

Kompetenzverbund und Datenmanagement im Rechnerverbund NRW

Christian Bischof, Rheinisch-Westfälische Technische Hochschule Aachen
Wilhelm Held, Westfälische Wilhelms-Universität Münster
Bruno Lix, Universität Essen

09. Juli 2001

1 Hintergrund

Eine funktionierende IT-Infrastruktur ist für das wissensbasierte, kooperativ agierende Unternehmen Universität von herausragender Bedeutung. Immer komplexere Systemlösungen sind erforderlich, um IT-Dienste zu erbringen, die den qualitativ und quantitativ ständig steigenden Ansprüchen der Benutzer genügen.

Die Universitätsrechenzentren sind auch im Wettbewerb mit externen Dienstleistern gefordert, ein immer umfangreicheres Spektrum von IT-Services auf hohem Niveau mit hoher Verfügbarkeit zu planen, bedarfsgerecht auszubauen und zu betreiben. Konkret verlangen die Benutzer vor allem zuverlässige Mail-, Web-, Rechen-, und Datendienste sowie Beratung bei der Konfiguration und dem Einsatz dieser Dienste. Darüber hinaus führt der ansteigende Einsatz von Softwarewerkzeugen aller Art zu einer stetig wachsenden Nachfrage an Anwenderberatung, Schulung und methodischer Unterstützung. Dieser Nachfrage nach immer mehr, immer zuverlässigeren und immer anspruchsvolleren Diensten stehen keinesfalls steigende, sondern bestenfalls konstante Personal- und Sachmittelressourcen gegenüber. Weil diese Entwicklung schon seit Jahren anhält, sind an den Hochschulen die Reserven, die durch interne Umorganisation und Effizienzsteigerung nutzbar gemacht werden können, in der Regel erschöpft. Eine weiterhin innovative und noch wirksamere Nutzung der vorhandenen Ressourcen erfordert organisierte hochschulübergreifende Zusammenarbeit, z. B. in Kompetenzverbänden.

Deshalb wurde im Rechnerverbund NRW (RV-NRW) ein einfach nutzbarer landesweiter Verbund von IT-Ressourcen eingerichtet, der derzeit konkurrenzlos ist. Die Benutzerverwaltung ist landeseinheitlich, die Daten sind automatisch dort, wo sie benötigt werden und die Ressourcen können landesweit in Anspruch genommen werden. Auf dieser Struktur aufbauend ist es möglich, Aufgaben, die in einzelnen Hochschulrechenzentren gleichermaßen anstehen, arbeitsteilig gemeinsam zu lösen, um so die dringende Weiterentwicklung der IT-Infrastrukturen und IT-Anwendungen voran zu bringen und die Verfügbarkeit von Diensten sicherzustellen. Im Gegensatz zum 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 programmatischen Innovationsanspruch der Hochschulen.

2 Die Bedeutung von Datenmanagement

Mit der Existenz des RV-NRW ist die Voraussetzung geschaffen, IT-Dienste von vornherein in diesen Rahmen zu integrieren. So ist mittlerweile die für die theoretische Chemie wichtige Beilstein-Datenbank für alle RV-NRW Nutzer verfügbar und durch eine Zwei-Server Installation in Aachen und Dortmund auch gegenüber Ausfällen relativ unempfindlich. Des weiteren wird

der in Aachen in der Beschaffung befindliche Hochleistungsrechner von Anfang an in den RV-NRW mit eingebunden.

Darüber hinaus hat der Arbeitskreis der Rechenzentrumsleiter in NRW (ARNW) eine große Zahl von Kooperationsfeldern identifiziert, auf denen jeweils mehrere Rechenzentren ihre Kompetenz im Rahmen des RV-NRW sinnvoll bündeln könnten. Insbesondere haben die Rechenzentren der Universitäten Aachen, Essen und Münster vereinbart, gemeinsam konkrete Schritte zur Realisierung eines verteilten Datenmanagements zu unternehmen. Die Organisation von Daten und der Zugriff auf diese Daten ist eine für numerisch und nicht-numerisch orientierte Benutzer gleichermaßen elementare Forderung. Mit der wachsenden Menge der Daten wird eine gemeinsame und auch langfristig sichere Datenhaltung immer mehr zur Voraussetzung für Kooperationsprojekte. Großvolumige, sicher betriebene Daten - Repositories, auf die alle Berechtigten von allen Orten gleichermaßen zugreifen können - ohne aufwändige, umständliche und fehleranfällige Datentransporte sind die geeignete und angemessene Form der Datenhaltung für solche wichtigen und erwünschten Kooperationen.

