedarfs im Bereich Hochleistungsrechnen zur Verfügung zu stellen.

3.2 Datensicherung

In Hochschulen stehen neben den zentralen Rechnersystemen noch sehr viele dezentral installierte Rechner in den Instituten zur Verfügung. Alle diese Systeme verfügen über eine oder mehrere mit Daten gefüllte Festplatten. Hier stellt sich das Problem der sicheren Langzeitar-

chivierung sowie der Datensicherung zum Schutz vor Hardwareausfällen und versehentlichem Löschen.

An großen Hochschulen kann von einer Anzahl von 5.000 bis 10.000 installierten PCs oder Workstations ausgegangen werden. Nimmt man einen mittleren Fehlerabstand für eine Festplatte von 40 Jahren an, fallen rein rechnerisch pro Jahr etwa 125 bis 250 Plattensysteme aus. Die Gefahr des Datenverlustes ist also gegeben. Dass es sich hier um durchaus eine reale Gefahr handelt, wurde vor kurzem im Rechenzentrum der RWTH Aachen demonstriert, als in einem RAID-System durch einen technischen Fehler Daten mit einem Volumen von 0,5 Terabyte zerstört wurden. Nur einem zuverlässig arbeitenden Backupsystem war es zu verdanken, dass die Benutzerdaten wieder restauriert werden konnten.

Der Einsatz von Bandmedien zur Sicherung erfordert allerdings einen beträchtlichen Aufwand an Personalleistung. Dieser führt dazu, dass diese eigentlich notwendigen Maßnahmen oft nicht oder nur unregelmäßig getroffen werden. Nimmt man zur Berechnung des Aufwandes zur notwendigen Datensicherung eine wöchentliche Datensicherung auf jeder Maschine an, bei der ein Mitarbeiter nur 10 Minuten beschäftigt ist (konservative Schätzung des Aufwandes), würde hochschulweit eine Arbeitsleistung von 833-1.667 Stunden pro Woche anfallen, d.h. 22 bis 44,5 Personen wären vollständig damit beschäftigt, diese Form von Grundversicherung vorzuhalten. Meistens wird diese Aufgabe sogar von wissenschaftlichen Mitarbeiterinnen und Mitarbeitern durchgeführt, so dass hier theoretisch von Personalkosten in Millionenhöhe auszugehen ist.

Als Alternative bietet sich die Sicherung bzw. die Archivierung an einer zentralen Stelle, d.h. durch das jeweilige HRZ an. Einige Zentren bieten den entsprechenden Dienst auch ihren Instituten an. Allgemein bleibt aber festzustellen, dass es sich hierbei um eine kosten- und personalintensive Dienstleistung handelt und dass hierfür notwendige Know-how keineswegs trivial ist.

Hier könnte die verteilte Datenhaltung über den RV-NRW mit einer verstärkten Kompetenzbildung an einigen Stellen für den effizienteren Einsatz von Personal sorgen. Weiterhin ist durch eine gemeinsame verteilte Nutzung der vorhandenen Backup- und Archivierungssysteme an verschiedenen Orten eine Erhöhung der Sicherheit bei Ausfall eines ganzen Zentrums möglich. Dies ist insbesondere für Archive essentiell. Die Sekundärsicherungen könnten also in eine andere Universität verlagert werden, um auch Katastrophen vorzubeugen.

Der Verbund könnte darüber hinaus einen landesweit verteilten Archivdienst zur Verfügung stellen, der es z.B. kooperierenden Arbeitsgruppen ermöglicht, auf ein gemeinsames extrem großes Datendepot zuzugreifen, um so ihre Arbeit effizienter gestalten zu können.

Da moderne Backup- und Archivierungssysteme inzwischen eine Verschlüsselung der Daten anbieten, ist ein Missbrauch auch bei Lagerung in anderen Universitäten auszuschließen.

3.3 MILESS

MILESS ist an der Universität Essen eine digitale Bibliothek für multimediale Dokumente, vorrangig für solche, die an der Universität selbst erzeugt worden sind. Sie dient vor allem der Unterstützung der Lehre, indem sie den Lehrenden die Möglichkeit gibt, den Studierenden zum Abruf über das lokale Netz der Hochschule oder über das Internet Foliensätze, Skripten, Videos, Animationen, Tests, Programme oder Übungsaufgaben als Begleitmaterial für Lehrveranstaltungen bereitzustellen.

MILESS ist aber auch technischer Kern eines im Entstehen begriffenen digitalen Universitätsverlages. Im ersten Schritt werden die in elektronischer Form abgegebenen Dissertationen

(bisher etwa 40) eingestellt. Derzeit wird ein Projekt in Angriff genommen, das vorsieht, das Videoarchiv des Medienzentrums zu digitalisieren und in MILESS einzubeziehen. Schon jetzt findet man mit einer einzigen thematischen Suchabfrage alle einschlägigen Materialien in MILESS selbst und im OPAC der Bibliothek. Das Videoarchiv wird darin einbezogen sein. Auf das nachgewiesene Material kann direkt zugegriffen werden, sofern es digital vorliegt und sofern der Nachfrager die erforderlichen Zugriffsrechte hat.

Grundsätzlich ist vorgesehen, dass die Autoren über ein Web-Interface ihr Material selbst einstellen und aktualisieren. Auf Wunsch wird Hilfe angeboten. Die Bibliothek stellt sicher, dass die Metadaten korrekt eingetragen werden.

Die Architektur von MILESS sieht eine zentrale Metadatenverwaltung und eine beliebig verteilte Objektdatenverwaltung vor, so dass auch Dokumente eingebunden werden können, die auf Fachbereichsservern liegen, trotzdem aber strenge referentielle Integrität gesichert ist. Objektdaten und Metadaten können nicht unabhängig voneinander verändert werden. Leere oder veraltete Verweise sind in MILESS nicht möglich.

