Zentraler File-Service auf der Convex C3820 ES

In den letzten Jahren ist die Datenflut auf den Rechnern der HUB - vom PC über die Workstation bis zu den zentralen Rechnern - enorm gewachsen. Damit einher geht das Problem für den einzelnen Nutzer, wie er seine Daten verwaltet, was in erster Linie bedeutet, wie kann er sie sichern, archivieren und im Zweifelsfalle auch einmal wieder herstellen, wenn etwas zerstört wurde; was aber auch bedeuten kann, wie er den ihm zur Verfügung stehenden Speicherplatz erweitern kann. Für das Rechenzentrum entsteht noch gleichzeitig das Problem, wie die Datenmenge (Files) der zentralen Rechner verwaltet werden bzw. wie der Drang nach Speicherplatz und der nur begrenzt zur Verfügung stehende Plattenspeicher auf den zentralen Rechnern miteinander in Einklang gebracht werden.
Um alle diese Probleme zu lösen wurde vom Rechenzentrum für die UNIX-basierenden Rechner - also Workstations und zentrale Rechner - ein Konzept eines File-Service erarbeitet, das jetzt mit der Einrichtung eines zentralen File-Service in einer ersten Realisierungsphase steht. Die Weiterführung des Konzeptes durch die Einbeziehung dezentraler Fileserver ist im folgenden Artikel beschrieben.

Darstellung des Konzeptes

- Ausgegangen wurde von einer Mehrstufigkeit des Systems, die beim Workstationarbeitsplatz des Wissenschaftlers beginnt, den Server eines Workstation-Clusters berücksichtigt und schließlich im gemeinschaftlich genutzten File-Server im Rechenzentrum mündet.

- Unterschieden wurden auf der Ebene des zentralen File-Service vier verschiedene Arten von Daten, die jeweils andere Formen der Verwaltung verlangen:
  - Aktive Daten sind Daten, auf die in einem gewissen Zeitraum zugegriffen wurde.
    Im Projekt SERVUZ wurden für diesen Zeitraum sechs Monaten angesetzt. Dabei wurde von einem Datenvolumen von 300 GByte in den nächsten fünf Jahren für die HUB ausgegangen, das in erster Linie von den Workstation-Clustern der Fachbereiche kommt.
  - Inaktive Daten sind Daten, auf die seit einem gewissen Zeitraum nicht zugegriffen wurde, die jedoch prinzipiell noch zu den Daten gehören, zu denen zugegriffen wird und die dabei noch verändert werden könnten.
    Als Zeitraum wurden hier sechs Monate angesetzt. Das Datenvolumen wurde für die nächsten fünf Jahre auf 150 GByte für die HUB geschätzt.
  - Sicherungskopien sind Daten die sich aus Backups zentraler Compute-Server und Backups von Nutzerdaten auf Workstations bestehen. Diese Daten fallen täglich an, müssen nur gespeichert werden und werden nur in Havariefällen benötigt.
    Das anfallende Datenvolumen wurde für diese Daten auf 250 GByte geschätzt.
  - Archivdaten sind solche Daten, die nur für Lesevorgänge bestimmt sind und bei denen die Häufigkeit des Zugriffes geringer ist als einmal in sechs Monaten.
    Der Umfang dieser Daten wurde auf 200 GByte geschätzt.

- Um in einem zentralen File-Server diese Daten verwalten zu können, wurde für diesen von einem hierarchischen System ausgegangen, wobei die Hierarchie mindestens aus einem primären Speichermedium, von dem aus sich Files im direkten Zugriff verarbeiten lassen, und aus einem sekundären Speichermedium besteht. Die Nutzung des zentralen File-Servers sollte über NFS (Network File System - das Filesystem des Servers wird direkt in die nutzende Anlage eingebunden) oder über FTP (File Transfer Protocol - die nutzende Anlage kopiert die Files vom oder zum Server) erfolgen, wobei NFS eine erhöhte Netzbelastung nach sich zieht und deswegen nur begrenzt angewendet werden sollte.

