Ist Datensicherung eine Wissenschaft?

Daten, von denen keine Sicherungskopie existiert, sind per Definition unwichtig (anonym)

Datensicherung ist bei der heutigen Datenabhängigkeit jeder Produktion zentraler als es vielen bewusst ist. Das noch immer zunehmende Datenwachstum sowie das lückenhafte Wissen um professionelle Sicherungsplanung und Verfahren schafft eine bedrohliche Diskrepanz und führt oft zu Vermeidungshaltung. Die Berücksichtigung einfacher Grundregeln ermöglicht jedoch die schrittweise Professionalisierung der Datensicherung. Damit wird auch der potentiellen Existenzbedrohung der Firma durch einen größeren Datenverlust vorgebeugt. Mit der Kombination aktueller Technologien lässt sich ein professionelles Sicherheitsniveau erreichen, das zudem Transparenz in der Datenhaltung schafft.

Data protection plays a more central role today in production than ever before, thanks to our dependence on data. An increasing amount of data, combined with patchy understanding of professional data security planning, creates a dangerous discrepancy that often leads to a position of avoidance. Taking basic ground rules into consideration allows gradually professionalizing data security. Thus, the threat of data loss destroying a company is prevented. With a combination of current technologies, a new level of professional data security can be reached, also offering transparency of data storage.

Backups are one area where „Paranoia is prudent“ (James Pond)

Der Vorgang des Kopierens (von lat. copiare = mehrfaches Abschreiben) ist schon fast so alt wie die Schrift selbst. Klostermönche sorgten im Mittelalter dafür, dass die Literatur aus der Antike durch Abschriften so weit verbreitet wurde, dass ein bedeutender Teil alle Feuer, Kriege und andere Unglücke überstand. Diese Funktion hat auch ein heutiges Backup. Doch ist Datensicherung heute noch nötig, wo alle Systeme immer zuverlässiger werden? Tatsächlich führt die steigende Komplexität sowohl der Einzelkomponenten als auch des Workflows insgesamt sogar zu höherer Anfälligkeit. Das größte Risiko stellen jedoch die Benutzer selbst dar. Gerade im hektischen Produktionsalltag passieren alle Fehler, die passieren können. **Datensicherung ist nötig, um die unübersehbaren Wechselwirkungen und Fehlerquellen abzudecken.**

Grundregeln der Datensicherung

1. Automatische Sicherung
2. Mehrfache/mehrstufige Sicherung
3. Vollständigkeit für Einsatzzweck
4. Medienwechsel Disk/Tape
5. Auslagerung
6. Restore-Test

Was kostet es, nicht zu sichern?

Was kostet Ausfallzeit? Im Falle eines Falles, das heißt, wenn Produktionsdaten nicht mehr erreichbar sind, betrifft das zunächst den Produktionsablauf, letztlich also die Kunden. Je nach Geschäftstyp kann man sie kürzer oder länger vertrösten. Je länger der Ausfall dauert, umso mehr Kunden erfahren davon. Zusätzlich sind die eigenen Mitarbeiter zum Nichtstun verurteilt. Die Produktion kann nicht weitergeführt werden. Außerdem beeinflusst wiederholter Ausfall auch die Atmosphäre und Arbeitseinstellung der Mitarbeiter negativ. Gehen Daten ganz verloren, kann der Schaden extrem kostspielig bis zu unbezifferbar sein und die Firma als Ganzes gefährden. Amerikanische Untersuchungen zeigen, dass ein Großteil der Firmen, die von einem größeren Datenverlust betroffen sind, später Konkurs anmelden muss [1].

Geplantes Erschrecken

Die Sicherung kann immer nur so gut sein wie die Ausfallszenarien, die ihr zugrunde liegen. Daher ist es sinnvoll, sich hier vom Ende, nämlich dem schlimmsten Szenario, aus zu nähern. Schon der lokale Blick auf die eigene Serverinstallation gibt einen guten Anfangspunkt. Zum Beispiel kann ein Defekt in einem Netzteil einen Server in Brand setzen. Auch die benachbarten Systeme sind dann davon betroffen, denn möglicherweise sind alle Systeme im gleichen Raum und in der Nähe. Der primäre Speicher und seine lokale Sicherung versagen dann eventuell gleichzeitig. Zusätzlich geht die Information über die konkrete Konfiguration dieser Systeme dabei verloren. Daraus folgt, dass der Zugriff auf wichtige Informationen zur Konfiguration des gesamten Setups im Notfall von entscheidender Bedeutung ist. Dazu gehört auch die Konfiguration der Server und ihrer Dienste, SAN-Systeme und des Netzwerks. Im Notfall steht eben keines der Systeme mehr zur Verfügung, um nachzusehen, wie die Konfiguration war, welche IP-Adressen genutzt werden oder wie andere Spezifikationen lauteten. Eine umfassende Dokumentation sowie ein Schritt-für-Schritt-Plan ist daher notwendig, um alle nötigen Informationen zur Wiederherstellung zu bündeln. Dabei gilt es folgende Fragen zu beantworten:

- Was passiert, wenn der zentrale Speicher zerstört wird?
- Was passiert, wenn die Datenbanken mit allen Adressen, Verträgen und Finanzdaten zerstört werden?
- Wie weit entfernt müssen Daten ausgelagert werden, um dieses Szenario zu überstehen?

Ein schlüssiges Disaster-Recovery-Planning erarbeitet Antworten auf diese drängenden Fragen [2].

Nur eine **automatische Sicherung** schützt zuverlässig. Jeder manuell anzustoßende Prozess kann vergessen und verschoben werden, besonders wenn Zeitdruck besteht. Doch gerade dann ist die Wahrscheinlichkeit für Nutzerfehler erhöht und damit eine Sicherung besonders nötig.

Mehrfache bzw. mehrstufige Sicherungen schaffen zusätzliche Sicherheit. Eine mehrfache Sicherung ermöglicht die Auslagerung von Daten. Eine mehrstufige Sicherung kann durch eine geschickte Kombination die Vorteile verschiedener Verfahren und Medien nutzen.

Bei **Disk2Disk2Tape** nutzt man die jeweiligen Vorteile von Disk und Band (Tape). **Disk-to-Disk** bietet eine schnelle Sicherung und schnelle Wiederherstellung von Dateien.