MILESS ist seit 1997 in gemeinsamer Arbeit von Rechenzentrum und Bibliothek auf der Basis der IBM Digital Library (jetzt Content Manager) entwickelt worden. Der Beitrag des Rechenzentrums besteht in Programmierung, DV- und Netztechnik, derjenige der Bibliothek in ihrem Know-how in Recherche, Erschließung, Katalogisierung und Kategorisierung von Dokumenten. MILESS hält sich an internationale Standards (Dublin Core, Open Archive Initiative) und kann beliebige Fachklassifikationssysteme (wie z. B. PACS) einbeziehen. Die Anforderungen der Bibliothek haben zu einem sehr komplexen Datenmodell geführt, das aber die sämtlichen bibliothekarischen Anforderungen an fachgerechte Dokumentation abbildet. Das System ist seit 1998 im produktiven Einsatz und enthält viele hundert Dokumente aus allen Fachbereichen.

MILESS ist Bestandteil der CampusSource-Initiative NRW und steht anderen Einrichtungen zur kostenfreien Nutzung unter den Lizenzbedingungen der General Public License (GPL) zur Verfügung. Anderen Hochschulen, die eine solche Digitale Bibliothek aufbauen wollen, wird im Rahmen des RV-NRW die Möglichkeit geboten, die vorhandene und funktionierende Hard- und Software-Infrastruktur der Universität Essen zu nutzen, um zu erproben, ob und in welcher Weise das System ihren Anforderungen genügt. Sie können sich dabei auf die Anwendung konzentrieren, ohne sich um den Aufbau der notwendigen Hardwareinfrastruktur und die Installation und Pflege der Softwareumgebung kümmern zu müssen. Vorerst ist dieses Angebot auf die Erprobungs- und Einführungsphase begrenzt. Ob ein dauerhafter kooperativer Betrieb möglich und sinnvoll ist, könnte in einer frühen Phase des RV-NRW festgestellt werden. Gegebenenfalls wären dann weitere Vereinbarungen zu treffen.

3.4 Smart-Karten

Smart-Karten stehen an vielen Hochschulstandorten des Landes vor der Einführung. Überall fallen nahezu gleiche Aufgaben an. Karten müssen personalisiert und zertifiziert werden. Verloren gegangene Karten müssen unverzüglich, vielleicht sogar Tag und Nacht gesperrt werden können (Revoke). Schon hierbei lässt sich vieles gemeinsam machen. In Ausnahmefällen, etwa bei defekten Geräten oder zur besseren Auslastung einzelner Geräte für Personalisierung und Zertifizierung, kann kooperiert werden.

Smart-Karten sollen den Zugang zu Computern sicherer machen. Dies ist für die verschiedenen Betriebssysteme zu erproben, einzuführen, zu dokumentieren: Das sind in jeder Hochschule wiederkehrende Aufgaben. Wenn der Zugang zum Computer sicherer geworden ist,

können zahlreiche Anwendungen direkt vom Benutzer angestoßen werden, die bisher nur durch persönliches Erscheinen und mit Brief und Siegel möglich waren: Etwa Anmeldungen zu Prüfungen, Ausdruck von Bescheinigungen und Zeugnissen, Immatrikulation und Rückmeldungen. Hierzu sind Web-Seiten zu erstellen und Anwendungsprogramme zu modifizieren. Wiederum handelt es sich um Aufgaben, die in jedem Hochschulstandort anfallen. Der Zugang zu Räumen kann über die Smart-Karte geregelt werden, Zahlfunktionen werden automatisiert. Für alle diese Gemeinsamkeiten bietet der RV-NRW eine Software-, Daten- und Kommunikationsplattform.

3.5 Public-Domain-Software

Public-Domain-Software spielt in vielen Bereichen der wissenschaftlichen Anwendungen eine große Rolle, da die entsprechenden Produkte häufig im Hochschulbereich für den Hochschulbereich entwickelt werden.

Die Software ist allerdings nicht so kostenlos wie es auf den ersten Blick erscheint. Man zahlt zwar kein Geld für das Produkt an sich, muss jedoch für die Installation einen erheblichen Personalaufwand einkalkulieren, der sich einerseits aus der großen Vielfalt der vorhandenen Lösungen ergibt und andererseits durch erhöhten Aufwand bei der Übersetzung für verschiedene Architekturen bedingt ist. Diese Aufgaben fallen in allen Hochschulen immer in der gleichen Weise an.

Somit sind die Konzepte des RV-NRW besonders geeignet, in diesem Bereich der Anwendungen durch Arbeitsteilung zu Einsparungen und höherer Qualität der Dienste zu kommen. Dazu bieten sich zwei Vorgehensweisen an, die aber durchaus auch als Mischform realisiert werden können.

1. Aufbau von produktbezogenen Kompetenzzentren

Hochschulen mit entsprechender Erfahrung stellen einzelne Produkte im Rechnerverbund für alle Plattformen zur Verfügung. Durch die größere Spezialisierung kann die Software besser gepflegt und insbesondere immer auf dem aktuellen Stand gehalten werden. Dies ist der Ansatz, der in der Regel auch bei kommerzieller Software verfolgt wird.

2. Aufbau von architekturbezogenen Kompetenzzentren

Die Übersetzung von Public-Domain-Software verlangt häufig detailliertes Wissen über die Compiler bestimmter Architekturen und ihrer Varianten. Deshalb kann es sehr nützlich sein, dieses Wissen zur Bereitstellung einer ganzen Produktpalette unter einer einzelnen Architektur, die in einer Hochschule intensiv vertreten ist, zu nutzen.

Die vorgeschlagene Struktur zur Softwareverteilung lässt die Realisierung beider Konzepte zu.

3.6 Kompetenzzentrum für Statistik-Software

Die Modellierung Auswertung von Daten ist eine der wichtigsten Techniken des wissenschaftlichen Rechnens. Dabei spielt insbesondere Statistiksoftware eine wichtige Rolle. Auch hier gibt es eine große Vielfalt an Produkten, die von einer einzelnen Institution kaum allein bereitgestellt werden kann, zumal es sich hierbei in der Regel um kommerzielle Produkte handelt, die viel Geld kosten.

