Backup – Datensicherheit für alle

Christoph Weickmann
weickmann@cms.hu-berlin.de

Backup, Datensicherung, Restore, LAN Backup, LANfree Backup

Die Notwendigkeit der Datensicherung erkennt der Nutzer meist erst, wenn es zu spät ist. An der HU wird Datensicherung schon seit vielen Jahren betrieben. Das Backupsystem TSM (Tivoli Storage Manager) ermöglicht LAN- und LANfree Backup.

Backup

Stimme des Nutzers: Mich interessiert nicht das Backup – ich muss nur wissen, wo ich meine Daten wiederfinde.

Unter dem Backup von Daten verstehen wir die Sicherungskopie(n) einzelner Dateien oder ganzer Datenträger (Filesysteme, Filespaces, Laufwerke usw.). Das heißt, die Daten des Nutzers existieren nach der Sicherung mindestens zweimal – zum einen beim Nutzer und zum anderen auf externen Datenträgern, die sich möglichst außerhalb des direkten Zugriffs des Nutzers befinden – dies kann z. B. eine Tapelibrary (Roboter) an einer zentralen Stelle sein. Diese Daten bleiben so lange erhalten, wie sie auch beim Nutzer existieren. Werden sie durch ihn gelöscht, werden sie auch auf den externen Datenträgern gelöscht – allerdings zu einem späteren Zeitpunkt (vgl. unten), so dass der Nutzer die Möglichkeit hat, seine Daten aus der Tapelibrary wieder zu restaurieren (restore).

Was kann meinen Daten schon passieren?

Für gewöhnlich interessiert sich der normale Anwender der Rechentechnik nicht dafür, wo sich seine Daten befinden – sie sind »eben da«. Dies wird sich spätestens dann ändern, wenn er zu den Daten zugreifen will und sie sind »eben nicht da«. Ab diesem Zeitpunkt wird er sich dafür interessieren, was mit ihnen alles passieren kann:

- die Platte, auf der sich die Daten befinden, kann »crashen«
- ein Deletekommando an der falschen Stelle
- ein Softwarefehler kann Daten löschen
- Einfluss von Viren
- die Daten wurden geändert und ein älterer Stand wird benötigt

Gleichgültig was es war – der Nutzer hat seine Daten, die oft nicht mehr rechtzeitig reproduzierbar sind (Examensarbeiten, Forschungsergebnisse), verloren.

Wie in [1] beschrieben, wird vom CMS ein großer Aufwand betrieben, um den Nutzer von der Verwaltung seiner Daten zu entlasten und um die Daten sicher gegenüber Plattenfehlern zu machen. Wenn wir damit auch 99,..% Sicherheit erreichen sollten, bleiben doch noch die Probleme des eigenen Löschens usw. übrig bzw. geht nach Murphy die eine Platte kaputt und der Blitz schlägt bei der anderen Platte (Spiegel) ein.

Um diese Gefahren zu vermeiden, existiert das Backup.

Warum Backup zentral vom CMS?

Eigentlich könnte jeder Nutzer seine Daten selbst sichern. Er arbeitet fleißig am PC und sichert zum Feierabend seine Daten auf einer Diskette oder einem anderen Medium.

Wenn dies so funktionieren würde, könnte der CMS sich viel Arbeit sparen. Wo liegt der Irrtum bei dieser Herangehensweise:

- Der Nutzer sichert nicht zum Feierabend (dies ist ein Erfahrungssatz).
- Die Haltbarkeit des Sicherungsmediums ist indiskutabel.
- Im Bedarfsfall: auf welcher Diskette habe ich meine Datei gesichert?
- Welcher Stand meiner Datei liegt wo?

Um diese Probleme zu umgehen, könnte sich jeder Nutzer sein eigenes Backupsystem schaffen – er besorgt sich entsprechende Hard- und Software und sichert fröhlich drauflos. Dies würde bedeuten, dass er sich weniger um seine eigentliche Arbeit kümmert, sondern stattdessen aufwendige Datensicherung betreibt, abgesehen von den Ausgaben für die Hard- und Software.

Um hier eine klare Linie an der Universität zu erhalten, hat der CMS schon vor mehreren Jahren ein zentrales Backup-

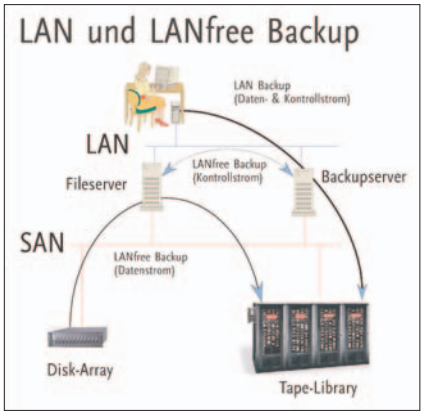

Abb. 1: LAN- und LANfree Backup

system aufgebaut, das aus mehreren Servern und Tapelibraries besteht und von den Nutzern gut angenommen wird [2].