Dr. med. Marc M. Batschkus ist Arzt für Medizinische Informatik, Hochschuldozent und Wissenschaftler. Bei Archiware verantwortet er den Bereich Business Development und Marketing Science/Medicine/Media.

Die Sicherung auf Tape geschieht vom Disk-Sicherungsvolume. Von dort aus kann das Tape seine überlegene Streamingfähigkeit nutzen und auf die gesamte Sicherung zugreifen. Durch den Kostenvorteil von Tape pro TB kann hier – gegenüber Disk – die Vorhaltezeit wesentlich verlängert werden. Zusätzlich kann eine Auslagerung bereits geschriebener Tapes einen professionellen Sicherheitsgrad gewährleisten [3].

Vollständig zu sichern ist nicht ganz so einfach wie es sich anhört. Das beginnt mit einer Auflistung aller Server und Arbeitsplatzrechner (Workstations) sowie ihrer Festplatten, genauer der darauf eingerichteten Partitionen. Eine Netzwerk- und Client-Verwaltungssoftware kann Listen von Rechnern und ihrer Software anlegen. Die zentrale Verwaltung von Softwarelizenzen sowie Images zur schnellen Installation erleichtern die Wiedereinrichtung. Die Dokumentation sollte mehrfach, außer Haus und immer erreichbar gesichert sein. Besonders aufwendig konfigurierte Rechner mit zahlreichen Tools, Plugins und Anwendungen sind im Notfall nur mit viel Zeitaufwand zu rekonstruieren. Es lohnt sich daher diese vollständig zu sichern. Für Mac-OS-X-Rechner kann hier eine boot-fähige Sicherung die Wiederaufnahme der Produktion extrem beschleunigen.

Der **Wechsel des Speichermediums,** also die Verwendung unterschiedlicher Technologien wie Disk und Tape, dient der Risikominimierung. Während eine Disk („Online Storage“) im Betrieb von Nutzerfehlern, Malware und Defekten betroffen sein kann, sodass sogar der gesamte Speicher in einem Zug gelöscht werden kann, ist Tape („Offline Storage“) praktisch nur durch unsachgemäße Lagerung und Gewalteinwirkung zu zerstören.

Auslagerung (Off-Site-Speicher) gehört zu jeder professionellen Sicherungsstrategie. Größere lokale Schäden im Bereich einer Firma haben wahrscheinlich auch Auswirkungen auf die lokale Sicherung. Zusätzlich kann auch ein Schaden, der allein innerhalb der IT-Anlage auftritt, das Backup beschädigen oder unbrauchbar machen. Denkbar sind hier sowohl Anwenderfehler wie Fehlkonfiguration, Update oder Umbau der Hardware als auch lokale Schäden wie Brand, Blitzschlag, Stromausfall usw. Für alle genannten und weitere Fälle stellt die Auslagerung von Sicherungen die Rettung dar. Damit erreicht man eine professionelle Sicherheitsklasse, die jedoch die Organisation einer Medienrotation erfordert. Als Medium sind Bänder (Tapes) wesentlich leichter und sicherer auszulagern als Disks. Bei Tape-Bibliotheken, sogenannten Libraries, kann man das durch Magazinwechsel auch mit mehreren Tapes leicht durchführen. Eine professionelle Medienrotation benötigt mindestens drei Tape-Sätze. Einer befindet sich am Auslagerungsort, einer wird aktuell in der Firma beschrieben und einer befindet sich auf dem Transportweg zum/vom Auslagerungsort. Auf diese Weise ist auch die Beschädigung beim Transport abgedeckt.

Wie steht es mit einer Sicherung in der Cloud?

Cloud-Computing ist sicherlich einer der aktuellsten Technologietrends. Hinter dem Begriff verbergen sich zahlreiche Technologien und Szenarien. Nur einige haben mit Sicherung zu tun. Doch was kann eine professionelle Sicherung in der Cloud, worauf ist zu achten und für welche Einsatzgebiete eignet sie sich?

Zunächst gilt es die gesamte Kette von Übertragungen zu betrachten. Für die Sicherung benötigt man ausreichend Upload-Bandbreite. Typische DSL-Leitungen haben eben gerade dort technologiebedingt ihre Asymmetrie, das heißt, eine geringere Upload- als Download-Bandbreite – je nach Typ im Verhältnis 10:1 bis 24:1 [4].

Datenleitungen mit größeren Upload-Bandbreiten bzw. symmetrische Leitungen sind noch immer ein beträchtlicher Kostenfaktor. Zusätzlich gilt es zu beachten, ob und gegebenenfalls welche Einschränkungen vom Provider vorgegeben werden. Das kann eine Durchsatzdrosselung ab einer bestimmten Datenmenge sein, der Ausschluss bestimmter Datenarten (zum Beispiel Videodateien) und anderes mehr. Im Falle eines Falles geht es dann darum, die Daten möglichst schnell wieder zurückzubekommen, das heißt, die maximale konstante Download-Rate ist zu testen. Diese kann je nach Tageszeit und Auslastung des Providers oder auch des Netzes dazwischen beträchtlichen Schwankungen unterliegen. Bei einem Sicherungsvolumen von mehreren TB können hier durch diese Schwankungen allein schon viele zusätzliche Stunden Wartezeit entstehen. Die Nutzung mehrerer Cloud-Anbieter erfordert für jeden Dienst eine eigene Software. Leichter lassen sich eigene Erfahrungen mit Cloud-Storage mit dem kostenlosen Tool „WingFS“ von Archiware sammeln. Es ermöglicht die übergreifende Verwendung mehrerer der gängigsten Cloud-Services [5].

Ein ganz anderes Thema ist die Sicherheit der Daten beim Provider. Alle US-Firmen und auch deren europäische Niederlassungen haben den amerikanischen Behörden auf Anfrage („Patriot Act“ und „FISA Amendments Act of 2008“) Zugang zu den Daten einzuräumen – und dürfen das ausdrücklich den betroffenen Kunden auch nicht im Nachhinein mitteilen.

Hier wird allen möglichen Verwendungen und menschlichen Fehlern Tür und Tor geöffnet. Deutsche Firmen, die ihre Kundendaten in dieser Weise verwalten, verletzen wahrscheinlich sogar deutsches Datenschutzrecht. Microsoft und Google haben diese Weitergabe bereits bestätigt [6].