Neben der Bereitstellung von Softwareprodukten für alle Arbeitsplatzrechner innerhalb und außerhalb des Rechnerverbunds kann es auch erforderlich sein, dezidierte Server zur Verfü-

gung zu stellen, auf denen die Software landesweit eingesetzt werden kann. Gründe hierfür wären ein erhöhter Bedarf an Ressourcen, der durch eine Workstation nicht abgedeckt wird, oder lizenzrechtliche Gründe, die eine allgemeine Verbreitung nicht kostengünstig zulassen. Zentrale Anwendungsserver lassen sich wegen der Infrastruktur des RV-NRW leicht realisieren.

Der Einsatz von Statistiksoftware setzt viel Erfahrung im Bereich der Angewandten Statistik und detaillierte Kenntnisse bei der Methodenauswahl voraus. Deshalb ist es im Rahmen des Rechnerverbunds sehr wichtig auch einen Beratungs- und Betreuungsservice als Kooperation der beteiligten Zentren aufzubauen.

Die hier vorgestellten Überlegungen lassen sich natürlich auf andere Anwendungsbereiche, wie z.B. numerisches Rechnen, symbolisches Rechnen, im Wesentlichen übertragen.

3.7 Visualisierung

Die Visualisierung von Daten spielt zur Erfassung von komplexer werdenden Zusammenhängen im technisch-wissenschaftlichen Bereich eine immer größere Rolle. Die Zahl der hierfür verfügbaren grafischen Auswertetools nimmt deshalb stetig zu. Vor diesem Hintergrund ist es für den Anwender sinnvoll, auf ein möglichst großes Spektrum an Software zugreifen zu können, um zumindest die Eignung für sein Problem prüfen zu können. Ein einzelnes HRZ ist hier bei der Vorhaltung der Software schnell überfordert und kann aufgrund der Personallage kaum eine adäquate Unterstützung leisten. In einem heterogenen Verbund besteht Zugriff auf ein breiteres Software-Spektrum. Durch verstärkte Kompetenzbündelung wird auch in diesem Feld die Wahrscheinlichkeit der adäquaten Unterstützung bzw. die Existenz eines etablierten Benutzerforums größer.

Einige Hochschulen entwickeln zur Zeit neue Techniken im Bereich der Visualisierung (Virtuelle Realität). Die für diese Technologie notwendige Investitionsmittel sinken durch neue Hardwarelösungen zur Zeit drastisch. Der RV-NRW bietet hier eine Basis, die es ermöglicht, eine größere Gruppe an den Ergebnissen partizipieren zu lassen und über gemeinsame Projekte Fortschritte im eigenen Forschungsbereich zu erlangen.

3.8 Beratung, VoIP, UMS, Call Center und Trouble Ticket

Anbieter von Dienstleistungen, dazu gehören vor allem auch die HRZ, benötigen leistungsfähige Kommunikationsinstrumente. Sie müssen auf allen möglichen Wegen erreichbar sein: Über Telefon im Festnetz oder Handy in allen Varianten, über E-Mail, Telefax, Anrufbeantworter, Videokonferenz usw.. Diese Kommunikationswege werden derzeit im UMS (unified messaging system) zusammengeführt. Zusätzlich wird noch die analoge Übertragung in die digitale überführt und alles im VoIP (Voice over internet protocol) auch über die Datennetze geleitet. Mit VoIP und UMS lassen sich ortsunabhängig Call Center¹¹ einrichten, über die Dienste auch hochschulübergreifend angeboten werden können. Entsprechenden Bedarf gibt es vielfach. Fragen der Benutzer zu einzelnen Betriebssystemen oder zur Anwendungssoftware stellen sich z.B. in Bochum, Bonn, Düsseldorf, Duisburg, Siegen und Hagen, aber auch in allen anderen Universitäten, gleichermaßen. Dem Benutzer ist es in der Regel unwichtig, wer ihm seine Fragen beantwortet. Die Kontakt-Stellen müssen unter einer einheitlichen Kennung erreichbar sein. Die den Dienst erbringenden Personen müssen sich leicht in diesen Ser-

¹¹ Ein schrecklicher Begriff, der besser durch *Kontakt-Stelle*, *Bereitschaftsdienst* oder *Anlaufstelle* zu ersetzen wäre.

vice einklinken können. Das ist mit modernen Kommunikationsinstrumenten flexibel gestaltbar, auch über Hochschulgrenzen hinweg.

Kontakt-Stellen der einzelnen Hochschulstandorte werden sich also gut ergänzen: An einer Stelle nicht sogleich lösbare Fragen können mit Experten an anderen Orten diskutiert werden. Die Bereitschaftszeiten können, wenn viele sich die Aufgaben teilen, ausgedehnt werden. Mit einem landeseinheitlichen Trouble Ticket-System lassen sich gemeldete Fehler registrieren und verfolgen. Der Fragende kann sich selbst vom Bearbeitungszustand ein Bild machen. Nichts muss verloren gehen zwischen den an den Kontakt-Stellen Beteiligten. Mit herkömmlichen Telefonanlagen stößt man dagegen schnell an Grenzen. Derartige Dienste können im RV-NRW schnell eingeführt und erprobt werden. Die Kosten für die notwendige Software lässt sich zwischen den HRZ aufteilen, Hardware wird in der Regel vorhanden sein.

Analoge Kontakt-Stellen oder Bereitschaftsdienste sind übrigens in vielen Bereichen einer Universität notwendig. Erfolgreiche Praxiserprobungen zwischen den HRZ lassen sich danach bei erfolgreichem Verlauf im Rahmen des RV-NRW auf Verwaltungen, Bibliotheken oder Kliniken übertragen. Selbst in Fachbereichen wird man wichtige Anwendungsmöglichkeiten finden.

