

Backup-Strategie für den Dokumentenserver

Niels Fromm | fromm@cms.hu-berlin.de

Ein wichtiger Teil des Konzeptes für die Sicherheit der auf dem Dokumentenserver abgelegten Daten ist eine Backup-Strategie, die diese Daten vor Verlust oder Beschädigung schützt. Ein Verlust von Daten oder deren Beschädigung kann durch fehlerhafte Hard- oder Software, unbeabsichtigte menschliche Fehler, aber auch durch gezielte Manipulation, etwa durch Schadprogramme wie Viren oder Würmer, entstehen. Um in einem solchen Fall die verlorenen oder beschädigten Dateien wiederherstellen zu können, muss es eine Kopie dieser Dateien, ein Backup, geben.

für den Dokumentenserver. Der Computer- und Medienservice (CMS) bietet daher für die gesamte Universität einen Backup-Service [1] an. Dieser Service bietet die Möglichkeit, alle Daten eines Servers automatisch jede Nacht in den Backupspeicher der Universität zu übertragen. Dieser Speicher besteht aus zwei Magnetband-Speichertürmen mit rund 200 TB Speicherkapazität, wobei sich ein Speicherturm im Rechnerraum in Mitte, der andere im Erwin Schrödinger-Zentrum in Adlershof befindet.

Der Inhalt beider Speichertürme wird jede Nacht synchronisiert, so dass



Abb. 1: Backup-Speicherturm des CMS im Erwin Schrödinger-Zentrum

Der folgende Artikel bietet eine Übersicht über die Backup-Strategie der Humboldt-Universität zu Berlin und beschreibt den Ansatz der Arbeitsgruppe Elektronisches Publizieren für die Sicherung der Daten auf dem Dokumentenserver.

Die Notwendigkeit von einem Backup der Daten besteht für alle Server der Humboldt-Universität zu Berlin, nicht nur

zu jedem Zeitpunkt alle Daten sowohl in Mitte als auch in Adlershof komplett vorhanden sind. Dadurch ist sichergestellt,

dass alle Daten selbst bei komplettem Ausfall eines Speicherturmes noch verfügbar sind.

Aus diesem Backupspeicher lassen sich beschädigte oder verlorene Dateien der letzten 60 Tage wiederherstellen. Der Backup-Service des CMS bietet seinen Nutzern zudem eine Versionierung, d. h. die vier letzten Versionen einer Datei aus diesem Zeitraum können zurückkopiert werden. Daten von Servern, die den Backup-Service nutzen, sind also dreifach redundant gespeichert, einmal auf dem lokalen Speicher des Servers selbst und zusätzlich zweimal auf den zentralen Backup-Speichertürmen.

Das beschriebene Backup der Daten von Servern auf einen zentralen Backupspeicher auf Magnetbandbasis ist die traditionelle, aber nach wie vor aktuelle und praktisch überall in Rechenzentren eingesetzte Methode, da nur sie die relativ kostengünstige Speicherung von sehr großen Datenmengen ermöglicht.

Ein Nachteil der Speicherung des Backups auf Magnetbändern ist die im Vergleich relativ geringe Geschwindigkeit der Medien. So kann das Wiederherstellen von größeren Datenmengen einige Zeit dauern.

Für das schnelle Wiederherstellen von Servern im Havariefall, also z. B. bei Ausfall von Komponenten des Servers, ist diese Art des Backups nicht unbedingt geeignet. Für Server wie den Dokumentenserver, die nach Möglichkeit jederzeit ihre Dienste anbieten sollen, müssen zusätzliche Sicherungsmaßnahmen getroffen werden. In Bezug auf die Sicherheit der Daten ist dafür vor allem ein ausfallsicheres, hochverfügbares Speichersystem für die Server nötig.

Der CMS hat aus diesem Grund vor einiger Zeit ein sogenanntes Storage Area Network (SAN) aufgebaut (siehe [2], [3]). Es besteht aus Festplattenspeichersystemen, die mit den Servern über ein eigenes Netz verbunden sind. Die Verbindung eines solchen Speichernetzwerkes wird typischerweise über Lichtwellenleiter (LWL), also Glasfaserkabeln hergestellt, sie ist damit unabhängig von der Anbindung des Servers an das normale Netzwerk. Ein an ein SAN angeschlossener Server kann also Speicher nutzen, der an be-

liebiger Stelle in diesem Netz angeboten wird, der Speicher muss nicht mehr in dem Server selbst vorhanden sein.

Dies ermöglicht eine zentrale Administration und vor allem Planung und damit effiziente Nutzung der gesamten Festplattenkapazitäten. In vielen Fällen ist dieser Speicher auch leistungsfähiger als lokal in Server eingebaute Festplatten.