Wie weit sich die Interessen von Nutzern und Dienstanbietern widersprechen, wird eindrucksvoll in der Dissertation von Christopher Soghoian nachgewiesen [7].

Schlussfolgerung

Eine Cloud-Sicherung kann kein traditionelles Onsite-Backup ersetzen. Höchstens als zusätzliche Auslagerung ist es zu betrachten, wenn die genannten Kriterien sinnvoll geklärt werden können und die zu übertragende Datenmenge begrenzt ist [8].

Der **Test des Restore-Vorgangs** sollte regelmäßig vorgenommen werden. Die Sicherungssoftware ist nur ein Teil dieses Vorgangs. Kritischer sind andere, besonders auch organisatorische Aspekte.

Welche Schritte spielen beim Restore-Vorgang eine Rolle?

Zunächst löst jemand den Restore von Dateien, Verzeichnissen oder auch ganzen Speicherbereichen oder Servern aus. *Wer kann/darf das in einer Firma?* Der nächste Schritt ist die Entscheidung, wohin etwas wiederhergestellt werden soll. Dabei macht es einen großen Unterschied, ob eine Datei, ein Serververzeichnis oder ein ganzes SAN wiederhergestellt werden soll:

- *Gibt es ausreichend Reserve-Storage für diesen Fall?*
- *Fällt ein Server aus, gibt es virtuellen oder physischen Ersatz?*
- *Wer hat Zugriff auf die nötigen Ressourcen, Passwörter und Accounts?*
- *Wie ist das genaue Vorgehen für welchen Ausfall?*
- *Wie oft wird dieser getestet?*

Das sind viele Fragen, doch sie genau und zwar vorab zu betrachten und Antworten zu dokumentieren ist nötig, um im Notfall zielgerichtet vorgehen zu können. Die Lieferung eines neuen Rechners oder Servers ist ein günstiger Zeitpunkt für einen Restore-Test. Mit geringem Aufwand kann dieser als Ersatz-Hardware eingesetzt werden, um aus dem Backup beladen zu werden, bevor er für seinen eigentlichen Einsatzzweck konfiguriert wird.

Die gesamte IT-Struktur sollte solide aufgesetzt und betrieben werden. Im Detail

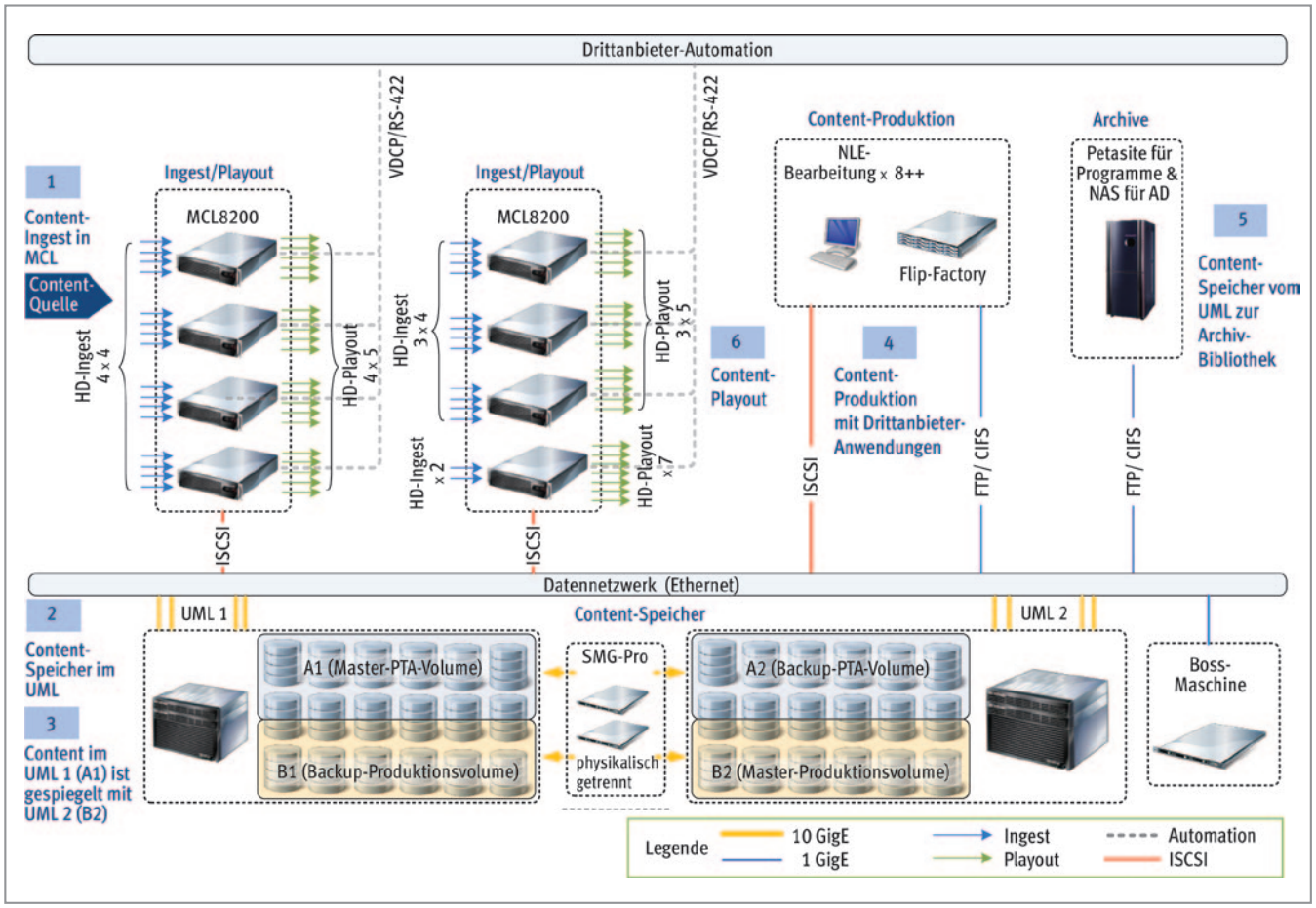

Bild 1. Beispiel für die Verwendung einer Verfügbarkeitslösung mit Cloning zwischen Speicher A1 und A2 bzw. B1 und B2: MTV3 Finnland mit XOR-Speicher und P4-Synchronize

heißt das, dass die Umgebungstemperatur sowie Feuchtigkeit kontrolliert werden müssen, um Korrosion und Alterung durch zu feuchte und elektrostatische Ladung durch zu trockene Luft zu vermeiden. Alle Komponenten wie Server und RAIDs enthalten nur für sie zertifizierte Teile. Der Administrator ist für seine Systeme geschult und dokumentiert deren Setup und Änderungen sowie Ausfallpläne. Erst diese Kombination professioneller Bedingungen vermindert Risiken und sorgt für ein hohes Maß an Zuverlässigkeit. Damit reduziert sich auch die Wahrscheinlichkeit, dass auf das Backup zurückgegriffen werden muss.