Um diese Erwartungen insgesamt erfüllen zu können, erfordert dies auf technischer Seite die Bereitstellung großer Datenmengen mittels eines skalierbaren Datenhaltungskonzeptes, wie z.B. eines storage area networks (SAN), die Archivierung großer Datenmengen und die Datensicherung (Backup). Der Betrieb eines solchen Systems erfordert leistungsfähige Helpdesk-Systeme und bei verteilten kooperierenden Standorten auch verlässliche und flexible Kollaborationsmechanismen.

Jeder dieser angesprochenen Themenkreise ist für sich allein komplex und aufwändig. Andererseits sind hochverfügbare und sichere Datenhaltung eine der elementarsten Anforderungen eines Benutzers an den IT-Dienstleister. Die Rechenzentren in Aachen, Essen und Münster beabsichtigen deshalb ein verteiltes Datenmanagement in einer eng koordinierten Art und Weise zu implementieren und so die Kernzelle für einen im RV-NRW ausbaufähigen Kompetenzverbund zu bilden. Durch eine Abstimmung von lokalen Bedürfnissen mit denen des RV-NRW in der Konzeptionsphase und kontinuierlich darüber hinaus kommt lokal zu leistende Arbeit auf natürliche Art und Weise dem RV-NRW zugute und es lassen sich die erheblichen Synergieeffekte aufzeigen und nutzen, die beim verteilten Datenmanagement entstehen.

3 Die Kernzelle Aachen/Essen/Münster

Die Rechenzentren in Aachen, Essen und Münster beabsichtigen, ein verteiltes Datenrepository mit verteilter Archivierung auf Basis von ADSM/TSM (von IBM/Tivoli) aufzubauen. Dabei nutzen sie die spezifischen Kompetenzen, die sie im Bereich Datenmanagement bereits erworben haben (siehe Abbildung 1).

In Essen und Münster liegen derartige Erfahrungen im Aufbau und Betrieb eines zentralen Backup- und Archivierungsdienstes auf der Basis von ADSM/TSM seit 1994 vor. In Aachen gibt es seit langem und insbesondere im Zusammenhang mit dem Betrieb der Landeshöchstleistungsrechner Kompetenzen und Erfahrungen mit der Speicherung und Verwaltung großer Datenmengen aus Projekten des wissenschaftlichen Höchstleistungsrechnens. Münster hat als erste Hochschule im Land die verteilten Arbeitsplatzrechner und Server in eine hochschulweite DCE - Infrastruktur mit verteiltem Filesystem (DFS) einbezogen und damit Datenhaltung und Datenverarbeitung getrennt. Essen befasst sich seit einiger Zeit intensiv mit

der Umstrukturierung des hochschulweiten Datenmanagements durch eine getrennte Verwaltung von Anwendungen und Daten nach den Prinzipien von Storage Area Networks (SAN) und Network Attached Storages (NAS).

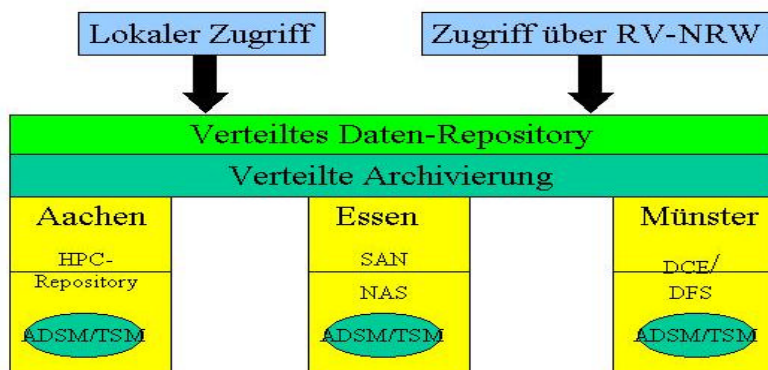


Abbildung 1

Ein Beispiel für Synergieeffekte, die durch gemeinsame Planung und abgestimmten Betrieb möglich werden, ist ein verteilter Archivierungsdienst, der bereits in die Tat umgesetzt ist. Dabei wird den Bedürfnissen des Datenschutzes dadurch Rechnung getragen, dass die eigentlichen Archivdaten an einem Standort, die zur interpretationsfähigen Wiederherstellung unerlässliche Datenbank mit den beschreibenden Metadaten an einem anderen Standort dupliziert werden. Nur mit der Information aus Essen *und* Münster können z.B. die Aachener Daten rekonstruiert werden (siehe Abbildung 2).