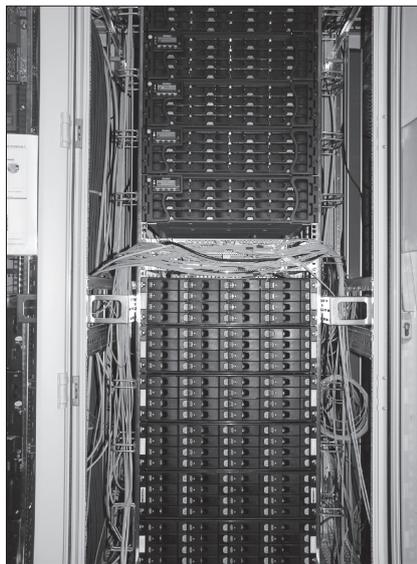


Abb. 2: SAN-Festplattensysteme im Erwin Schrödinger-Zentrum

Das SAN der Humboldt-Universität zu Berlin ist vor allem unter dem Gesichtspunkt der Ausfallsicherheit entwickelt worden. Das Konzept des CMS für die Benutzung des SAN sieht Redundanz auf jeder Ebene vor, damit der Ausfall einer Komponente nicht den Ausfall des gesamten Systems nach sich zieht. So muss zunächst jeder Server zweifach an das SAN angeschlossen werden und auch die gesamte Verbindung zwischen den Servern und den Speichersystemen ist redundant ausgelegt. Damit ist der Speicher für den Server auch dann noch verfügbar, wenn ein Anschluss an das SAN gestört ist.

Der Speicher im SAN besteht aus redundanten Festplattensystemen, die auf verschiedene Gebäude der Universität verteilt sind und den Ausfall einer einzelnen Festplatte ohne Verlust von Daten überstehen können. Als zusätzliche Sicherheit werden jeweils zwei dieser Festplattensysteme gespiegelt und den Servern als ein Speichersystem angeboten. Alle Daten werden daher gleichzeitig auf zwei der redundanten Speichersysteme geschrieben, die sich auch an unterschiedlichen Orten befinden können. Somit kann auch ein gesamtes Speichersystem ausfallen, ohne die Sicherheit der Daten zu gefährden. Weiter bieten die Speichersysteme im SAN zusätzliche Möglich-

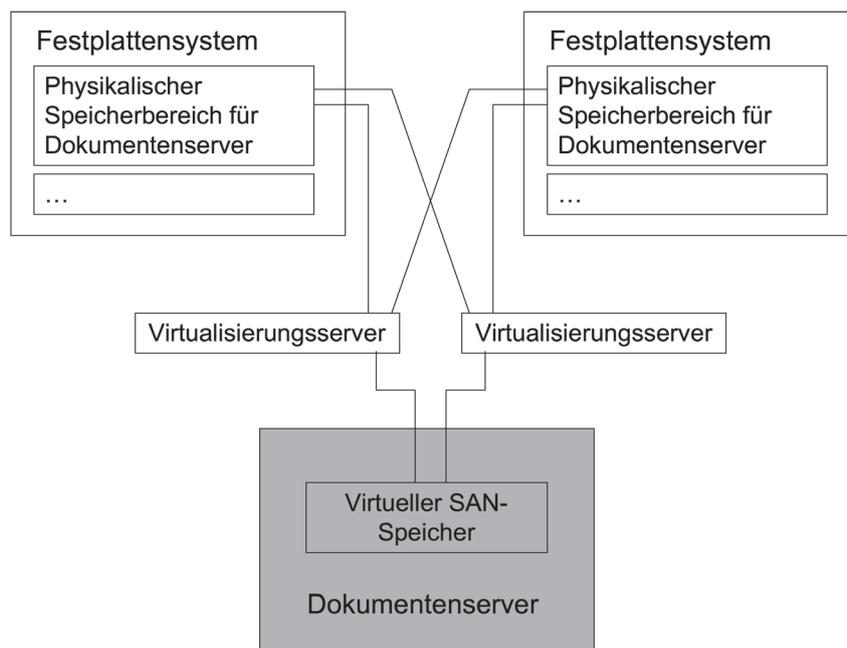


Abb. 3: Anbindung Dokumentenserver – SAN

keiten der Datensicherung wie Snapshots, die bei der Backup-Strategie für den Dokumentenserver Anwendung finden.