Backup-Planung beginnt vom Ende her

Bei der grundlegenden Entscheidung für ein Sicherungsverfahren geht es zuerst um die Ausfalltoleranz. Es geht um die Frage: Wie lange komme ich ohne die Daten aus? Diese Angabe wird im Datensicherungsjargon *Recovery-Time-Objective* (RTO) genannt. Es ergeben sich drei Hauptkategorien:

- zeitkritisch, also Minuten,
- Stunden und
- einen oder mehrere Tage.

Danach ist die Vorhalte- oder Retention-Zeit zu entscheiden, also wie weit man zeitlich im Datenbestand zurückgehen können will:

- kurz (Tage),
- mittel (Wochen) oder
- lang (Monate bis Jahre).

Daraus folgt die Entscheidung für das Sicherungsmedium bzw. eine Kombination von Medien wie zum Beispiel

- Disk = kurz bis mittel (Tage bis Monate) und
- Tape = lang (Monate bis Jahre).

Wie lange kann man ohne die Daten auskommen?

Die Sicherungsmethode wird ganz wesentlich davon bestimmt, wie lange man ohne einen jeweiligen Datenbereich auskommen kann (RTO). Zeitkritische Daten, die bei einem Ausfall binnen Minuten wieder benötigt werden, können zum Beispiel nur mit einer Datenverfügbarkeitslösung gesichert werden. Diese auch Ausfallsicherung genannte Methode erlaubt das „Umschalten“ auf den geklonten bzw. gespiegelten Speicherbereich – hier entfällt ein Restore-Prozess. Das Eingreifen eines Administrators ist jedoch erforderlich, um die Daten wieder online zu bringen. Eine solche Ausfallsicherung ist extrem attraktiv – besonders auch für SAN-Systeme, also Speicher, auf den mehrere Nutzer mit hoher Last zugreifen können. Ausreichender sekundärer Speicher, auf den geklont werden kann, ist nötig sowie dessen möglichst direkte Anbindung an das SAN. Es ist darauf zu achten, dass tatsächlich alle Dateiattribute übertragen werden, also ACLs, extended Finder-Attributes, Xsan-Attributes usw. **(Bild 1)**.

Ausgelöst durch eine Modernisierung wurden bei MTV3 SD- und HD-Speicherbereiche zusammengelegt, hoch performanter XOR-Speicher redundant angelegt und mittels Cloning mit P4-Synchronize (Archiware) verfügbar gehalten. Darüber hinaus kann P4-Synchronize alle Dateiversionen und Snapshots des gesamten Speichers vorhalten.

Bei geringeren Anforderungen an die Ausfalltoleranz, also wenn die Wiederherstellungszeit auch mehrere Stunden betragen darf, kann ein traditionelles Backup eingesetzt werden. Hierbei wird immer auch eine

Laufzeit für den Wiederherstellungsprozess benötigt, die direkt abhängig ist von der angeforderten Datenmenge. Handelt es sich um viele TB, so kann die Wiederherstellung mehrere Stunden betragen. Von Vorteil ist es, den Nutzern selbst (eventuell selektierten) Zugriff auf die Sicherung zu geben. Kleine Missgeschicke können auf diese Weise selbst behoben werden und der Administrator wird entlastet.

Backup-Arten

Unterschiedliche Sicherungsarten stehen zur Verfügung, um die jeweils beste Sicherung für die jeweiligen Gegebenheiten zu erstellen. Da die Begriffe innerhalb verschiedener Lösungen unterschiedlich verwendet werden, ist es sinnvoll, sich gründlich mit der jeweils verwendeten Lösung und ihren Begrifflichkeiten auseinanderzusetzen.

Zunächst gibt es die **Vollsicherung**. Wie der Name bereits suggeriert, werden dabei alle Daten gesichert. Die **inkrementelle Sicherung** sichert alles, was seit der letzten Sicherung neu erstellt oder geändert wurde. Die **differentielle Sicherung** sichert, was sich seit der letzten Vollsicherung geändert hat. Zusätzlich gibt es noch Sonderfälle wie die **permanente inkrementelle Sicherung,** die ohne vorheriges Voll-Backup auskommt. Diese heißt in P4-Backup von Archiware *Progressive-Backup* und sorgt mit zusätzlichen Mechanismen für ausreichende Redundanz der Dateien ohne Vollsicherung, das heißt, wenn die Datenmenge für eine Vollsicherung bei gegebenem Zeitfenster zu groß ist. Das kann leicht passieren. In Abhängigkeit vom Durchsatz und Zeitfenster sind an einem Wochenende (Fr. 18 h – Mo. 6 h = 60 h) bei 100 MB/s Durchsatz etwa 21 TB zu sichern. In der heutigen Speicherlandschaft ist das ein eher geringer Wert. Da der Durchsatz nicht im gleichen Maße wie die Speichergröße wächst, ist davon auszugehen, dass zunehmend Backup-Zeitfenster an ihre Grenzen stoßen werden.

