

**50 JAHRE RECHENZENTRUM /  
COMPUTER- UND MEDIENSERVICE**

DER HUMBOLDT-UNIVERSITÄT ZU BERLIN



# ÜBERBLICK ZUM FILESERVICE AB 1990

Frank Sittel

54



Backup-Server und -Library (1999-2010)

## Einleitung

Die ersten Jahre des ORZ (Organisations- und Rechenzentrum) wurden durch Einzel- oder Groß-Rechentechnik dominiert, die zentral bedient und gewartet wurde. Für die nur sehr vereinzelt existierende Büro-Computer-Technik wurde von Seiten des ORZ Hilfe bei der Inbetriebnahme und technischen Wartung gegeben, aber dezentrale Dienste, wie wir sie heute verstehen (wie Vernetzung, Mail, Backup), gab es überhaupt nicht. Erst das Ende der DDR, der Neustart ins RZ und die damit verbundenen Erstbeschaffungsmittel ermöglichten den Aufbau einer vernetzten, dezentralen Rechnerumgebung. Und die Betonung liegt auf „Netz“ – ohne den Aufbau eines universitätsweiten LAN und später SAN ist die Explosion von Diensten, wie wir sie heute

um uns haben, schlicht unmöglich. Mit der Vernetzung und dem Einzug der Arbeitsplatzrechner ergab sich der Bedarf nach Unterstützung dieser Systeme durch das RZ. Stück für Stück wurden universitätsweite Dienste etabliert. Der Fileservice verstand sich im Rahmen dieser Entwicklung als ein Dienst, der den Angehörigen der Universität einen sicheren Ort anbietet, an dem sie ihre Dateien ablegen, wiederfinden und diese im Disaster-Fall auch restaurieren können.

## Banyan VINES ab 1990

Im Herbst des Jahres 1990 wurden einige PCs beschafft, eine Handvoll BNC-Kabel durch die Räume des RZ gezogen und in einer Kammer neben dem Rechnerraum zwei Tower-Rechner aufgestellt, auf denen ein Stapel Disketten mit der Aufschrift „Banyan/VINES“ eingespielt wurde und schon war die Humboldt-Universität im Besitz des ersten Netzwerk-Dienstes, der darauf ausgerichtet war, eine gemeinsame Nutzung von Platten, Druckern, Gateways und Servern zu ermöglichen. Das Hervorstechende an VINES (VIRtual NETwork System) war, dass es sich dabei um ein service- und nicht serverorientiertes System handelte, d. h. ein Nutzer am PC musste überhaupt nicht wissen, wo sein Platten lagen, wo der Drucker angeschlossen war und wie er an seine Mail kam – er rief einfach „Her mit der Temp-Disk der Abteilung 4“ und schon kam sie durch das Netzwerk angefliegen! Ermöglicht wurde dies durch den ersten Verzeichnis-Dienst der HU in einer Zeit, in der noch niemand ahnte, was Verzeichnis-Dienste überhaupt sind –

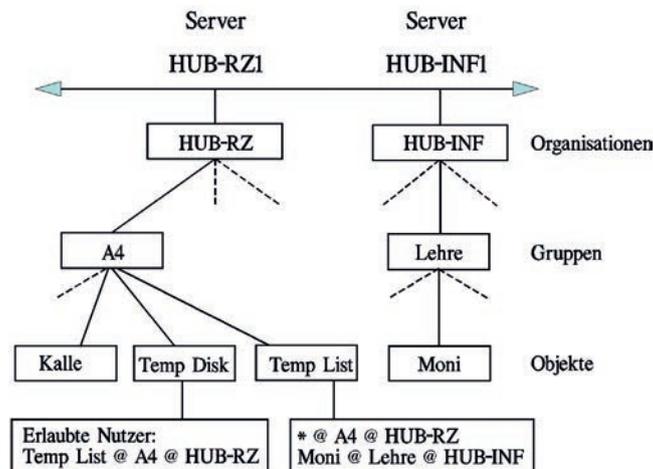
„Street Talk“ von VINES. Das dreistufige Namenskonzept war genial einfach und intuitiv: objekt@gruppe@organisation, wobei Objekte sowohl Nutzer als auch Dienste und Listen sein durften. An den Diensten wurden die Berechtigungen verwaltet. Der Ruf nach der Disk sah dann so aus:

```
C:\> setdrive t:
„Temp Disk@A4@HUB-RZ“
```

Und dieser Zugriff funktionierte von jedem PC, auf dem ein VINES-Client installiert war – egal wo er stand und wie er hieß – unterstützt wurden DOS, Windows, OS/2 und Apple Macintosh. Die Street-Talk-Struktur hinter diesem Zugriff sieht in etwa so aus, wie in der folgenden Abbildung dargestellt.