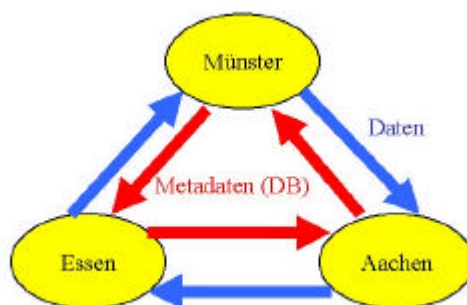


Abbildung 2

Diesem Schritt sollen weitere folgen. Konkret ist angedacht, dass Aachen sich dem Bereich Backup und Archivierung federführend zuwendet, Essen die Implementierung von NAS und SAN vorantreibt, und Münster neben dem weiteren Ausbau der DCE/DFS Infrastruktur auch Helpdesk- und Kollaborationssysteme evaluiert. Um ihre Beiträge zum Kompetenzverbund in der vorgesehenen Weise leisten zu können, müssen alle drei Hochschulen ihre Infrastrukturen zum Datenmanagement ausbauen. Entsprechende HBFG-Anträge sind in Aachen und Essen in Vorbereitung, Münster wird im nächsten Jahr folgen.

Durch die ständige abgestimmte Arbeit an der gemeinsamen Aufgabe und den damit einhergehenden Erfahrungsaustausch werden Kompetenzen gebündelt und (z. B. auch durch Schwerpunktbildung) gestärkt. Die Vereinbarungen zwischen Aachen, Essen und Münster sehen

auch gegenseitige personelle Hilfeleistungen (z.B. bei Urlaub oder Krankheit des lokalen Administrators) vor. Damit wird (zumindest teilweise) das Problem gelöst, dass schwierige Störfälle in komplexen Diensten nicht angemessen behoben werden können, wenn der (oftmals) einzige Mitarbeiter, der das erforderliche vertiefte Wissen hat, nicht zur Verfügung steht.

4 Landesinteresse

Die Förderung dieses Kompetenzverbundes liegt im landesweiten Interesse: Die spezifischen Kompetenzen der drei Hochschulen betreffen drei der wichtigsten aktuellen Entwicklungen im Bereich Datenmanagement. Alle Hochschulen haben in unterschiedlicher Weise und auf unterschiedlichen Entwicklungsstufen Schritte zur Kooperation unternommen. Sie werden aus der Bündelung der vorhandenen Kompetenzen in dem entstehenden Kompetenzverbund in der Form Nutzen ziehen und ein landesweites Referenzzentrum bilden.

Von übergeordnetem Interesse ist ebenfalls, dass hier zum ersten Mal hochschulübergreifende Zusammenarbeit in einer wichtigen Produktionsumgebung (nicht nur in einem experimentellen Einzelprojekt) stattfindet. Von nicht zu unterschätzender Bedeutung ist, dass diese Kooperation und ihre geplante Weiterentwicklung Gebrauch von der Leistungsfähigkeit des Wissenschaftsnetzes macht.

5 Ausblick

Im Rahmen eines kooperativen und abgestimmten Verbundes können die überall gleichermaßen anfallenden Aufgaben auf einen größeren Personenkreis verteilt werden und so eine höhere Verfügbarkeit von Diensten erreicht werden. Dies ist notwendig vor dem Hintergrund der Bedeutung der von den Rechenzentren geleisteten Dienste, der i.a. dünnen Personaldecke und der durch die nach wie vor große Nachfrage nach IT-Spezialisten verstärkte Personalfluktuations. Der Mehraufwand, der sich hochschulübergreifend durch die Betreuung einer größeren Anzahl von Nutzern ergibt, ist verglichen mit der grundsätzlichen Beschäftigung mit einem Thema an sich eher kleiner. So ist durch die Vermeidung der Replikation von Arbeiten in einem kooperierenden Verbund eine verstärkte Kompetenzbündelung bei den Mitarbeitern möglich. Auch ist es i.a. beträchtlich preisgünstiger, vorhandene Systeme zu vergrößern als sie komplett neu aufzubauen.

Für die Datendienste ist neben Maßnahmen zum ununterbrochenen Betrieb der Server und Dienste eine örtlich verteilte Datenhaltung fester Bestandteil der Ausfallsicherung. Die jetzt zur Verfügung stehenden modernen und leistungsfähigen Weitverkehrsnetze, hier insbesondere das G-WiN ermöglichen effiziente und modernere Verfahren zum verteilten Datenmanagement, die im Rahmen des Kompetenzverbundes erprobt werden. Damit wird auch die Möglichkeit eröffnet, dass entweder Hochschulen an diesen Datendiensten teilnehmen, ohne die dafür erforderliche technische Infrastruktur und personelle Kompetenz selbst aufbauen zu müssen, oder dass sie durch Einbringen von eigenen Kapazitäten die Leistungsfähigkeit des Datenverbundes weiter stärken.