Wann, wie oft und in welcher Kombination die Sicherungsarten verwendet werden, hängt von den spezifischen Anforderungen ab. Folgende Fragen sind dabei zu stellen:

- **Wann kann eine Sicherung laufen, ohne die Produktion einzuschränken?** Bei zunehmenden Produktionszeiten, die teilweise aus verschiedenen Zeitzonen beschickt werden, kann das Festlegen eines Backup-Zeitfensters eine Herausforderung sein. Möglicherweise muss die Infrastruktur aufgerüstet werden, um Produktion und Sicherung gleichzeitig unterstützen zu können. Das kann auch über die Einführung von Teilnetzen oder exklusive Kommunikationswege für verschiedene Speicherbereiche geschehen. Generell ist es sicherer, nachts zu sichern, da die meisten Dateien dann in einem stabilen Zustand sind und zudem die Infrastruktur weniger belastet ist.
- **Wie oft muss gesichert werden, um den möglichen Verlust tolerabel zu halten?** Die Frequenz der Sicherungsläufe bestimmt den maximal möglichen Datenverlust (RPO = Recovery Point Objective). Unterschiedliche Speicherbereiche und Daten haben meist auch unterschiedliche Verlusttoleranzen. Eine Segmentierung verschiedener Datenbereiche kann hier die effiziente Versorgung nach unterschiedlichen Anforderungen unterstützen. Eine tägliche Sicherung erscheint als das absolute Minimum in jeder Produktionsumgebung.
- **Was ist die beste Kombination von Sicherungsverfahren?** Hierbei spielen mehrere Faktoren eine Rolle. Zunächst bestimmt die Größe des zu sichernden Datenbereichs die Laufzeit einer Vollsicherung. Die Rate der geänderten Dateien bestimmt das gleiche für die Inkrementsicherung. Daraus sowie aus dem RPO lässt sich ein Schema erstellen und mit den Produktionsabläufen koordinieren, um Last zu verteilen.

Disk oder Tape?

Sind kurze Zugriffszeiten zu realisieren, so spricht das für eine Disk-Lösung. Stehen allerdings die Vorhaltezeit, also wie weit man zeitlich im Backup zurückgehen kann, im Vordergrund, so ist Tape eindeutig überlegen. Die Verwendung von Disk und Tape schafft eine zusätzliche Sicherheitsebene, da beide Technologien unterschiedliche Anfälligkeiten haben. Sowohl beim Preis pro TB als auch bei den laufenden Energiekosten, dem Raumbedarf und der Dauerhaftigkeit der Medien befindet sich Tape hier eine Klasse über der Disk. LTO-Tape hat sich als Standard in der IT etabliert und wird weltweit in großen Firmen, Banken und Versicherungen eingesetzt.

Neben seinen Kostenvorteilen sind besonders auch die Leistungsfähigkeit (Durchsatz LTO-6 = 140 MB/s, nativ), Kapazität (2,5 TB, nativ), seine Sicherheitsmechanismen (read-after-write, Error-Correction, Auto-Speed, Servo-Tracks) und seine Dauerhaftigkeit (30 Jahre Lesbarkeit) überzeugend. Tape skaliert zudem einfach. Durch das Einlegen leerer Tapes steht binnen weniger Minuten neuer Speicher zur Verfügung, ohne Konfigurationsänderungen und Umbau. Für große Datenmengen ist LTO-Tape derzeit ohne Alternative [9].

Disk und Tape

Disk2Disk2Tape heißt ein Sicherungsverfahren, bei dem man die jeweiligen Vorteile von Disk und Tape als optimale Kombination nutzt. Disk-to-Disk bietet eine schnelle Sicherung und schnelle Wiederherstellung von Dateien. Die Sicherung auf Tape geschieht vom Disk-Sicherungsvolume. Von dort aus kann Tape seine hohe Durchsatzrate nutzen und auf die gesamte Sicherung zugreifen. Durch seinen Kostenvorteil kann hier die Vorhaltezeit gegenüber der Disk wesentlich verlängert werden. Zusätzlich kann eine Auslagerung bereits geschriebener Tapes einen professionellen Sicherheitsgrad gewährleisten.

Eine weitere Kombination von Disk und Tape ist in der Produktion häufig von Vorteil und zwar bei der Verwendung von Media-Asset-Management-(MAM-)Systemen. Diese verwenden häufig SAN-Speicher. Bei ihnen ist die Verfügbarkeit besonders wichtig, da das MAM die zentrale Drehscheibe des gesamten Workflows darstellt. Die Kombination mit einem Tape-Archiv bietet sich ebenfalls an, um den zentralen Speicher zu entlasten und abgeschlossene oder selten verwendete Assets ins Archiv zu migrieren. Das kann auch automatisiert geschehen [10].

Besonders komfortabel ist es, wenn die Nutzer innerhalb des MAM-Systems den Archiv- und Restore-Vorgang auslösen können. Ein aktuelles Beispiel für eine solche Integration von MAM und Archiv bietet die Kölner moosystems.com. Deren „CP Archive App“ verbindet das Cantemo-Portal-MAM und P4-Archive von Archiware. Bei entsprechend konsequenter Verwendung amortisiert sich die Investition in ein Archiv allein durch die eingesparte SAN-Speichererweiterung. Zusätzlich kommen drastisch reduzierte Energie- und Kühlkosten hinzu. Über längere Zeit berechnet ergeben sich bedeutende Einsparpotentiale bei der Verwendung von Tape- gegenüber Disk-Systemen. [11]

Balance zwischen Backup und Archiv

- **Backup** = sekundäre Kopie von primären Daten
- **Archiv** = primäre Kopie von sekundären Daten (Curtis Preston)

Im Gegensatz zum Backup dient das Archiv

der zeitlich (meist) unbefristeten Sicherung von Daten. Es bildet die Gesamtheit aller abgeschlossenen Produktionen. Technologiebedingt kann eine Datei nicht aus dem Archiv bzw. auf dem Tape gelöscht werden. Nur das ganze Tape kann gelöscht werden. Ein Archiv wächst somit immer an, während ein Backup sich zyklisch erneuert und etwa eine konstante Größe behält. Da jedoch die Datenmenge auf den zu sichernden Systemen wächst, nimmt auch die Backup-Größe zu. Oft ist das unnötig, da viele Daten, die angelegt werden, nicht oder nur selten wieder abgefordert werden. Niemand mag das gern zugeben, aber das gilt tatsächlich in fast allen Umgebungen [12].

Hier ergibt sich ein beträchtliches Einsparungspotenzial, da diese Daten in ein Archiv ausgelagert werden können. Von dort können sie jederzeit bei Bedarf wieder zurückgespielt werden, kosten aber keinen wertvollen und aufwendigen Hauptspeicherplatz. Zudem verringert sich dadurch die Größe des Backups beträchtlich und das Archiv amortisiert sich dadurch schon nach wenigen Jahren.