3.9 Aus- und Weiterbildung

Viele HRZ leisten einen erheblichen Beitrag zur allgemeinen IT-Ausbildung der Studierenden und Bediensteten ihrer Universitäten. Sie unternehmen auch große Anstrengungen zur Weiterbildung ihrer eigenen Mitarbeiterinnen und Mitarbeiter. Die HRZ werden dazu zukünftig die Instrumente der multimedialen Aus- und Weiterbildung verstärkt nutzen können, wenn sich die Mitarbeiterinnen und Mitarbeiter im RV-NRW besser kennengelernt haben. Denn eine vertrauensvolle Zusammenarbeit ist unverzichtbare Voraussetzung bei der gemeinsamen Erstellung von Multimedia-Lehreinheiten. Darüber hinaus ist der RV-NRW besonders geeignet, um kommerzielle Computer Based Trainings-Kurse (CBT) gemeinsam über Universitätsgrenzen hinweg anzubieten.

3.10 Zukünftige Möglichkeiten

In kurzer Zeit werden sich die Möglichkeiten des RV-NRW bei den Mitgliedern der beteiligten Universitäten herumgesprochen haben. Fachbereiche, Verwaltungen oder Bibliotheken werden ihre Kooperation unter Nutzung des RV-NRW verbessern können und auf die Möglichkeiten dieser Infrastruktur mit neuen Ideen zurückgreifen.

4 Kooperation und Konkurrenz

Derzeit sind bereits viele HRZ dem RV-NRW beigetreten: Aachen, Bochum, Dortmund, Düsseldorf, Duisburg, Essen, Münster und Siegen; Bonn und Hagen haben diesen Beitritt für 2001 in Aussicht genommen.

Allein die Anzahl der Beteiligten sagt aber noch nichts über die Aufgabenverteilung und den Nutzen der Einzelnen aus. Der RV-NRW stellt eine Gratwanderung zwischen Kooperation und Konkurrenz der Universitäten dar, die beide vom Gesetzgeber gewollt sind und sich im Laufe der Zeit auch an anderen Beispielen zeigen werden. Die Kooperation im Rahmen des RV-NRW dient dazu, das Leistungsniveau aller Partner durch Nutzung von Synergien zu erhöhen, um im Wettbewerb mit den Hochschulen anderer Bundesländer und anderer Länder in Europa und der Welt die Nase vorn zu haben oder vorn dabei zu sein. Der faire Wettbewerb wird dann auf höherem Niveau auch zwischen den Hochschulen in NRW stattfinden.

Um diesen Prozess der Kooperation in Gang zu setzen, soll zunächst relativ großzügig mit dem Einsatz von Personal, Hardware und Software der einzelnen HRZ umgegangen werden; in absehbarer Zeit muss das Geben und Nehmen der einzelnen Hochschulen im RV-NRW jedoch ausgeglichen sein. Wer nach einem Jahr immer noch nichts gibt, wird von den anderen HRZ nichts mehr bekommen können. Dies muss den Leitungen der partizipierenden Hochschulen deutlich gemacht werden. Wer dauerhaft nur wenig geben kann, wird dauerhaft nur wenig nehmen können. Schon jetzt ist abzusehen, dass einige HRZ aufgrund ihrer unzureichenden Ausstattung nur am Rande mitwirken können. Dabei werden im Rahmen der Kooperation des RV-NRW die unterschiedlichen Schwerpunkte in den einzelnen HRZ, aber auch in den Hochschulen insgesamt, eine deutlich breitere Servicepalette für die Benutzer in NRW ergeben. Gemeinsam werden die Hochschulen z.B. über ein sehr hohes Rechenpotential (allerdings noch nicht das höchstmögliche) verfügen, welches die einzelnen Hochschulen niemals erreichen könnten. Für viele andere Ressourcen und Dienstleistungen gilt dies gleichermaßen.

5 Anhang

5.1 Leistungsfähige Rechensysteme in Deutschland

In der folgenden Tabelle werden als Auszug aus der TOP500 Liste (<http://www.top500.org>) die 55 leistungsfähigsten in Deutschland installierten IT-Systeme aufgeführt (Stand: November 2000). Diejenigen Systeme aus dem akademischen Bereich sind fett dargestellt worden.

Rang (D) (intern.)		Hersteller	Leistung (Gflops)	Installation
1	7	Hitachi	1035	Leibniz Rechenzentrum
2	23	Cray	671	Deutscher Wetterdienst
3	40	Cray	447	Forschungszentrum Jülich
4	50	Cray	355	Max-Planck-Gesellschaft MPI/IPP
5	51	Cray	341	Universität Stuttgart
6	74	Cray	253	Konrad Zuse Zentrum
7	75	<i>IBM</i>	250	<i>Bayer AG</i>
8	83	Cray	234	Forschungszentrum Jülich
9	91	<i>IBM</i>	212	<i>Bayer AG</i>
10	102	<i>IBM</i>	171	<i>Deutsche Telekom AG</i>
11	105	IBM	171	PIK
12	108	<i>IBM</i>	167	<i>Metro MGI Informatik</i>
13	110	IBM	164	GWDG
14	121	<i>IBM</i>	151	<i>BASF</i>
15	126	Eigenbau	143	TU Chemnitz
16	127	<i>IBM</i>	139	<i>Bayer AG</i>
17	130	<i>IBM</i>	138	<i>DeTeCSM</i>
18	131	<i>IBM</i>	138	<i>DeTeCSM</i>
19	149	IBM	132	Universität / FZ Karlsruhe
20	167	NEC	121	Universität Stuttgart
21	171	<i>Sun</i>	118	<i>Bank</i>
22	185	Hitachi	115	Universität Stuttgart & DLR
23	192	<i>IBM</i>	109	<i>Thyssen Krupp Informationssysteme</i>