Auf diese Art wurden die Nutzer daran gewöhnt, ihre wichtigen Daten auf VINES-Laufwerke abzulegen. Dort wurden sie nämlich gesichert – naja, in der Anfangszeit wurden sie einmal pro Woche auf ein Tape geschrieben. Nichtsdestotrotz war dies der Anfang eines wirklichen Fileservice. Die Nutzer vertrauten dem VINES ihre Daten an und hatten meistens nicht die geringste

Ahnung, wo sie tatsächlich landen würden und trotzdem die Hoffnung, sie zurückzubekommen. In den folgenden 10 Jahren war VINES so erfolgreich, dass es zu seinen besten Zeiten auf 102 Servern knapp 50 Organisationen mit 8000 Nutzern beherbergte.



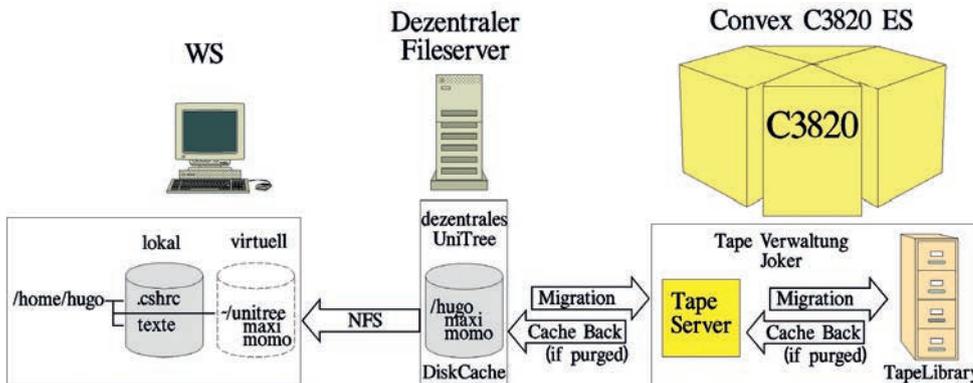
Street-Talk-Struktur

## Unix-Fileservice 1993 – 1999

Anfang der 90er Jahre waren Unix-Workstation auf RISC-Basis sehr populär, da sie in Prozessor- und Grafikleistung den PCs noch deutlich überlegen waren. Die Anzahl wuchs vor allem in den naturwissen-

1 TB passte – zu dieser Zeit war Terabyte noch ein Synonym für unendlich. Das Bestechende an diesem Ungetüm (Convex + Library) war die Software, die das alles zusammenhielt: Unitree. Es war unsere erste Begegnung mit einem automatischen

zierten Wert, dann werden alte und große Dateien gelöscht, bleiben aber in der File-System-Ansicht erhalten. Greift man auf gelöschte Dateien zu, dann werden sie vom Band zurück auf den Cache kopiert und man kann sie stark verzögert benutzen. Der besondere Witz unserer HSM-Lösung bestand aber darin, dass sich der Tape-Server nicht nur mit einem, sondern mit beliebig vielen Disk-Caches verbinden konnte. Das führte 1993 zur Beschaffung von acht Sun-Servern mit Disk-Arrays, die wir in die Institute verteilten (Wirtschaftswissenschaften, Physik, Chemie, Mathematik ...). Jeder von diesen Servern hatte neben einem „normalen“ Disk-Bereich von 5–10 GB, der per NFS verteilt wurde, einen zusätzlichen Disk-Cache für Unitree und dieser konnte so lange beschrieben werden, wie es die Tape-Library hergab – es ging also „unendlich“ viel hinein. Zudem hatten wir eine Backup-Lösung der Firma „Software Mogul“ mit dem Namen SM-arch im Einsatz, die nicht mit der Tape-Library kommunizierte, sondern mit dem HSM von Unitree. Damit hatten wir für den Unix-Bereich erstmals eine umfassende File-