Die Idee dabei war, an den beiden großen Standorten (Mitte und Adlershof) der Universität jeweils einen Backupserver mit Tapelibrary zu installieren. An jedem Standort wurden die dort befindlichen Server gesichert. Gleichzeitig wurden die gesicherten Daten aus der Tapelibrary des einen Standorts zur Tapelibrary des anderen Standort kopiert. Dies ist eine Funktionalität des bei uns eingesetzten Backupsystems TSM (**T**ivoli **S**torage **M**anager von IBM). Die doppelte Sicherung wird von uns gemacht, damit wir unabhängig von Tapefehlern oder sonstigen Katastrophen an einem Standort sind – man erhält die Daten dann eben vom anderen Standort.

Die Tapes werden regelmäßig auf ihre Les- und Schreibbarkeit geprüft, so dass sich dort ankündigende Probleme rechtzeitig erkannt und korrigiert werden können.

TSM ist ein Server-/Clientsystem. Auf den Clients – den Nodes in der TSM-Sprache – wird eine Software installiert (der Softwareclient), die in Zusammenarbeit mit dem Server die Daten zum Server transportiert. Wir unterstützen Clients für Unix, einschließlich Linux, Windows und Mac.

Mit der Installation des Clients ist für den Nutzer bereits die ganze Arbeit getan. Das Backup wird vom Backupserver automatisch angestoßen; wenn ein Fehler auftritt, erhält der Nutzer eine Mail, worauf er in den Logfiles nach Ursachen suchen kann. Bei Problemen kann er sich mit uns in Verbindung setzen, damit wir diese gemeinsam lösen.

Da die Software nach Lizenzen für einzelne Nodes bezahlt wird, bieten wir den Backupservice in erster Linie Arbeitsgruppen an, die ihre Daten schon zentralisiert haben, z. B. auf File-, Web- oder Datenbankservern. Dadurch wird mit einer geringeren Anzahl von Lizenzen einer größeren Zahl von Nutzern geholfen.

Ein paar Zahlen

1. Gesicherte Nodes: >300
2. Netto-Platz in den Libraries: 100 TB
3. Datenaufkommen pro Nacht: 150 – 200 GB
4. Versionen, die von einer Datei aufgehoben werden: 4
5. Tage, nach denen eine ältere Version einer Datei gelöscht wird: 31
6. Tage, die eine gelöschte Datei aufgehoben wird: 62

Die letzten drei Punkte bedürfen vielleicht noch einer näheren Erläuterung:

- Punkt 4: Wenn vier Versionen einer Datei gesichert sind und es kommt eine fünfte hinzu, so wird die älteste gestrichen.
- Punkt 5: Eine ältere Version einer Datei wird höchstens 31 Tage aufgehoben, danach wird sie gestrichen. Wenn also eine Datei an drei aufeinanderfolgenden Tagen geändert und gesichert wird, so wird es nach 34 Tagen nur noch die letzte gesicherte Datei geben. Die letzte gesicherte Version der Datei wird nie gestrichen.
- Punkt 6: Löscht der Nutzer seine Datei, so wird die letzte gesicherte Version der Datei noch 62 Tage aufgehoben, er hat also so lange Zeit, die Datei zu restaurieren.

LAN- und LANfree Backup

Wie an den obigen Zahlen zu sehen ist, bringt TSM ein hohes Datenaufkommen mit sich, das das LAN sehr belastet.

Es gibt unter TSM zwei Datenströme, die zu beachten sind:

- Kontrollstrom
- Datenstrom

Der Kontrollstrom enthält die Metadaten (Grösse, Zugriffsrechte, wann zuletzt geändert usw.) zu jeder Datei. Diese Metadaten werden zusammen mit dem Aufbewahrungsort (Tape, Platte) in der Datenbank von TSM verwaltet. Der Datenstrom enthält die eigentlichen Nutzerdaten.

Wie aus Abb. 1 erkennbar, fließen beim LAN Backup alle Daten von den Platten zum Node (in der Abb. ein Fileserver) und von dort über das LAN zum Backupserver und zur Library. Bei 150 – 200 GB ergeben sich daraus drei Nadelöhre:

- Fileserver
- LAN
- Backupserver

Beim LANfree Backup sind die Wege der Ströme getrennt. Der Kontrollstrom (die geringere Datenmenge) nimmt seinen Weg über das LAN zur Datenbank des Backupservers, während der Datenstrom vom Node über das SAN direkt zur Library fließt. Es entfallen also die bottlenecks LAN und Backupserver. Erreicht wird das dadurch, dass der Backupserver dem Fileserver über das SAN temporär ein Laufwerk seiner Library zum ausschließlichen Gebrauch zuordnen kann.

Der Weg des LANfree Backups wird von uns für alle Server, die sich im SAN befinden [1], angewendet – dies ist vom Umfang der Daten her der größere Teil. Alle anderen Nodes werden nach wie vor auf dem traditionellen Weg über das LAN gesichert – was von der Anzahl der Nodes her der größere Teil ist.

Die Masse der Nutzerdaten nimmt also den kurzen Weg über das SAN.

Literatur

[1] Sittel, F.: Institute ans Storage Area Network. *CMSJ25*, 2004, Seite 59 – 61

[2] Weickmann, C.: TSM – 1999 – 2001, Zwei Jahre Datensicherung. *RZ-Mitteilungen* Nr. 22, 2001, Seite 15/16