24	193	IBM	108	BASF
25	196	Fujitsu	106	Leibniz Rechenzentrum
26	197	Sun	105	Mannesmann Mobilfunk
27	200	IBM	102	Deutsche Telekom AG
28	207	IBM	100	Daimler Chrysler
29	210	Compaq	99	Universität Wuppertal
30	215	Eigenbau	96	Universität Tübingen
31	216	IBM	95	Thyssen Krupp Informationssysteme
32	238	IBM	87	BMWAG
33	286	Sun	80	Deutsche Telekom AG
34	317	IBM	79	Thyssen Krupp Informationssysteme
35	318	IBM	79	Deutsche Bank
36	320	IBM	77	COMDIRECT Bank
37	322	IBM	77	RWE
38	324	NEC	77	Universität Stuttgart
39	327	IBM	76	Deutsche Telekom AG
40	334	Sun	75	T-Online
41	339	Sun	74	Debis Systemhaus
42	344	IBM	73	Thyssen Krupp Informationssysteme
43	346	IBM	71	Deutsche Bank
44	352	IBM	71	Europäisches Patentamt
45	360	IBM	69	DeTeCSM
46	364	IBM	68	BASF
47	365	IBM	68	Dresdner Bank
48	400	HP	63	Volkswagen AG
49	409	IBM	62	Deutsche Telekom AG
50	438	IBM	60	Merck
51	441	NEC	60	DLR
52	460	Sun	59	Deutsche Telekom AG
53	476	Cray	58	Alfred Wegener Institut
54	493	NEC	56	Volkswagen AG
55	495	IBM	55	BASF

5.2 Erläuterungen zu DCE/DFS

DCE-RPCs

DCE-RPCs (RPC=remote procedure call) bilden das Fundament der DCE-Zelle. Sie benutzen TCP/IP als Kommunikationsprotokoll und sind authentifiziert, d. h. die jeweiligen Kommunikationspartner können sich gegenseitig ihrer Identität sicher sein. Bei Bedarf kann der RPC-Datenverkehr DES-verschlüsselt¹² werden.

Cell Directory Service

Die Zelle ist die administrative Einheit des DCE. Rechner gehören genau einer (oder keiner) Zelle an. Rechner realisieren DCE-Dienste und nehmen sie in Anspruch. Sowohl der Rechner

¹² DES = Data Encryption Standard

als auch der Dienst sind authentifiziert. Der Cell Directory Server (CDS) ist der Verzeichnisdienst der Zelle. Rechner und Dienst-Anbieter tragen sich im CDS ein. Ein DCE-Client fragt den CDS-Server, welcher Rechner einen bestimmten Dienst zur Verfügung stellt. Eine lokale Parametrisierung der DCE-Clients ist nicht notwendig. CDS-Server sind aus Gründen der Ausfallsicherheit replizierbar. Kommunikation mit anderen Zellen ist möglich.

Security Service

Die Security Server authentifizieren mit Kerberos- Technik Rechner, Server und Benutzer. Sie skalieren mühelos mit wachsender Benutzer-Zahl und sind replizierbar. Passwörter werden nicht im Klartext über das Netz transportiert, Passwort-Änderungen sind praktisch sofort wirksam.

Distributed Time Service

Eine Kerberos-basierende Authentifizierung setzt synchrone Rechneruhren voraus, die mit Distributed Time Services (DTS) realisiert werden. DTS-Server und -Clients bieten ausgefeilte Techniken der Zeitsynchronisation, auch über LAN-Grenzen hinweg.

Distributed File System

DFS ist die wichtigste DCE-Anwendung. DFS-Server exportieren »filesets«. Filesets werden in den Dateibaum montiert und sind für den Benutzer transparent. Filesets können repliziert werden, d. h. von einem Fileset gibt es eine schreibbare Version und ggfs. viele nur lesbare Versionen. Filesets können im laufenden Betrieb vergrößert und von einem Server zum anderen verschoben werden. Eine verteilte Datenbank, die FLDB (=Fileset Location Data Base), hält nach, welcher Server welches Fileset exportiert. Aus Nutzersicht stellt sich ein DFS-Dateisystem wie ein „normales“ Unix-Dateisystem dar. Der Dateizugriff kann zusätzlich mit ACLs über das unter Unix Übliche hinaus gesteuert werden. Damit wird der Dateizugriff von fremden Zellen aus möglich. Die DFS-Clients benutzen einen Cache zur Effizienzsteigerung. Auch DFS-Clients müssen nicht lokal parametrisiert werden, da sie über den CDS die FLDB-Server finden und über die FLDB schließlich den DFS-Server.

5.3 Softwareinstallation

Im Rahmen des verteilten Dateisystems sollen zunächst einige grundsätzliche Überlegungen zu Anwendungen und Architekturen angestellt werden. Danach soll das Konzept zur Installation von Anwendungssoftware vorgestellt werden.

Anwendungen und Architekturen

Zunächst muss deutlich zwischen einer Anwendung und ihrer speziellen Implementierung unterschieden werden. Jedes Anwendungssoftwareprodukt benötigt bestimmte Voraussetzungen, bevor es eingesetzt werden kann. Dabei handelt es sich um eine Kombination von Software und Hardware, die vorhanden sein muss, um die Lauffähigkeit des Systems zu garantieren. Diese Voraussetzungen sollen als Architektur bezeichnet werden.

Klassifikation von Anwendungen

Bei der Beschäftigung mit einer Anwendung ist die Architektur für den Anwender zunächst zweitrangig, da er an der Lösung von Problemen interessiert ist und dies in erster Linie durch die Software erfolgt, während die Architektur die Grundlage der Realisierung bildet.

Für den Verwalter von Anwendungssoftware sind erst einmal organisatorische Kriterien vorrangig, mit denen sich die Systeme architekturunabhängig strukturieren lassen. Wichtige Begriffe sind dabei

- Hersteller
 - Produkt
 - Version

Diese Auflistung ist hierarchisch zu sehen. Ein Softwarehersteller bietet oft mehrere Produkte an, die sich gemeinsam verwalten lassen. Diese Einteilung ist aber nicht notwendig vorgegeben. Häufig stellt ein Anbieter auch nur ein Produkt her, so dass eine weitere Unterteilung an dieser Stelle nicht sinnvoll ist. Andererseits können Produkte des gleichen Herstellers so unterschiedlich sein, dass sie besser getrennt verwaltet werden. Im Mittelpunkt der Betrachtung steht natürlich das einzelne Produkt. Darunter können sich noch verschiedene Versionen (Releases) befinden. Aus Kompatibilitätsgründen ist es oft erforderlich, noch eine ältere Version für eine Übergangszeit bereit zu halten.