Anzubietende Dienste

Entsprechend der oben skizzierten vier Arten von Daten können auch drei grundsätzliche Arten von Diensten unterschieden werden, die das Rechenzentrum mit dem zentralen File-Service anbietet (Voraussetzung dafür ist, daß eine hinreichend schnelle Verbindung - z. B. durch den FDDI-Ring - zum Rechenzentrum besteht):

- Erweiterung des Speicherplatzes für Nutzerfiles auf den lokalen Rechenanlagen des Rechenzentrums - das können Workstations aber auch die ALLIANT FX /2800 sein - und auf den Workstations der Fachbereiche. Dies kann z.B. dadurch erreicht werden, daß der Nutzer Files auf dem zentralen File-Server ablegt (Filespeicherung) und nur zur Abarbeitung auf seine Anlage holt (Filebereitstellung).
- Datensicherung für zentrale und dezentrale Computer. Diese kann nachts erfolgen, um eine Überlastung des Netzes zu vermeiden. Um der Begrenztheit der Speichermedien gerecht zu werden, kann dieser Dienst nicht generell für jede

Workstation und für alle ihre Files erfolgen sondern nur für ausgewählte Anlagen (z.B. für die Nutzerverzeichnisse auf den geplanten dezentralen File-Servern).
- Archivierung von Files; da die Zugriffsgeschwindigkeit für diese Daten keine so große Rolle spielt muß die Archivierung nicht automatisch erfolgen. Wichtig ist nur, daß der Nutzer seine Daten in einer zumutbaren Zeit wieder zur Verfügung gestellt bekommt. Dies könnte auch über manuelle Eingriffe durch die Bediener am zentralen File-Server erfolgen.

Folgende notwendigen Anforderungen an den zentralen File-Server wurden gestellt:
Bereitstellung eines "unbeschränkten" Direktzugriffsspeichers auf der Basis eines hierarchischen, automatisch verwalteten Speichersystems mit primärem und sekundärem Speichermedium sowie Software zum Transport von Files zwischen primärem und sekundärem Medium.

- Direktzugriffsspeicher:
  . ca. 30 GByte (für etwa 8 - 10% der aktiven Daten)
  . ausbaubar auf 50 GByte
  . Sicherungsmöglichkeiten (z.B. in Form von Disk Arrays)
  . Transferrate von mindestens 3 MByte/s

- Sekundärspeicher:
  . Speichermedium muß häufig les- und beschreibbar sein
  . Automatische Fehlererkennung mit Korrekturmöglichkeit
  . ca. 700 GByte Größe; ausbaubar auf das Doppelte
  . Maße und Anforderungen an die Klimaumgebung des Systems sollten möglichst gering sein

- Migrations/Caching-Soft- und Hardware
  . automatische Migration und Caching von Files, einschließlich automatischer Wechsel von Sekundärdatenträgern
  . hohe Zuverlässigkeit der Geräte und Speichermedien
  . automatische Fehlererkennung, Führung von Fehlerstatistiken, Aussonderung fehlerhafter Datenträger
  . automatische Datenkomprimierung im Sekundärspeicher

Stand der Realisierung

Für die Realisierung der Aufgaben des zentralen File-Service wurde eine C3820 ES der Firma Convex beschafft, die auch noch Aufgaben des Compute-Service erfüllen soll. Die Entscheidung zugunsten der C3820 ES fiel, weil sie im Preis-Leistungs-Verhältnis die oben angegebenen Forderungen am besten erfüllte.
Bezüglich File-Service weist die Anlage folgende Parameter auf (die den Compute-Service betreffenden Parameter sind im vorhergehenden Artikel angegeben):

- Primärspeicher: 40 GByte Plattenspeicher, aufgeteilt auf 16 Platten mit je 2,5 GByte und 9,34 MByte/s Transferrate. Für den File-Service sind davon 30 GByte vorgesehen.
  - 5 GByte Nutzerbereich (Nutzerfilesysteme werden durch Quotas reglementiert)
  - 25 GByte sind für den zentralen File-Service im engeren Sinn vorgesehen

- Sekundärspeicher: VHS-Robotersystem RSS-48 mit zwei Laufwerken und 48 VHS-Kassetten; dadurch wird eine Speicherkapazität von 1,04 TByte erhalten.

- Verwaltungssoftware: Als Verwaltungssystem wird das Softwareprodukt UniTree der Firma DISCOS eingesetzt.
  UniTree ist ein hierarchisches File- und Speicherverwaltungssystem, das besonders für heterogene Rechnernetze mit einem zentralen Fileserver geeignet ist. Es ist so konzipiert, daß es bequem handhabbar ist für Systeme, die einen starken Zuwachs an Daten haben.