Datenfluß des UniTree (dezentral – zentral)

schaftlichen Instituten und Banyan VINES hatte für Unix als Klient nichts übrig. So wurde Ende 1992 schon in Hinblick auf diese zukünftigen Klienten ein zentraler Compute- und Fileserver Convex C3820 beschafft, ein Vektorrechner mit zwei riesigen Vektor-CPU's und Anschluss an eine Tape-Library. Diese Library hatte zwei Laufwerke und 48 VHS-Bänder, auf die etwas mehr als

HSM-System (Hierarchical Storage Management) und dieses System hatte es in sich. Unitree reserviert einen Disk-Bereich (Cache) für sich, in dem man lokal, per NFS oder FTP Dateien einspielen kann. Diese Dateien werden nach einer vorgegebenen Zeit über einen Tape-Server-Prozess in die Library kopiert. Füllt sich dieser Disk-Bereich bis zu einem vorher spezifi-

service-Lösung mit realem Plattenplatz per NFS für Home-Directorys, ein langsames, aber sehr großes HSM und ein Backup-System. Aus einem Vortrag stammte die Veranschaulichung unseres HSMs. Was uns dieses dezentrale Server-System außerdem nahebrachte, war, dass die Festplatten nicht im selben Gehäuse wie der Server untergebracht werden mussten. Die „SPARC Storage Arrays“ waren Stand-Alone-Geräte, die mit Glasfaser-Leitungen hunderte von Metern entfernt von den eigentlichen Servern stehen konnten, was wir in einem Fall tatsächlich auch so benutzt haben. Außerdem war auf den Leitungen die irrsinnige Transfer-Leistung von 200 MBit/s – das war doppelt so schnell wie FDDI. Die Erfahrung aus dieser Zeit minderte unsere Berührungsangst mit dem SAN im Jahre 2000 deutlich.

### Das Jahr 1999

Das Jahr 1999 war in der Öffentlichkeit vor allem durch Katastrophen-Szenarien geprägt, was wohl zur Jahrtausendwende so alles passieren könnte. Uns ereilten zwei wirkliche Katastrophen. Die Firma Hewlett-Packard, die 1995 Convex übernommen hatte, gab uns am Anfang des Jahres die Garantie, dass sowohl die Convex C3820 als auch die angeschlossene Tape-Library den Jahreswechsel nicht überstehen würden. Das zweite Desaster passierte im Herbst – die Firma Banyan kündigte ihre Produktpalette auf (vor allem VINES) und benannte sich in ePresence um, die dann nur noch Kunden beraten wollte, wie sie aus VINES herauskommen können. Zumindest hat sie kurz vor ihrem Zusammenbruch noch die letzte Version 8.5 und Jahr-2000-Patche herausgereicht, so dass an dieser Front kein sofortiger Handlungsdruck bestand. Der größte Teil der VINES-Server lief noch bis 2003 und dann kam es zu einem sukzessiven Wechsel in Richtung Windows-Fileservice unter dem Active-Directory-Service.

Deutlich schlechter sah es mit unserem Unix-Fileservice aus. Vor allem HSM und der Backup-Dienst drohten, völlig auszufallen. In einer Hauruck-Aktion wurden nicht nur Mittel bewilligt und eine Ausschreibung vorbereitet, sondern wir erhielten zudem

### Backup und HSM mit TSM seit 1999

Eines der Dinge, die wir aus dem Unitree/SM-arch-Zeiten gelernt haben, war, dass die Nutzer HSM zwar ganz nett fanden, aber eigentlich viel dringender eine gute Backup-Lösung suchten. Da SM-arch



Verkabelung an einem Fiber-Channel-Switch

die Chance, den Fileservice neu auszurichten. Zwei Neulinge hielten Ende 1999 Einzug ins RZ: das Backup-System TSM von IBM und das SAN.

