Fileservice für die Institute auf Basis eines Storage Area Network

Frank Sittel
sittel@cms.hu-berlin.de

Christoph Weickmann
weickmann@cms.hu-berlin.de

Fileservice, Storage Area Network, SAN, Speichervirtualisierung, Backup

Der folgende Artikel beschäftigt sich mit der zu realisierenden Struktur des Fileservices für die Institute der Humboldt-Universität an den Standorten Adlershof und Mitte.

Begünstigt durch den Umzug des CMS nach Adlershof und die damit verbundenen Erstausstattungsmittel erfährt der weitere Ausbau des universitätsweiten Fileservice in Abstimmung mit der Medienkommission der Humboldt-Universität einen erheblichen quantitativen und qualitativen Anstoß. Wie bereits in den RZ-Mitteilungen [1] beschrieben, wird von uns vor allem die Entwicklung eines Storage Area Networks (SAN) vorangetrieben. Die wesentlichen Gründe, ein solch komplexes Unternehmen in Angriff zu nehmen, sind:

- Zunehmend größere Festplatten ermöglichen es selbst kleinen Arbeitsgruppen, sich Arrays in TeraByte-Größe hinzustellen. Für einen LAN-bezogenen Backup/Restore ist dies ein Problem, vor allem, wenn ein solches Array einmal verloren gehen sollte.
- Die Erwartungen an eine permanente Verfügbarkeit eines Fileservice sind sehr hoch.
- Die Verwaltung, vor allen Dingen die Fernwartung, ist in einem Speicher-Netzwerk sehr viel einfacher zu bewerkstelligen.
- Ein SAN gestattet ein LAN-free Backup. Bei einem LAN-free Backup wird der Weg der Daten vom Weg der Metadaten (dies sind die Verwaltungsdaten, über die TSM weiß, wo sich die Nutzerfiles befinden) getrennt. Die Metadaten wandern nach wie vor über das LAN, während die eigentlichen Daten vom Client via SAN direkt auf die Volumes des Backupservers gespeichert werden. Es wird also der »normale« Netzverkehr durch das Backup nicht mehr behindert.

In Vorbereitung der Installation unseres SAN haben wir bisher Fileserver und Storageserver (IBM RSS 2102 und Infortrend 6300 mit Chaparral-Router VFS226) eingesetzt und die Laufwerke unserer Backup Library auf SAN aufgerüstet. Mit dieser Konfiguration versorgen wir bis heute die Institute für Chemie, Informatik und Mathematik, die Kulturwissenschaften, die Sozialwissenschaften sowie die Universitätsbibliothek mit insgesamt 3,1 TByte Disk aus dem SAN-Bereich.

Im Storage-Bereich wird ein zusätzliches Feature angeboten – die Storage-Virtualisierung. Die Abbildung 1 soll das Prinzip verdeutlichen. Zwei Storage-Virtualisierungs-Server (SVS) importieren alle Festplatten aus den Storage-Zonen, »filtern« deren Inhalt und geben sie in die Client-Zonen an die SAN-Clients wieder heraus. Sie arbeiten in dem Sinne redundant, dass der Ausfall eines Servers nicht zum Plattenverlust des SAN-Clients führt.

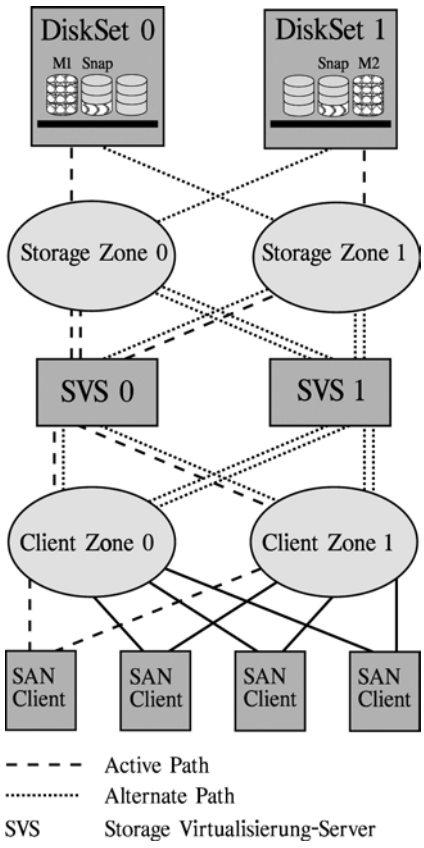

Abb. 1: Prinzip der Storagevirtualisierung

Wir benutzen als Virtualisierungssoftware IPStor der Firma FalconStor.