Klassifikation von Architekturen

Der Begriff Architektur wird in der Regel mit Hardware-Architektur gleichgesetzt. Aus der Sicht der Anwendungssoftware ist er aber weiter zu fassen. Auf jeden Fall gehört das Betriebssystem mit dazu, evtl. auch eine vorhandene grafische Benutzeroberfläche. Die Architektur lässt sich also folgendermaßen gliedern:

- Softwarearchitektur
- Hardwarearchitektur

Die Hardwarearchitektur wird im wesentlichen durch die Maschine bestimmt, auf der eine Anwendung laufen soll. Entscheidend ist dabei die Prozessorlinie. Unterschiede bei Implementierungen, Herstellern und Maschinenanbietern sind für die Anwendungssoftware in der Regel nicht relevant. Ausnahmen wären Produkte, die z.B. nur auf einer 64-Bit Variante laufen. Eine Sonderrolle bezüglich Prozessorarchitektur und Adressbreite spielt lauffzeitintensive Software, bei der sinnvollerweise die bestmögliche Laufzeitoptimierung zum Tragen kommt.

Die Softwarearchitektur ist im Wesentlichen durch das Betriebssystem bestimmt. Allerdings können hier zusätzliche Komponenten wie eine grafische Oberfläche eine bedeutende Rolle spielen. Auch unterschiedliche Betriebssystemversionen sowie unterschiedliche Wartungsstände können die Einsetzbarkeit von Anwendungssoftware beeinflussen.

Die hier vorgestellten Strukturen decken natürlich nicht alle Spezialfälle ab, aber sie bieten zumindest ein ungefähres Raster, in das sich die meisten Anwendungen einordnen lassen.

Beziehung zwischen Anwendungen und Architekturen

Zwischen Anwendung und Architektur ist ebenfalls eine Hierarchie definiert. Die Anwendung ist dabei höher anzusiedeln, da sie einerseits architekturunabhängige Teile (z.B. Dokumentationen oder Beispiele) enthalten kann und in der Regel auch auf mehreren Architekturen verfügbar ist. Diese hierarchische Anordnung unterstützt die Sicht des Anwenders aber auch die des Softwareverwalters, für den eine produktbezogene Sicht in der Regel günstiger ist.

Das Konzept zur zentralen Softwareverteilung

Das Konzept zur zentralen Softwareverteilung beruht auf drei Prinzipien:

- Im verteilten Dateisystem wird logisch genau ein zentrales Verzeichnis erstellt, das allen Rechnern im Netz zur Verfügung gestellt wird. In dieses Verzeichnis werden dann alle Anwendungssoftwareprodukte installiert.
- Der weitere Aufbau des zentralen Verzeichnisses berücksichtigt die oben vorgestellte Hierarchie zwischen Anwendungen und Architekturen.
- Der Zugriff auf die Software erfolgt von zentraler Stelle.

Das Basisverzeichnis

Die Wahl des Namens für das Basisverzeichnis ist sorgfältig vorzunehmen, damit er möglichst lange unverändert bleiben kann. Änderungen an dieser Stelle haben nämlich einen großen Einfluss auf das gesamte Konzept. Der Name muss natürlich im Rechnerverbund eindeutig sein, damit eine zentrale Verwaltung möglich wird; er sollte keine Hinweise auf den Typ des vor Ort verwendeten Dateisystems (wie NFS, DFS oder lokal) oder andere technische Einzelheiten enthalten. Es muss außerdem sichergestellt sein, dass der Name auf den beteiligten Rechnern nicht schon vergeben ist und in Zukunft für die Softwareverteilung reserviert bleibt.

Die logische Wurzel für die Softwareinstallation im RV-NRW lautet: `/nrwfs/SW`.

Diese Wahl erfüllt die oben aufgestellten Kriterien. Jede konkrete Realisierung lässt sich mit Hilfe einfacher Maßnahmen (Setzen symbolischer Links) an diese Namenswahl anpassen.

Verzeichnisse für Produkte

Der interne Aufbau des Verzeichnisses `/nrwfs/SW` ist ebenfalls landesweit einheitlich. Dieses Verzeichnis muss in sich geschlossen bleiben, d.h. es darf keine Verweise auf Verzeichnisse außerhalb dieses Baums geben. Symbolische Links sind also immer relativ zur Wurzel anzugeben.