auf Unitree nicht wirklich gut skalierte, sahen wir uns nach einer leistungsfähigen Backup-Lösung um, die zudem auch eine HSM-Komponente besaß. Wir fanden sie im Tivoli Storage Manager (TSM) von IBM, der außer Backup- auch HSM- und Archivierungskomponenten besaß. Der eigentliche Reiz von TSM besteht im Backup-Verfahren „incremental forever“, was bedeutet, dass es einen Full-Backup nur ein einziges Mal gibt – beim Start. Danach

werden nur noch die geänderten Daten gesichert. Außerdem gehörte es ab sofort zu unseren Design-Prinzipien, alle wichtigen Daten (und alle Daten sind wichtig) doppelt an zwei Standorten zu halten. So wurden zwei Backup-Server und -Librarys für die Stand-

Library in Adlershof durch eine von Quantum ersetzt wurde, deren Laufwerke nun über das SAN angesprochen wurden. Der neue TSM-Server erreichte auch seine Cache-Arrays nur noch über Fiber-Channel-Verbindungen. Bei der Umrüstung 2009 hatten

## Das SAN ab 2000

Einer der wesentlichen Gründe, ein SAN (Storage Area Network) aufzubauen, ist, die Datensicherheit deutlich zu erhöhen. Durch ein Netzwerk wird es erst möglich, im großen Stil Duplikate der Daten über Gebäudegrenzen hinweg zu erstellen. Doch zuerst mussten die technischen Voraussetzungen geschaffen werden. Angefangen haben wir zu Beginn des Jahres 2000 mit zwei IBM-Speichertürmen, die 60 SCSI-Platten fassten, je 1,5 TB Plattenvolumen lieferten und mit zweimal 1-GBit an die Switche angeschlossen waren, mit zwei 16-Port-1-GBit-FC-Switchen der Firma Brocade und einer Handvoll Servern, die die Sun-Server ersetzten. Da war noch nicht viel mit Redundanz – 2000 haben wir vor allem die Technologie geübt und diverse Tests durchgeführt. Eines wurde bald klar – ein bezahlbarer Ausbau der Speicherkapazitäten im SAN war mit FC- oder SCSI-Platten nicht möglich. So haben wir schon 2000 probiert, ob und wie IDE (oder PATA) Arrays verwendet werden können. Bald war der erste Selbst-Bau-Turm fertig, der 12 IDE-Arrays

über den Umweg von zwei SCSI-FC-Routern mit einer Kapazität von 6,5 TB ans SAN brachte. Einen Produktions-„Test“ bestand dieses System erfolgreich und so wurden in den folgenden drei Jahren weitere Türme des Selbst-Bau-Typs installiert. Schon 2001 wurde klar, dass eine direkte Zuweisung von Speicher an unsere Klienten (die Server der Universität) mit wachsender Speicher- und Klienten-Zahl ein exponentiell wachsendes Management-Problem darstellt. So begannen wir 2002, verschiedene Management-Lösungen zu untersuchen. Pünktlich 2003 – zum Umzug des CMS ins Erwin Schrödinger-Zentrum – konnte die Grundstruktur unseres SANs dank der Erstbeschaffungsmittel für dieses Gebäude realisiert werden, wie sie qualitativ noch heute Bestand hat: zwei unabhängige FC-Netzwerke, Verbindung der Standorte Mitte und Adlershof per WDM, Einsatz von hochverfügbaren Core-Switchen, Virtualisierung des Speichers als Management-Mittel, Gebäude- und Standort-Vernetzung in Ring-Topologien usw. In diesem System ist alles

*Der Computer- und  
Medienservice der Humboldt-Universität atmet den freien und toleranten Geist  
des universitären Lebens, er bildet im Campus ein  
Stück Heimat. Das ist nicht selbstverständlich  
für eine solche Einrichtung.*

