

TSM – 1999–2001

Zwei Jahre Datensicherung

TSM wird seit zwei Jahren an der HU für die zentrale Sicherung und Archivierung der Daten angewendet. In dieser Zeit entstand ein stabiler Betrieb, der von mehr als 300 Clients zurzeit genutzt wird.

Wie sicher sind meine Daten?

Was kann dem normalen Nutzer mit seinen wertvollen und meist nicht reproduzierbaren Daten passieren? Sie können verändert, gelöscht, verarbeitet, weitergegeben oder bei zwischenzeitlicher Nichtbenutzung langfristig aufgehoben werden. So lange dies unter der bewussten Steuerung des Nutzers geschieht, ist alles in Ordnung (im Allgemeinen) – aber ...

Dieses aber wird für den Nutzer im Fall des Eintretens sehr interessant und kann ihn, wenn er nicht dafür Vorsorge getroffen hat, langfristig beschäftigen, vor allem dann, wenn ein weiterer Fall eintritt, der noch nicht betrachtet wurde – die Havarie einer Platte.

Die einzig sichere Vorsorge, die der Nutzer treffen kann, ist die regelmäßige Sicherung seiner Daten. Sichern heißt hier, dass er seine Daten auf einem zweiten, unabhängigen Medium speichert, um sie im Fall der Fälle von dort wieder restaurieren zu können.

Daraus ergeben sich für ihn einige Überlegungen und Aktivitäten, die ihn in seiner eigentlichen Arbeit stören.

- Er kann z. B. seine Daten auf seinem System spiegeln, wenn er die Möglichkeiten, die meist mit Geld verbunden sind, dafür hat. Spiegeln nützt ihm allerdings nichts, wenn er seine Daten irrtümlich löscht oder wenn er auf einen älteren Stand der Daten zurückgreifen muss.
- Er kann die Daten auf einem Magnetband sichern. Dazu gehört aber auch, dass er eine sehr gute Verwaltung über die gesicherten Daten hat, um sie jederzeit kurzfristig auf seinen Bändern zu finden – wer kann von sich aus schon behaupten, dass ihm dies möglich ist. Und er muss natürlich auch seine von ihm erdachte Sicherung regelmäßig anstoßen.

Nicht erledigt wird mit den angegebenen Sicherungsarten auch die längerfristige Aufbewahrung von zurzeit nicht genutzten Daten, die die Platten des Nutzers unnötig belasten.

Wie hilft mir das Rechenzentrum?

Im Rechenzentrum haben wir uns schon seit längerem mit diesen Problemen beschäftigt, nicht zuletzt deswegen, weil bei uns große Datenmengen anfallen. In den 90er Jahren wurde vom Rechenzentrum für die Universität das HSM-System (Hierarchical Storage Management) UniTree mit dem nachgeordneten Backup-System SM-arch angeschafft. TSM, der Tivoli-Storage-Manager, löste im Herbst 1999 an der Humboldt-Universität diese bis dahin mit Erfolg eingesetzten Systeme ab.

Durch TSM wurde der Schwerpunkt vom HSM auf die Datensicherung und Archivierung verschoben. Dies ergab sich aus einer die Nutzung der Systeme betreffenden Analyse. TSM ermöglicht auch eine gewisse HSM-Arbeit.

Was ist TSM und wie kann ich es nutzen?

TSM ist ein Server-/Clientsystem, das die Daten der Clients – in der TSM-Sprache Nodes – auf dem TSM-Server abhängig von der geforderten Funktionalität sichert oder archiviert. Bei uns werden Unix, Windows 95–2000 und Mac unterstützt. Auf die HSM-Funktionalität soll hier nicht weiter eingegangen werden. Für die Funktion des Backups bzw. der Archivierung besitzt der Server nicht nur eine große Plattenkapazität (bei uns ca. 200 GB), sondern noch eine dahinterliegende Library – dies ist bei uns ein Tape-Roboter mit einer nominalen Kapazität von 20 TB. Der Server verkehrt mit den zu ihm gehörenden Clients über Routinen, die auf diesen installiert werden müssen. Wie man diesen Dienst des RZ nutzen kann, ist unter <http://www.hu-berlin.de/rz/filesv> bzw. <http://www.hu-berlin.de/rz/filesv/tsm.1.1.html> zu finden.

TSM bietet die Möglichkeit, die auf einem TSM-Server gesicherten Daten auf einen zweiten Server zu kopieren und damit doppelt zu sichern. Dies schützt vor eventuellen Tape-Fehlern und ist primär für archivierte Daten wesentlich, da diese im Allgemeinen auf dem Client nicht mehr vorhanden sind. Da das RZ der HU neben dem TSM-Server in Mitte einen analog konfigurierten Server in Adlershof besitzt, wird diese Möglichkeit der „server-to-server-Kopplung“ so genutzt, dass jeder dieser beiden Server die Daten in seiner Umgebung sichert und anschliessend eine Kopie der gesicherten Daten zum anderen Server macht.

Wenn der Nutzer die TSM-Software auf seinem Client installiert hat und das erste Mal einen passwortgeschützten Kontakt mit „seinem“ Server aufgenommen hat, kann er entscheiden (in Zusammenarbeit mit dem zentralen Administrator), ob das Backup vom Server aus (automatisch) erfolgen soll oder ob er selber den Zeitpunkt des Backups bestimmen will (manuell). Das manuelle Backup ist bei uns zwischen 5 und 7 Uhr nicht möglich, da zu dieser Zeit bestimmte Systemarbeiten durchgeführt werden.