Daten, die noch für die Produktion benutzt werden bzw. an denen weitere Änderungen wahrscheinlich sind, gehören auf den Produktionsspeicher und in dessen Backup integriert. Ist eine Produktion jedoch vollständig abgeschlossen, so sollte sie vom Produktionsspeicher ins Archiv verlagert werden. Dort wird sie mit Sicherungsmechanismen des Archivs wie Tape-Cloning abgesichert und kann anschließend vom Produktionsspeicher gelöscht werden.

Wird jedoch vorzeitig archiviert, also sind Dateien regelmäßig aus dem Archiv wieder herzustellen und nach Bearbeitung erneut zu archivieren, dann wächst das Archiv überproportional. Das Löschen von Dateien aus einem Tape-basierten Archiv ist technologiebedingt nicht möglich. Die Entscheidung für den richtigen Archivierungszeitpunkt ist daher von wesentlicher Bedeutung für einen möglichst ökonomischen Archivbetrieb.

Sonderfall Laptop-Sicherung

Bei der Erstellung und Bearbeitung von Inhalten kommen oftmals Laptops zum Einsatz. Die Einsatzorte wechseln und zusätzlich kommen freie Mitarbeiter ins Spiel, die Inhalte unterwegs erstellen. Da die Produktion von der Vollständigkeit der Inhalte abhängt, sollte die Sicherung der Inhalte auch mobiler Arbeitsplätze zentral organisiert werden. Datenverlust, auch von freien Mitarbeitern, kann die gesamte Produktion gefährden. Eine flexible, automatische und für Laptops angepasste Sicherung, die alle relevanten Daten zentral in der Firma sichert, ist notwendig, um hier Risiken zu minimieren. Der Administrator sollte mit dieser zusätzlichen Aufgabe möglichst wenig Arbeitsbelastung haben. Kleinere Arbeitsgruppen können flexible Regelungen (Open Policies) verwenden, um darüber zu entscheiden, was wann und wie oft gesichert wird. Große Firmen benötigen starre Regeln (Closed Policies), um ein konsistentes Sicherungsniveau zu erreichen. Eine Lösung wie „P4 Backup2Go“ (Archiware) kann offene wie geschlossene Sicherungsanforderungen über Templates abdecken und jederzeit neue Arbeitsplätze in Minuten integrieren. Auch hier erleichtert eine Nutzer-Restore-Funktion zur Wiederherstellung einzelner Dateien oder Verzeichnisse die Arbeit für den Administrator.

Besonders bei der Laptop-Sicherung kann es je nach Anforderung nötig sein, Geschäftsdaten zu verschlüsseln. So lassen sich die Geschäftsführungs- und Vertragsunterlagen ebenfalls in die Routinesicherung einbeziehen. Besonderes Augenmerk ist darauf zu legen, dass im Notfall der Schlüssel auch ausgedruckt– zum Beispiel im verschlossenen Kuvert im Tresor – hinterlegt wird. Die Verschlüsselung von Mediendaten wird in den allermeisten Fällen eher produktionsbehindernd sein und damit unterbleiben.

Herausforderung am Set

Bei der Produktion mit bandlosen Kameras entstehen schon am Set zahlreiche Sicherungsfragen. Hier ist mit einer mobilen Anlage für maximale Sicherung der soeben aufgenommenen Daten zu sorgen. Zunächst sollten Speicherkarten in mindestens zwei Kopien auf Festplatten überführt werden. Maximale Sicherheit ergibt sich durch die Verzögerung der Wiederverwendung der Karte um einen Tag. Eventuell auftretende Inkonsistenzen mit den Originaldaten können so abgeglichen werden.

Für diese verantwortungsvolle Aufgabe existiert bereits eine neue Berufsbezeichnung, der Data-Wrangler. Hier gibt es eigene Verfahren, zum Beispiel für die Markierung von Speicherkarten und deren Handling, die denen der Datensicherung ähneln. Die gesamte Produktion hängt von dieser Person ab, da durch ein Versehen die Arbeit des gesamten Teams vernichtet werden kann. Eingebürgert hat sich hier die Verwendung von Checksummen, die jedoch selbst wenig zur Sicherheit beitragen und außerdem überlegt eingesetzt werden sollten. Eine der Kopien sollte im Verlauf auf LTO-Tape übertragen und dieses möglichst sofort vom Set entfernt und eingelagert werden. In den USA wird die Einlagerung auf LTO-Tape teilweise ausdrücklich von der Set-Versicherung verlangt.

Für die gesamte Sicherung ist die Entscheidung, mit welcher Software gesichert wird, wichtig. Neben den Anschaffungs- und Wartungskosten ist auch auf die Einfachheit der Software zu achten. Früher oder später wird es nötig sein, dass mehrere, eventuell auch nur kurz eingewiesene Mitarbeiter die Sicherung zu überwachen haben. Mit komplexen Tools, selbst modifizierten Skripten und Open-Source-Lösungen entstehen hier unnötige Gefahrenquellen. Prinzipiell sollte Sicherungssoftware so ausgelegt sein, dass man auch nach mehreren Monaten, in denen keine Anpassungen nötig waren, sich schnell wieder im Interface zurechtfindet. Die Unterstützung aller relevanten Plattformen sowie Speichertechnologien vergrößert die Flexibilität. Auch die Skalierbarkeit, wie die Software mitwachsen kann, ist ein wichtiges Kriterium. Ein weiteres Kriterium ist die Möglichkeit, den Nutzern selbst einen Restore-Zugang zu ihrem Backup zu bieten.

Bei der Beschaffung und Konzeption von neuen Systemen und Workflows wird viel zu selten nach den Auswirkungen für die Datensicherung gefragt. Eine umfassende Planung bezieht die Datensicherung von Anfang an mit ein, genauso wie Wartungsintervalle, Verbrauchsmaterial und andere Folgekosten. Ein Automatismus, durch den jede neue Hardware auf ihre Backup-Notwendigkeit untersucht wird, ist bereits eine gute Katastrophenvorsorge.

Ausgehend von klassischen Informatikprinzipien ist Datenvermeidung der wichtigste Schritt, um eine Produktion überschaubar und damit auch gut sicherbar zu halten. Hier ist ein Blick auf die Produktionsformate angebracht und deren tatsächlichen Nutzen. Welcher Codec kann die nötige Qualität liefern – bei gleichzeitiger Reduktion der Datenmenge? Welche Auflösung ist tatsächlich nötig? Die Versuchung ist groß, mit einer 4K-Kamera auch 4K aufzunehmen, selbst wenn es dafür derzeit noch kaum Auslieferungsmöglichkeiten gibt.