  UniTree verwaltet ein virtuelles Speichersystems, das bei uns aus dem Primärspeicher - dem Disk-Cache - und dem Sekundärspeicher - den VHS-Kassetten besteht. Das Filesystem, auf dem UniTree arbeitet, unterscheidet sich von einem Standard-Unix-Filesystem, um so den Verwaltungsaufgaben besser gerecht zu werden. UniTree hat folgende Eigenschaften:

  - Keine logische Begrenzung der Filegröße
  - Ausgefeilte Fehlerbehandlung bei I/O-Fehlern.
  - Automatische Migration (Sichern von Files aus dem Primärspeicher in den Sekundärspeicher) und ein automatisches Caching (der umgekehrte Vorgang - d.h. ein File, der sich nicht im Primärspeicher befindet, dort aber angesprochen wird, wird automatisch eingelagert). Damit wird eine Unabhängigkeit vom Speicherort des Files erreicht.
  - Halten von mehreren Versionen eines Files
  - UniTree unterstützt als Verbindung zwischen seinem virtuellen Filesystem und dem Unix-Filesystem NFS und FTP
  - Plattenverwaltung; wenn der Primärspeicher einen bestimmten Füllstand besitzt, werden Files gestrichen (eventuell vorher noch migriert)

- Automatische Operationen, wenn ein Robotersystem verfügbar ist (bei uns RSS-48)
- Automatisches Backup aller Daten, die zur Verwaltung von UniTree gehören.
- Offenes System; UniTree ist nicht abhängig von einzelnen Computerproduzenten, es ist vollständig portabel. Dies ist eine besonders wichtige Eigenschaft, wenn es in heterogenen Umgebungen - wie z.B. die HUB - eingesetzt wird.

Da UniTree seine Daten im virtuellen Speicher verwaltet, muß sich der Nutzer die Files für seine eigene Verarbeitung bereitstellen bzw. nach der Verarbeitung wieder abspeichern. Dies kann dadurch geschehen, daß über NFS der Primärspeicher gemountet wird oder daß der Nutzer sich seine zu benutzende Files mittels FTP aus dem virtuellen Speicher bereitstellt oder wieder in ihm speichert. Durch die Filebereitstellung erhält der Nutzer die Möglichkeit, seinen Speicherplatz, der im Nutzerbereich reglementiert ist, zu erweitern.

Realisierungsstand der Dienste

Für die Realisierung der Dienste werden die 25 GByte Speicherplatz, die für den File-Service im engeren Sinne vorgesehen sind, aufgeteilt:
- 15 GByte Disk-Cache; dies ist der Primärspeicher von UniTree
- 10 GByte Scratch-Bereich

Erweiterung des Speicherplatzes:
Dieser Dienst wird durch die Filespeicherung und -bereitstellung in Verbindung mit Migration und Caching realisiert. Für die Filebereitstellung ist der Scratch-Bereich vorgesehen, in dem die Files mittels FTP durch den Nutzer bereitgestellt werden müssen.

Datensicherung:
Bei der Datensicherung muß zwischen der Sicherung von Nutzerdaten des Compute-Servers (Daten aus den 5 GByte Nutzerbereich der C3820 ES) und der von Daten aus dem dezentralen File-Service unterschieden werden. Für letzteres wurde ein Remote System Backup (RSB) beschafft, das die Zusammenarbeit zwischen Backup und UniTree-Verwaltung realisiert.
Die zur Sicherung vorgesehenen Daten der dezentralen Fileserver können auch nachts, wenn die Belastung im Netz nicht so hoch ist, in das UniTree-Filesystem geschafft und damit auch auf den VHS-Kassetten gespeichert werden. In der ersten Ausbaustufe, wenn die dezentralen Fileserver noch nicht beschafft sind, werden wir die Erfahrungen für diesen Sicherungsmechanismus mit den lokal angeschlossenen Workstations des Rechenzentrums und den Workstations, die in Verbindung mit den SERVUZ-Projekt beschafft wurden und über eine hinreichend schnelle Verbindung (FDDI, Ethernet) zum zentralen File-Server verfügen, sammeln.

Datenarchivierung:
Zur Datenarchivierung sind zum jetzigen Zeitpunkt noch keine endgültigen Aussagen zu treffen, da sie sehr eng mit der Datensicherung zusammenhängt und auf diesem Gebiet erst die entsprechenden Erfahrungen vorliegen müssen. Ein vorläufiges Konzept sieht folgendes vor:
Die Archivierung erfolgt auf VHS Kassetten, jedoch nicht über UniTree. Die Kassetten sind so aus dem Roboter entfernbar und gesondert aufhebbar.
Da für diese Kassetten gilt, daß sie nur noch ganz selten angefaßt werden, muß geklärt werden, wie die Lesbarkeit gesichert wird. Weiter muß noch geklärt werden, ob der Vorgang der Archivierung in den Händen des Rechenzentrums liegt, oder ob geeignete Werkzeuge geschaffen werden können, die das dem Nutzer überlassen.

Christoph Weickmann