Automatisches Backup ist die favorisierte Arbeitsweise. Sie wird auf alle Systeme angewendet, die rund um die Uhr aktiv sind. Durch sie wird der Nutzer am meisten von Überwachungsarbeiten für die Daten-



Abb. 1: TSM-Nutzer GUI

sicherung befreit. Die manuelle Arbeitsweise wird auf Systeme angewendet, die im Allgemeinen am Arbeitsplatz des Nutzers stehen und bei denen nicht davon ausgegangen werden kann, dass sie immer eingeschaltet sind – also PCs und kleinere personengebundene Workstations.

Bei der manuellen Datensicherung muss die Aktivierung vom Nutzer aus erfolgen. Er kann entscheiden, ob er bei der Verbindungsaufnahme jedes mal ein Passwort eingeben will oder ob sich das System selber ein Passwort generiert. Dies ist eine Option im Konfigurationsfile.

Bei der automatischen Datensicherung nimmt der Server zu bestimmten (voreingestellten) Zeiten den Kontakt zum Client auf. Dabei überprüft er, ob es sich um den bei ihm registrierten Client handelt und ob das Passwort noch aktuell ist. Ist dies nicht der Fall, wird die Verbindung abgelehnt und der zentrale Administrator erhält eine Information. Nach der erfolgreichen Verbindung erfolgt ein inkrementelles Backup.

Haben sich zwischen zwei inkrementellen Backups Files auf dem Client verändert, werden diese Files gesichert. Die alten Daten verfallen jedoch nicht, sondern werden nach zentralen Vorgaben noch einige Zeit aufgehoben. Es gibt hierfür unter TSM die Begriffe *active version* und *inactive version*. Eine ausführlichere Beschreibung dieser Begriffe ist unter <http://www.hu-berlin.de/rz/filesv/tsm.1.1.html> zu finden. Bei uns werden vier Versionen eines Files für eine gewisse Zeit aufgehoben. Löscht der Nutzer ein File, hat er noch zwei Monate Zeit, es aus dem Backup zu restaurieren.

Unabhängig von *automatisch* oder *manuell* lässt das System das Passwort nach einer voreingestellten Zeit (90 Tage) verfallen. Danach muss der Nutzer bei Kontaktaufnahme das Passwort ändern bzw. es wird auto-

matisch ein neues eingetragen abhängig von der jeweiligen Option in den Konfigurationsfiles. Dies soll gewährleisten, dass nur der Eigentümer von Daten an diese kommt und kein Fremder – auch nicht der zentrale TSM-Administrator!

Muss der Nutzer Daten restaurieren, hat er grundsätzlich zwei verschiedene Möglichkeiten. Er kann mit einem Graphical User Interface (GUI; Aufruf `dsm`) oder mit dem Command Line Interface (CLI; Aufruf `dsmc`) arbeiten. Die Arbeit mit dem CLI ist etwas gewöhnungsbedürftig – man sollte sich aber mit ihm beschäftigen, da er günstig für die Restauration von großen Datenmengen (großen Verzeichnissen, ganzen Platten) ist. Sehr brauchbar ist seine Helpfunktion (Aufruf: `dsmc help`), mit der man sich über alle Optionen und Möglichkeiten informieren kann.

Der GUI (Abb. 1 zeigt den GUI für einen PC) ist im Allgemeinen selbsterklärend. Man kann ihn nicht nur zum *Backup/Restore* bzw. *Archive / Retrieve* benutzen, sondern auch für Verwaltungsarbeiten. Hier soll nur auf die Helpfunktion hingewiesen sein und die Möglichkeit der Veränderung in den Konfigurationsfiles mittels *Edit_Preferences* bzw. der Änderung des Passwortes mittels *Utilities_Change Password...* erwähnt werden.

Beim Restaurieren von Daten muss der Nutzer auch beachten, in welcher Version die zu restaurierenden Daten vorliegen. Standardmäßig sieht er im GUI und CLI nur die *active version*. Wenn er eine ältere Version braucht, muss er das im GUI über *Restore_View_Display active/inactive Files* fordern.

Wie entwickelt sich TSM an der HU weiter?

Für den Nutzer werden die weiteren TSM-Entwicklungen nicht vordergründig sichtbar sein. Mit dem Umzug des RZ in das IKA nach Adlershof wird die Einführung des SAN vorangetrieben (vgl. Artikel *Storage-Area-Network an der Humboldt Universität* in diesem Heft). Dort werden wir für die großen zentralen File-Server das LANfree Backup einführen, durch das das LAN von den Datenströmen, die beim Backup fließen (Stand heute: ca. 140 GB pro Tag von über 300 Clients), entlastet wird.

Christoph Weickmann
weickmann@rz.hu-berlin.de

So war es zu lesen in den RZ-Mitteilungen Heft Nr. 19/2000 *ADSM (Adstar Distributed Storage Manager) ist ein IBM-Produkt, das während der Zeit der Beschaffung von Tivoli – einer IBM-Tochter – übernommen wurde. Dabei wurde ADSM zu TSM (Tivoli Storage Manager). Es werden also gewiss in den nächsten Zeiten die Begriffe ADSM und TSM gleichbedeutend benutzt.*