Für uns werden die SVS folgende sinnvolle Tätigkeit vollbringen:

- Sie können aus den relativ großen RAID-5-LUNs (jede ca. 670 GByte) kleine Scheiben herausschneiden und so die weniger speicherhungrigen Clients befriedigen, die sich mit 200–300 GByte begnügen wollen. Andererseits können sie mehrere LUNs zu einer Riesenplatte zusammenfassen, wenn das gewünscht sein sollte.
- Die SVS beherrschen das LUN-Masking, d. h. sie vergeben die virtuellen Disks genau an die Clients, die sie auch bekommen sollen – jeder Client sieht exakt seine Platte und sonst nichts.
- Wir werden uns nicht auf die Funktionsfähigkeit unserer RAID-5-Systeme verlassen, sondern lassen die SVS je 2 RAID-5-LUNs aufeinander spiegeln. Der SAN-Client erhält von uns immer ein gespiegeltes RAID-5-Ensemble als eine Disk, d. h. er muss sich um die Spiegelung nicht mehr selbst kümmern.
- Die SVS können außerdem jede Nacht Snapshots der virtuellen Disks machen, d. h. sie frieren den Datenbestand zu einem bestimmten Zeitpunkt ein und schreiben die Veränderungen in eine separate Snapshot-Area. Diese solchermaßen eingefrorenen Platten können den Clients als read-only-Disk zur Verfügung gestellt werden, aus denen ein Nutzer irrtümlich gelöschte Daten aus der »Vortags«-Disk selbst restaurieren kann. Ob wir dieses Feature anbieten, wird von der Gesamtlast im System abhängen.

Mit der Installation im Erwin Schrödinger-Zentrum werden 64 RAID-5-Controller mit insgesamt 4864 MByte Cache ca. 40 TByte Disk (netto, d. h. nach Abzug der RAID-5-Parity-Kapazität und der überall vorhandenen Reserve-Disks) kontrollieren und diese dann mit einer maximalen SCSI-Bandbreite von 5,5 GByte/sec ins SAN exportieren. 280 Switch-Ports werden diesen Datenstrom als 20 TByte gespiegeltes Plattenvolumen an bis zu 40 Server weiterleiten. In drei Tape-Robotern werden 19 Laufwerke über einen Tape-Bestand von über 110 TByte (unkomprimiert) herrschen.

Es werden folgende Komponenten eingesetzt:

SAN

- 12 Brocade-Switches (SilkWorm 12000 und 3800, durchgängig 2 Gbit/s) werden in zwei physisch getrennten Netzwerken (Fabrics) so angeordnet, dass alle angeschlossenen Komponenten (Plattensysteme, Tape-Roboter, SVS, Clients) mit je der Hälfte ihrer Fiber Channel Interfaces in je einer Fabric vorhanden sind. Über eine Dual-Pathing Software wird sichergestellt, dass selbst beim Ausfall einer kompletten Fabric die Clients und SVS über die verbliebene Fabric einen Weg zu ihren Storage-Einheiten finden.
- Das SAN der Standorte Adlershof und Mitte wird über vier DWDM-Kanäle (je 2 Gbit/s) miteinander verbunden.

Storage

- Als kleinste Speichereinheiten kommen 60 Infortrend 63x0-Systeme zum Einsatz. Sie besitzen acht Ultra-ATA-Hard-Disk, die im RAID-5-Modus plus Reserve-Disk betrieben werden. Als externes Interface besitzt solch ein System einen Ultra-160-SCSI-Kanal.
- Insgesamt zehn Chaparral-Router VFS226 bringen die Disk-Arrays auf zwanzig 2 Gbit/s-Kanälen zu den SAN-Switches.
- Vier Dell-Server mit je zwölf 2 Gbit/s-Fiber-Channel-Ports arbeiten in den Standorten Mitte und Adlershof als redundante SVS und wirken zusätzlich zu den bereits oben beschriebenen Möglichkeiten als einheitliches Speicher-Management-System.

Fileserver

- Als Fileserver werden vom CMS die Plattformen Sun, IBM, Win2k und Redhat-Linux direkt unterstützt.

Backup-System

- Als Backup-System wird an der HU seit mehreren Jahren der IBM Tivoli Storage Manager (TSM, vgl. auch [2]) eingesetzt. Wir betreiben bis jetzt 2 Server an den Standorten Mitte und Adlershof, von denen die Daten in je einer Library STK 9710 mit 6 Laufwerken DLT 7000 und einer Library IBM 3494 mit 5 Laufwerken Magstar 3590H1A gespeichert werden. Für den Umzug nach Adlershof und die damit erforderliche Einbindung des SAN wurde ein weiterer Backup-Server mit der Enterprise Edition des TSM installiert. Als Library verwenden wir hier eine Scalar 1000 der Firma Advanced Digital Information Corporation (adic) mit 6 LTO2-Laufwerken. Die an das SAN angeschlossenen Laufwerke (LTO) ermöglichen uns den LAN-free Backup.
- Im weiteren Ausbau des Systems planen wir, einen Server-free Backup über die SVS zu realisieren. Bei dieser Art des Backups werden die Metadaten über das »normale« LAN zum Backupserver gesendet, während die Daten direkt von den Platten unter Umgehung des Clients (der Fileserver) auf den Volumes des Backupservers gespeichert werden. Hierbei wird also der Client (Fileserver) durch das Backup überhaupt nicht belastet.

Durch die Verwirklichung des SAN-Konzeptes erwarten wir, dass der Datenhunger unserer Nutzer zukunftsweisend gestillt wird und die Sicherheit der einzelnen Daten optimal gewährleistet ist.

Literatur

[1] Sittel, F.: Storage-Area-Network an der Humbold-Universität. *RZ-Mitteilungen* 22/Nov. 2001, S. 10–14 bzw. http://www.rz.hu-berlin.de/rz/rzmit/rzm22/4.pdf

[2] Weickmann, Ch.: TSM – 1999–2001. *RZ-Mitteilungen* 22/Nov. 2001, S. 15–16 bzw. http://www.rz.hu-berlin.de/rz/rzmit/rzm22/5.pdf