Zusammenfassend lässt sich sagen, dass die größte Gefahr eine fehlende oder inadäquate Sicherung darstellt. Damit sind den Gefährdungen durch Nutzerfehler meist Tür und Tor geöffnet. Ein zunehmend zu beobachtendes Phänomen bei Entscheidern – bei der Beurteilung von Technologien – ist, dass aus der eigenen (privaten) Verbrauchererfahrung geurteilt wird, wie zum Beispiel der USB-Festplatte als private Sicherung.

Dabei werden professionelle Anforderungen und Notwendigkeiten ungenügend beachtet. Bei bestehenden Sicherungen entstehen die größten Probleme durch fehlendes Testen des tatsächlichen Restore-Ablaufs. Dabei kann der Unterschied zwischen dem geschätzten und tatsächlichen Zeitbedarf sogar mehrere Tage betragen. Eine genaue Schritt-für-Schritt-Dokumentation für den Notfall ist extrem wichtig. Gerade unter Druck kann nicht auf die Erfahrung und die richtige Einschätzung durch die Beteiligten gebaut werden.

Gefährliche Irrtümer

Zahlreiche Fehleinschätzungen verhindern adäquate Sicherungen oder verstellen den Blick auf Risiken.

„Wir sind gesichert – wir verwenden RAID." Eine verbreitete Fehleinschätzung, zu der RAID-Hersteller mit teils überzogenen Botschaften beigetragen haben. Zum einen sind 70 % der Dateiverluste auf Nutzer und deren versehentliches Löschen oder Umbenennen von Dateien, Verzeichnissen oder ganzer Volumen zurückzuführen, wovor RAID nicht schützen kann. Zum anderen nimmt die Ausfallswahrscheinlichkeit mit der Anzahl der Komponenten zu. Ein RAID besteht aus sehr kritischen Komponenten wie RAID-Controllern und einer größeren Anzahl von Festplatten. Das Zusammenspiel dieser Komponenten hat sehr enge Toleranzgrenzen. Ein RAID kann nur gegen den Ausfall einer oder, je nach RAID-Level, mehrerer Platten schützen. Kein Schutz besteht jedoch gegenüber Filesystem-Korruption, Löschen, Verschieben, Nutzerfehlern usw. Ohne Backup muss hier das ganze RAID-System im Notfall zum Datenrettungsservice gebracht werden. Die Häufigkeit der Anfragen zur Datenrettung hat mit RAID-6 drastisch zugenommen, da durch die vermeintliche Sicherheit kein Backup eingerichtet wurde [13].

Etwas Festplattenmathematik mag das RAID-Risiko verdeutlichen. Die Berechnung von „Mean Time to Data Loss" von RAID-5 und RAID-6 bezieht die Bitfehler-Wahrscheinlichkeit mit ein. Die Betrachtung ist nötig, um auch Datenverlust beim eventuellen Rebuild nach dem Ausfall einer Festplatte einzubeziehen. Dabei müssen alle Sektoren unbedingt fehlerfrei zur Verfügung stehen. Ein Rebuild-Vorgang kann bei einem großen RAID mehr als einen Tag (!) in Anspruch nehmen. Dabei ist die Leistung reduziert und die Ausfallgefahr erhöht und *nicht* durch RAID abgesichert. Bei einem RAID-5 mit 40 Festplatten beträgt dieser Wert nur ein (!) Jahr [14].

Resümee: RAID kann ein Backup nicht ersetzen, sondern benötigt selbst eine Sicherung.

„Tape ist veraltet." Diese Fehleinschätzung hat ihre Wurzel in den zahlreichen Vorläufertechnologien von LTO-Tape. Diese waren zum Teil langsam, unberechenbar und unsicher. LTO-Tape gehört jedoch zu den seltenen Glücksfällen der IT-Geschichte, in denen aus den Vorläufer-Technologien die besten Features kombiniert in eine neue Technologie mündeten. Zudem ist LTO in der derzeit sechsten Generation (LTO-6) am Start, mit einer Roadmap bis zur Generation 8 und stellt eine zuverlässige professionelle Langzeitsicherung dar. Gleichzeitig ist es das einzige Medium, das eine Auslagerung großer Datenmengen ermöglicht [15].

„Wir haben einen Snapshot." Das Snapshot-Feature virtueller Umgebungen wird dazu verwendet, frühere Zustände wieder herzustellen. Dieses Feature ist jedoch kein Backup. Es baut auf das fehlerlose Funktionieren der gesamten Virtualisierungsinfrastruktur auf und kann damit im Notfall eben genau nichts wiederherstellen.

„Wir sichern, wenn wir Zeit haben" oder *„Es läuft doch alles"* fallen in die Kategorie: Sicherung wird unterschätzt und als nicht zum Geschäfts- und Produktionsprozess gehörend betrachtet. Da die Produktions-, Kontakt-, Buchhaltungs- und E-Mail-Daten einer Firma heute den tatsächlichen Firmenkern repräsentieren, ist eine Sicherung dieser existentiell notwendigen Daten verpflichtender Bestandteil des Geschäftes selbst. Gesetzliche Verpflichtungen bestehen dazu bereits und machen Firmen haftbar für verlorene Geschäftsdaten. Eine Studie des BSI (Bundesamtes für Sicherheit in der Informationstechnik) zeigt die heterogene Sensibilisierung und Umsetzung von Sicherungsmaßnahmen in mittelständischen Firmen [16].

„Wir verwenden eine kostenlose/mitgelieferte Software." Oft handelt es sich dabei um Software für Einzelanwender. Eine professionelle Sicherung und Ansprüche waren und sind nicht das Ziel. Entsprechend zeigen sich die Einschränkungen: intransparente Abläufe, keine Angaben, was gesichert wurde, Löschen alter Sicherungen ohne vorherige Warnung, Fehlen von Reporting, fehlende Skalierbarkeit, beschränkt auf ein Betriebssystem, keine Übersicht zum Sicherungsstand und vieles andere mehr.