*Peter Zelt  
Wirtschaftswissenschaftliche Fakultät*

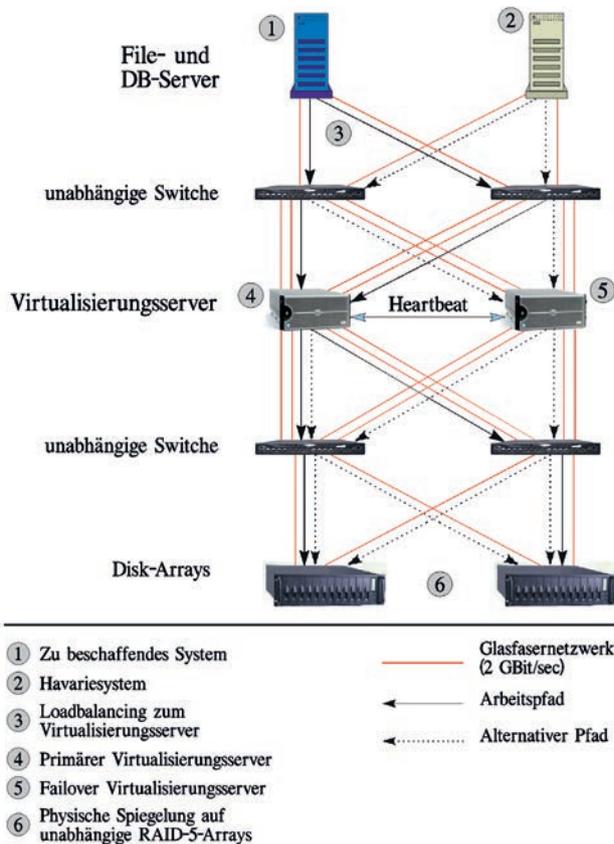
orte Mitte Hauptgebäude und Adlershof Informatik beschafft. In den Anfangsjahren wurden aus Kapazitätsgründen nur die Archiv-Daten auf den jeweilig anderen Standort gedoppelt – heute auch alle Backup-Daten. Die 1999 beschafften Backup-Systeme waren von der Struktur her Produkte des letzten Jahrtausends – monolithische Blöcke: IBM-Server mit per SCSI direkt angeschlossenen Tape-Librarys und lokalen Plattenarrays. Das wurde 2003 aufgebrochen, als die StorageTek-

dann alle Laufwerke an beiden Standorten FC-Anschlüsse und die Server griffen auf im SAN befindliche SAS-Arrays als TSM-Cache zu. 1999 hatte die beiden Librarys eine Kapazität von 40 TB – heute von 4 PB .

doppelt vorhanden, wie die Prinzip-Skizze aus einem älteren Vortrag zeigt.

Jede fehlerhafte Komponente kann in diesem System umgangen werden. Alle redundanten Komponenten sind über insgesamt fünf Standorte ver-

teilt. Die Virtualisierung spiegelt alle Daten und hält die Kopien in verschiedenen Gebäuden. Das Gesamtsystem SAN ist so ausgelegt, dass alle Daten ohne Verlust auch dann zugreifbar bleiben, wenn ein zentraler Standort vollständig ausfällt (z. B. abbrennt).



SAN-Struktur

In den Folgejahren bis 2009 wurden neue Arrays mit SATA-Platten bestückt und die Netzwerk-Geschwindigkeit erhöhte sich auf 4 GBit/s. Mit den Startmitteln beim Einzug ins Grimm-Zentrum im Jahre 2009 konnte die Geschwindigkeit im Netzwerk auf 8 GBit/s erhöht werden, es wurde die Anzahl der Virtualisierungsserver-Cluster auf vier und der verfügbare Speicher auf 1 PB erhöht. Damit waren die Grundlagen gelegt, statt einiger Dutzend nun hunderte von Servern zu unterstützen (derzeit 370).

Es gibt eine ganze Reihe von Möglichkeiten, die aus dem Einsatz unseres SANs erwachsen. So werden weiträumige Clusterungen von Servern ermöglicht. Im Windows-Fileservice, der seit 2003 die VINES-Server ersetzt, sind alle Cluster über die Standorte Adlershof und Mitte gespreizt. Beteiligt sind daran 31 Windows-Server mit 55 TB zugewiesenem Speicher für 31000 Nutzer. Außerdem erlauben die Virtualisierungsserver mit ihren Snapshot-Mechanismen Server- und LAN-free Backup. Alle zentralen Mail-Server der Univer-

sität werden ohne Beanspruchung der Server und des LANs auf separaten Maschinen gesichert. Da im SAN der Zugriff von mehreren Systemen auf eine Speicher-Ressource möglich ist, können damit Cluster-Filesysteme zum Einsatz kommen, wie im Compute-Cluster und in den PC-Saal-Clustern.

Das SAN ist der vorläufige Höhepunkt im Bestreben, Dateien einen sicheren Ort zu bieten und zudem ein äußerst stabiler und wichtiger Dienst des CMS geworden.