Der Dokumentenserver der Humboldt-Universität zu Berlin nutzt zurzeit zwei virtuelle Festplatten mit je 400 GB Kapazität aus dem SAN. Der Dokumentenserver befindet sich im Rechnerraum des CMS im Erwin Schrödinger-Zentrum in Adlershof. Die 800 GB der beiden virtuellen Festplatten befinden sich physikalisch in Festplattensystemen in den Rechnerräumen des CMS in Adlershof und im Hauptgebäude der Universität in Mitte. Jeden Monat einmal werden diese 800 GB des Dokumentenservers als zusätzliches Backup auf andere Speichersysteme im SAN der Universität abgelegt. Dieses zusätzliche Backup liegt in Speichersystemen im Erwin Schrödinger-Zentrum und im Gebäude der Physik in Adlershof. Weiterhin wird jeden Tag ein sogenannter Snapshot der Daten des Dokumentenservers im SAN angelegt. Ein Snapshot von einem Dateisystem kann zu jedem beliebigen Zeitpunkt erstellt werden. Jede noch so kleine Änderung an Daten in diesem Dateisystem nach diesem Zeitpunkt wird im Snapshot festgehalten und kann dadurch rückgängig gemacht werden. Da in einem Snapshot alle Änderungen an den Daten seit seiner Erstellung gespeichert werden, vergrößert sich ein Snapshot mit jeder Änderung, er wächst also mit der Zeit. Je mehr Daten sich in diesem Dateisystem verändern, desto schneller wächst die Größe des Snapshots.

Um nicht zuviel Speicherplatz zu verwenden, wird die Größe aller Snapshots für den Dokumentenserver auf ein TB begrenzt und bei Erreichen dieser Grenze werden automatisch die ältesten Snapshots gelöscht bzw. überschrieben. Der Speicherplatz für diese Snapshots befindet sich physikalisch in Speichersystemen im ESZ und im Gebäude der Informatik und reicht bei täglichen Snapshots für aktuell 180 Tage. Alle Änderungen an den im Dateisystem abgelegten Daten des Dokumentenservers der letzten 180 Tage können also rückgängig gemacht werden, also z. B. auch gelöschte oder beschädigte Dateien wiederhergestellt werden.

Selbstverständlich nutzt der Dokumentenserver auch den Backup-Service

des CMS, jede Nacht werden daher die Daten des Dokumentenservers auf den beiden Backup-Speichersystemen der Universität gesichert.

Die Backup-Strategie für den Dokumentenserver legt also fest:

- die Nutzung von Speicher im SAN und damit die redundante Speicherung dieser Daten im Erwin Schrödinger-Zentrum und in Mitte,
- die tägliche Erstellung von Snapshots von diesen Daten und die redundante Speicherung der Snapshots im Erwin Schrödinger-Zentrum und in der Informatik,
- das tägliche Backup der Daten in die redundanten Backup-Speichersysteme im Erwin Schrödinger-Zentrum und in Mitte
- und die monatliche Kopie aller Daten sowie deren redundante Speicherung in Mitte und der Physik.

Damit sind alle Daten des Dokumentenservers nach einem Monat physikalisch sechs Mal an vier verschiedenen Orten Berlins vorhanden, was einen sehr hohen Grad an Sicherheit für diese Daten bedeutet. Denn dass alle acht Kopien der Daten an den verschiedenen Standorten gleichzeitig verloren gehen oder zerstört werden, kann als nahezu unmöglich bezeichnet werden.

Die Anzahl der vorhandenen Kopien von Daten wird demnach als Maß für die Sicherheit dieser Daten vor Verlust oder Beschädigung gesehen. Aber die räumliche Trennung der verschiedenen Kopien ist ebenso wichtig wie die Anzahl der Kopien, denn existieren alle Kopien an einem Ort, so können sie auch gleichzeitig zerstört werden.

Bei der Erstellung der Backup-Strategie für den Dokumentenserver wurde daher versucht, eine hohe Anzahl von Kopien der Daten auf möglichst viele unterschiedliche Standorte zu verteilen, um ein Höchstmaß an Sicherheit zu erreichen. Dies ist mit der vorliegenden Strategie gelungen, wobei dies nur möglich ist, da die Humboldt-Universität zu Berlin mit dem SAN über die nötige technische Infrastruktur verfügt und die Datenmenge mit weniger als 800 GB noch so klein ist, dass die Erstellung von mehreren Kopien kein Problem darstellt.

Literatur

- [1] WEICKMANN, CHRISTOPH: *Backup – Datensicherheit für alle*. cms-journal Nr. 25, 2004, System- und Netzdienste S. 63–64
- [2] SITTEL, FRANK: *Storage-Area-Network an der Humboldt-Universität*. cms-journal Nr. 22, 2001, 10 Jahre RZ-Mitteilungen S. 10–14
- [3] SITTEL, FRANK: *Institute ans SAN*. cms-journal Nr. 25, 2004, System- und Netzdienste S. 59–61