„Das Backup ist durchgelaufen." Oft wird übersehen, dass eine Sicherung nur dann wirklich verwendbar ist, wenn der Restore-Prozess regelmäßig getestet wird. Hierbei geht es, neben der Überprüfung der gesicherten Daten, hauptsächlich um die tatsächlich nötigen organisatorischen Abläufe. Wohin kann wiederhergestellt (restored) werden, welcher Speicher steht im Notfall zur Verfügung, wie kann ein Ersatzserver aufgesetzt werden, wer hat die nötigen Zugriffsrechte, wie lange dauert der Restore tatsächlich, was muss zusätzlich im Netzwerk oder in der Infrastruktur angepasst werden, wo sind die nötigen Schritte dokumentiert?

„Wir haben ein NAS." Heutige Netzwerke können wesentlich mehr Daten transportieren als noch vor zehn Jahren. Allerdings ist die Menge der zu transportierenden Daten auch erheblich angestiegen. In den meisten Fällen sogar wesentlich stärker als der Durchsatz im Netzwerk. Das zeigt seine Auswirkung vor allem dann, wenn große Datenmengen im Netzwerk gesichert werden sollen. Neben Einschränkungen der Verwendbarkeit für andere Zwecke (zur gleichen Zeit) ist die Dauer der Übertragung meist zu groß. Bei wachsender Datenmenge nimmt sie weiter zu.

Wichtige und große Speicherbereiche sollten daher mit ihrer eigenen dedizierten Verbindung zur Sicherung angebunden oder direkt zum Beispiel im FC-(Fibre-Channel-) Netz mit der Sicherung platziert werden. Viele NAS-Systeme sind zudem Black-Boxes mit einer unzureichenden Transparenz bezüglich des verwendeten Kernels, RAIDs und der Hardware. Im Schadensfall sind hier unerwartete Hindernisse zu überwinden oder sogar Daten unwiederherstellbar.

LTO-Tape mit LTFS

Mit der 5. Generation von LTO-Tapes (LTO-5) wurde LTFS eingeführt. Hierbei handelt es sich um die Möglichkeit, ein Tape partitioniert zu nutzen – mit einer Daten- und einer Metadatenpartition. Theoretisch kann damit ein Tape ähnlich einer CD-RW/DVD-RW genutzt werden, beim Einlegen erscheint der Inhalt des Volumens und die Dateien können gelesen oder geschrieben werden. Der in diesem Zusammenhang von den Tape-Herstellern oft gezogene Vergleich zur Festplatte ist jedoch extrem realitätsfern.

Als Sicherungsmedium eignet sich LTFS – als Format – nicht. Es kann nicht alle Dateiattribute und Dateiarten sichern. Daher verwenden alle bekannten Backup-Software-Hersteller eigene Formate, die zudem für die Sicherung optimiert sind.

Ausblick

Die Zukunftsaussichten sind recht eindeutig, was das Datenwachstum angeht. Eine weitere Zunahme der Datenmenge ergibt sich durch

größere Produktionsformate sowie mehr Plattformen und führt dazu, dass fast jede Produktion heute Größenordnungen zu verwalten hat, die noch vor wenigen Jahren nur in Rechenzentren vorkamen. Daraus ergibt sich auch eine Diskrepanz zwischen Kompetenz und Erfahrung sowie den Anforderungen im Datenmanagement und Backup. Datenverluste werden vermehrt auftreten bis sich eine halbwegs professionelle Sicherungskultur etabliert hat.

Die Verwendung von SSD-Speicher wird rapide zunehmen und von der Industrie forciert werden. Hierbei sind die Ausfallsrisiken noch unklar. Disks werden noch billiger werden, wahrscheinlich bei weiter sinkender Qualität und steigender Anfälligkeit (außer in der teuersten Kategorie). Das LTO-Tape wird das professionelle Archivierungs- und Langzeitsicherungsmedium bleiben [17].

Da die Netzbandbreiten in viel geringerem Maße wachsen werden als die Menge der Daten, wird eine Sicherung im und über das Internet weiterhin für große Datenmengen inadäquat sein.

Eine professionelle lokale Sicherung bleibt auf absehbare Zeit die beste Lösung in der Medienproduktion.

Schrifttum

[1] www.bisimplified.com/_pdf/_newsletters/10_09_articles/article_4.pdf
[2] http://en.wikipedia.org/wiki/Disaster_recovery
[3] www.infoworld.com/d/storage/wp/long-term-data-protection-and-retention-finding-the-correct-balance-570
www.archiware.de/disk-to-disk-to-tape-d2d2t.33.0.html
[4] http://de.wikipedia.org/wiki/Asymmetric_Digital_Subscriber_Line
[5] www.wingfs.com
[6] www.bigbrotherawards.de/2012/.comm
www.zdnet.com/blog/igeneration/microsoft-we-can-hand-over-office-365-data-without-your-permission/11041
www.heise.de/newsticker/meldung/Auch-Google-uebermittelt-europaeische-Daten-an-US-Behoerden-1319347.html
[7] http://files.dubfire.net/csoghoian-dissertation-final-8-1-2012.pdf
[8] www.techrepublic.com/blog/five-apps/five-tips-for-backing-up-your-data-to-the-cloud/697?tag=content;siu-container
[9] www.spectralogic.com/index.cfm?fuseaction=members.docContactInfoForm&DocID=4255
[10] www.andre-aulich.de/perm/mit-press-tore-archive-automatisch-daten-auf-tape-verschieben
[11] www.clipper.com/research/TCG2010054.pdf
[12] www.ssrc.ucsc.edu/Papers/leung-usenix08.pdf
[13] www.searchstorage.de/themenbereiche/storage-security/datenrettung-forensik/articles/310761/
[14] www.heinlein-support.de/upload/slac08/Heinlein-RAID_Mathematik_fuer_Admins.pdf
Christof Windeck: Sinnvolle Redundanz – So setzt man RAID heute ein. c´t, 2/2012, S. 136.
Detaillieret Infos zu Ausfallswahrscheinlichkeiten http://www.heise.de/ct/12/02/links/136.shtml
[15] www.lto.org/technology/roadmap.html
[16] www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Studien/KMU/Studie_IT-Sicherheit_KMU.pdf?__blob=publicationFile
Der sehr umfangreiche Grundschutzkatalog enthält gut strukturierte Hinweise zu allen relevanten Themen.
www.bsi.bund.de/DE/Themen/ITGrundschutz/StartseiteITGrundschutz/startseiteitgrundschutz_node.html
[17] www.theregister.co.uk/2011/05/06/soirage_trifecta/