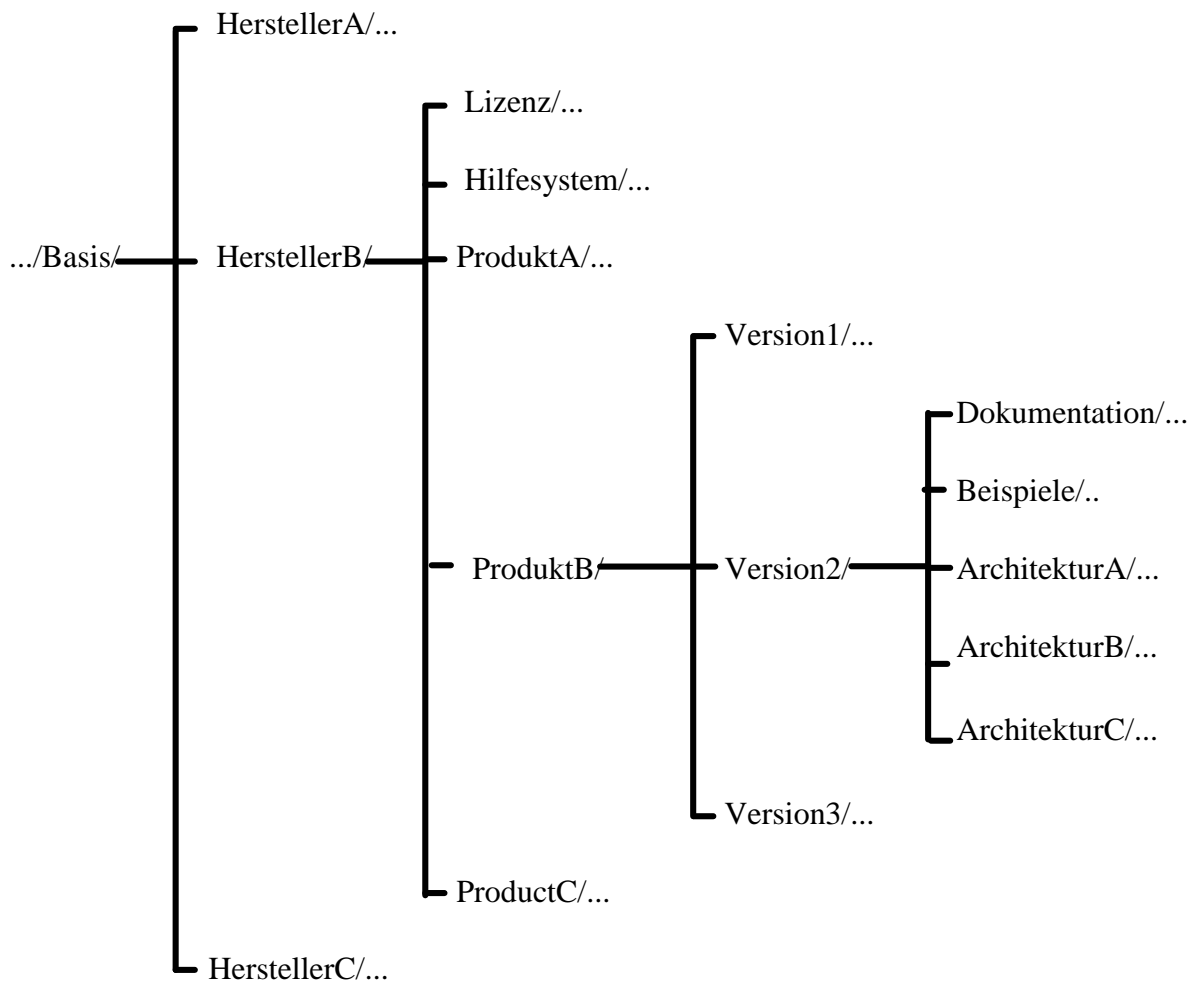
Unterhalb des Basisverzeichnisses liegen thematische Unterverzeichnisse wie „math“, „fem“, „cfd“ etc. Diese Gruppierung strukturiert die installierte Software. Die Zahl der Verzeichniseinträge wird dadurch nicht zu groß, und interessierte Benutzer können sehen, welche thematisch verwandte Software noch installiert ist. Bei Produkten, die mehreren Gebieten zuzuordnen sind, werden in entsprechenden Unterverzeichnissen symbolische Links angelegt. In diesen thematischen Verzeichnissen liegen die Unterverzeichnisse für jedes Produkt. Dabei kann der interne Aufbau mehr oder weniger hierarchisch sein. Häufig ist er vom Hersteller vorgegeben und dann sollte er in der Regel auch so übernommen werden.

Muss die Struktur selbst generiert werden, wird folgende Hierarchie festgelegt:

- Hersteller
 - Produkt
 - Version
 - Architektur

Ein Hersteller kann mehrere Produkte anbieten, die aber vielleicht auf gemeinsame Grundlagen wie Lizenzmanagement oder Dokumentationssystem zurückgreifen. Das Verzeichnis eines Produkts muss unbedingt eine Versionsverwaltung berücksichtigen. Aus Kompatibilitätsgründen müssen nämlich ältere Varianten häufig eine gewisse Zeit zusätzlich bereithalten werden. Dort muss eine weitere Aufteilung in architekturunabhängige Teile (z.B. Hilfe-Dateien) und architekturabhängige Teile (ausführbare Dateien, Bibliotheken, etc.) vorgenommen werden.

Beispiel:



Dieses Konzept ist stets pragmatisch zu handhaben. Wenn ein Hersteller beispielsweise verschiedene Produkte unterstützt, aber jede Architektur einzeln organisiert, dann kann es sinnvoller sein, direkt nach dem Namen des Herstellers den Namen der Architektur folgen zu lassen.

An dieser Stelle stellt sich auch die Frage nach der Benennung der Architekturverzeichnisse. Auch hier machen die Hersteller der Software oft Vorgaben, die leider sehr uneinheitlich sind. Daher führen wir hier gewisse Namenskonventionen ein. Die oben schon angesprochene Aufteilung in Hardware und Betriebssystem wird dabei durch ein Namenspaar in der Form

- softarch-hardarch

abgegeben (eine weitere Hierarchie ist hier nicht mehr angemessen). Beispiele wären:

- solaris2-sparc
- solaris2-ix86
- linux-alpha
- linux-ix86

Die genaue Differenzierung ergibt sich aus den Anforderungen der jeweiligen Software.

Zugriff auf Anwendungssoftware

Das bisher entwickelte Konzept kommt der Sichtweise des Softwareverwalters entgegen. Es erlaubt eine gut organisierte Verwaltung der Software im zentralen Dateisystem. Es sollen sich insbesondere nur wenige Dateien außerhalb des Installationsverzeichnisses eines Produkts befinden.

Die dort beschriebenen Vorteile dürfen jedoch nicht zu Lasten des Anwenders gehen. Der Zugriff auf eine Anwendungssoftware sollte möglichst komfortabel sein. Er erfolgt üblicherweise mit Hilfe eines Skripts. Würde dieses Skript im Installationsverzeichnis des Produkts stehen, dann müsste der Anwender entweder sehr lange Pfadnamen eingeben oder aber seinen Suchpfad bei jedem hinzukommenden Produkt ändern. Hinzu kommt, dass bei dieser Vorgehensweise dem Anwender die interne Struktur des Softwareverzeichnisses bekannt sein müsste. Dies muss aber verborgen bleiben, da es durchaus Änderungen unterworfen sein kann.

Die Notwendigkeit, einzelne Software-Pakete durch spezifische Initialisierungen zugänglich zu machen, muss vom jeweiligen Paket-Verantwortlichen erörtert werden. Abhängig von den Anforderungen der Software ist eine gezielte Initialisierung ggf. vorteilhaft gegenüber einer Überfrachtung der Umgebung, zumal bei der durch die Zusammenarbeit zu erwartenden Mehrung an installierter Software Kollisionen und Nebenwirkungen schwerlich zu überschauen sein werden.

Die Skripts zum Aufruf oder zur Initialisierung sind wiederum an zentraler Stelle abzulegen. Diese Stelle muss für alle Anwendungsprodukte (ggf. wiederum gruppiert nach Themenbereichen - s.o.) gemeinsam sein. Diese sollen architekturunabhängige Skripts, die auf jeder Plattform lauffähig sind, oder symbolische Links auf solche enthalten. Beispiele sind allgemeine Standard-Unix-Skripts oder insbesondere Skripts, die feststellen können, auf welcher Architektur sie aufgerufen wurden und dann selbständig die für die Architektur benötigten Programme korrekt aufrufen. Solche Skripts werden in einem Verzeichnis **bin** des globalen Softwareverzeichnisses abgelegt. Neben diesem Verzeichnis wird es weitere architekturunabhängige Verzeichnisse geben, insbesondere **man**, in dem Hilfeinformationen abgelegt werden, wenn die Dokumentation keine architekturspezifischen Besonderheiten erläutert. Hier sollen beispielsweise auch allgemeine Dokumentationen in einem eigenen Unterverzeichnis abgelegt werden.

Die Auswahl der jeweils "richtigen" Architektur durch das jeweilige Skript hat den großen Vorteil gegenüber einem verzeichnisbasierten Vorgehen, da hier auch wieder der der jeweiligen Software angemessene Differenzierungsgrad realisiert werden kann.

Das Basisverzeichnis für die Softwareverteilung enthält also die Unterverzeichnisse:

- Produkt- und architekturübergreifende Verzeichnisse
- Produktbezogene Verzeichnisse

Da das globale Basisverzeichnis im verteilten Dateisystem auf allen Clients verfügbar ist, stehen neben allen Produkten auch alle architekturunabhängigen bzw. abhängigen Verzeichnisse und die in ihnen enthaltenen Skripts, Binaries und Hilfe-Dateien zur Verfügung.

Der Anwender muss nun den Zugriff auf diese Verzeichnisse durch bestimmte Umgebungsvariablen, insbesondere \$PATH ermöglichen. Diese Setzungen sind aber nur einmal vorzunehmen, da auch alle hinzukommenden (!) Anwendungen automatisch zur Verfügung stehen. Solche Setzungen können beispielsweise im Profil des Benutzers eingetragen werden. Dabei ist es nicht schwierig sogar zu berücksichtigen, dass ein Benutzer häufiger mit Rechnern unterschiedlicher Architektur arbeitet. Dies kann im Profil abgefragt werden und entsprechende Setzungen können vorgenommen werden. Alle architekturabhängigen Verzeichnisse müssen dem Anwender natürlich in der richtigen Form automatisch zur Verfügung gestellt werden. Dies kann beispielsweise durch einen (zentralen!) Profilmechanismus realisiert werden.

Damit der Benutzer in seiner Pfad-Variablen nur einen (architekturabhängigen) Weg angeben muss und nicht für jedes Produkt einen eigenen, müssen natürlich die Skripts aller Produkte in einem Verzeichnis stehen. Dieses Verzeichnis kann dann sehr schnell unübersichtlich werden. Diesen Nachteil kann man umgehen, indem für jedes Produkt ein eigenes Unterverzeichnis mit dem Namen **bin.product** angelegt wird, wo alle Skripts für ein Produkt liegen. Eine Stufe höher werden dann die gleichen Namen als symbolische Links in das Produktverzeichnis zur Verfügung gestellt. Damit befinden sich alle Dateien für den Benutzer an einer Stelle, während der Verwalter wieder über eine produktbezogene Struktur verfügt. Dieser Mechanismus erlaubt es Benutzern auch, nur einzelne Produkte von zentraler Stelle abzurufen, beispielsweise, wenn es eine lokale Installation gibt, die einige der zentralen Produkte ebenfalls vorhält.

Aufgaben für die Softwareverwalter

Im Rahmen der oben bereits beschriebenen Aufgaben bestimmt der Softwareverwalter (Paketverantwortliche) insbesondere den internen Aufbau des entsprechenden Teils des Softwarebaums unter Berücksichtigung der festgelegten Kriterien. Er legt auch die notwendige Differenzierung der Architektur (Betriebssystem-Version, Prozessorlinie, Adressbreite, patch level etc.) fest und installiert die zugehörigen Skripten im Verzeichnis **/nrwfs/SW/bin**.

Schnittstelle zur Systemverwaltung

Die wichtigste Schnittstelle zur Systemverwaltung ist sicherlich das oben beschriebene zentrale Dateisystem. Es wird unter Beibehaltung der Namenskonventionen jeweils lokal realisiert.

Damit es effizient genutzt werden kann, sind dort natürlich für alle Softwareverwalter die entsprechenden Zugriffsrechte zu setzen. So sollte der Verwalter eines Produkts (und sein Vertreter) alle Rechte in seinem Pfad besitzen. Zusätzlich muss es möglich sein, in zentralen Verzeichnissen (z.B. bin) Dateien anzulegen und zu modifizieren. Hier sollte auch über eine geeignete Gruppenstruktur für die Softwareverwaltung nachgedacht werden.

Das zentrale Dateisystem muss aber auch von allen unterstützten Architekturen zugreifbar sein, da manche Installationen auf der für sie richtigen Architektur laufen müssen. Für Tests ist dies sogar unumgänglich. Im RV-NRW muss also für jede Architektur ein Referenzsystem für Installationen und Tests für alle Softwareverwalter zur Verfügung stehen. Dies betrifft insbesondere Rechner mit neuen Versionen von Betriebssystemen.

Mit der Systemverwaltung müssen auch weitere Schnittstellen abgesprochen werden. Zu nennen wäre insbesondere der schon erwähnte Profilmechanismus, der die Umgebung des Benutzers so einstellt, dass z.B. durch das Abfragen von bestimmten Umgebungsvariablen die für den Einsatz bestimmter Softwareprodukte benötigten Einstellungen vorgenommen werden können.