Speichertechnologien im Überblick

Im Delta des Datenflusses

Der Storage-Markt bietet eine breite Palette teils komplementärer, teils alternativer Technologien, um der stetig wachsenden Nachfrage nach Speicherplatz gerecht zu werden. Keine passt überall. Jens-Christoph Brendel

Foto: NASA

Was Tsai Lun, Minister am Hof des chinesischen Herrschers Ho Ti, antrieb, war eine Art Storage-Problem. Deshalb erfand er im zweiten Jahrhundert das Papier. Die Kapazität der alten Tontafeln genügte nicht mehr und die Schreibzugriffe waren nicht performant. Papier blieb 600 Jahre Geschäftsgeheimnis und verdrängte danach hierzulande im Spätmittelalter das Pergament. Dank des neuen Mediums ließen sich die Speicherkosten pro Seite entscheidend senken. Genauso – erwartetet man zum Anbruch des Informationszeitalters – würde die Harddisk den Hefter aus dem Feld schlagen.

Zwar ist das papierlose Büro bislang ein Mythos geblieben, doch von allen neu gewonnenen Informationen gelangen gegenwärtig über 90 Prozent auf magnetische Träger und nur Promille auf Papier [2]. Die Zielrichtung blieb über die Jahrhunderte dieselbe: Immer mehr flüchtige Informationen galt es, immer billiger, sicherer und schneller festzuhalten. Der Bedarf trieb die technische Entwicklung voran.

Eins aber hat sich in den letzten Jahrzehnten dramatisch geändert: Der Zuwachs an Information beschleunigte sich rasant. Schätzungen für den Anstieg reichen aktuell von 30 bis 75 Prozent im Jahr [1]. Dabei mag manches Szenario über den explodierenden Speicherbedarf von interessierter Seite überzeichnet sein, doch die Steigung der Wachstumskurve ist nicht zu übersehen.

Der Pegelstand der Daten-Polder

In der Folge ist an der Speicherfront ein Wettrüsten entbrannt. Eine nur ein Vierteljahrhundert alte Technik ist uns heute fast so fremd wie das China zur Zeit des Papiererfinders: 1980 bevölkerten Winchester-Plattenlaufwerke mit dem Outfit einer Waschmaschine, dem Preis einer Eigentumswohnung und der Kapazität eines daumengroßen Memory-Sticks unserer Tage die Rechenzentren. Heute beansprucht allein E-Mail – inzwischen die wichtigste Kommunikationsplattform in Unternehmen – mit weltweit über 30 Milliarden Sendungen pro Tag mehr als 400 000 TByte im Jahr [2].

Die Informationsflut zwingt nicht nur Großunternehmen, sondern auch den Mittelstand dazu, die Deiche zu erhöhen. Dabei ist „Viel hilft viel“ allerdings die falsche Devise. Gerade in Zeiten knapper Kassen kommt es eher darauf an, aus dem breiten Spektrum verfügbarer Technologien die für bestimmte Zwecke geeignetste auszuwählen.

Wegmarke

Eine erste Orientierung vermittelt die Speicherpyramide (Abbildung 1). An ihrer Spitze residiert der schnellste und teuerste Speicher: Register und Caches im direkten Zugriff der CPU. Zum Glück wird er nur in vergleichsweise homöopathischer Dosis benötigt, 1 GByte davon würde mehrere hunderttausend Dollar kosten. Seine Gegenstücke sind die Archivspeicher am Fuß der Pyramide, je GByte ein tausendstel Dollar. Da hier die Medien aber offline gelagert und erst bei Bedarf in die Reichweite eines Rechners gebracht werden, können ihre Zugriffszeiten im Bereich von Stunden oder Tagen liegen.

Am interessantesten ist das Mittelstück: Hier findet sich zum einen die Grauzone zwischen On- und Offline: Nearline. In ihr sind Speicher angesiedelt, die automatisiert innerhalb einiger Sekunden online geschaltet werden können. Ne-

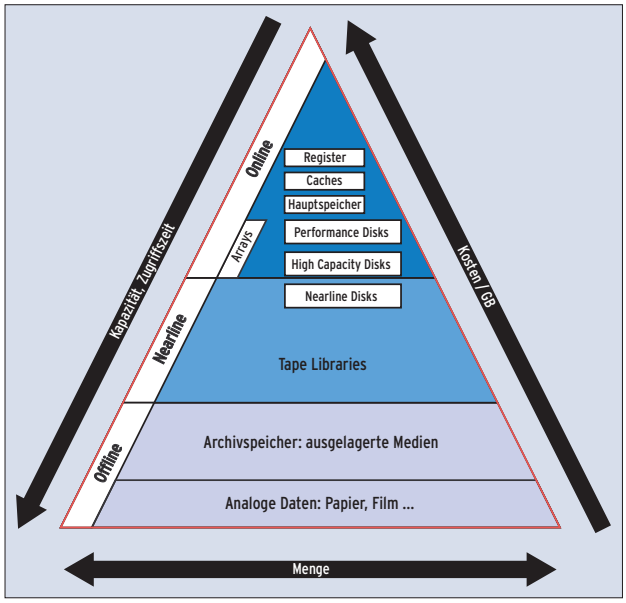

Abbildung 1: Die Speicherpyramide als Wegweiser in der Storage-Landschaft: Wichtige Kriterien für den Einsatz verschiedener Speichertechnologien sind hier übersichtlich zusammengefasst.

ben optischen Speichermedien ist dieser Sektor nach wie vor eine Domäne der Magnetbänder.

Am Fuß der Datenberge

Im Laufe ihrer über 50-jährigen Geschichte konnten die Bandspeicher ihr Volumen ständig vervielfachen. Auf einem der ersten Bänder, dem IBM Model 726, fanden anno 1952 gerade mal 1,4 MByte Platz, das Äquivalent einer Diskette. Derzeit streiten im Highend-Bereich mehrere Formate um die Vorherrschaft: SDLT 600 (Super Digital Linear Tape) mit einer Kapazität von 300 GByte pro Band, das LTO-3 (Linear Tape Open, Kapazität 400 GByte) und SAIT-1 (Super Advanced Intelligent Tape) mit 500 GByte Speicherplatz (jeweils unkomprimiert).

Im Entwicklungslabor konnten bereits Bänder hergestellt werden, die 1 TByte Daten aufnehmen – und auch dies wird bestimmt noch nicht das Ende der Fahnenstange sein, die SAIT-Roadmap zum Beispiel plant für das Jahr 2010 Version SAIT-4 mit 4 TByte [3].

Der Durchsatz eines solchen SDLT- oder LTO-Laufwerks schwankt je nach Art, Größe und Komprimierbarkeit der Daten, dürfte aber gegenwärtig im Durchschnitt zwischen 30 und 35 MByte pro Sekunde liegen. Damit sprengen Backup oder Recovery einer großen Datenmenge die normalerweise verfügbaren Zeitfenster (1 TByte benötigt so mindestens acht Stunden). Deshalb können große Tape-Libraries mit dutzenden parallel arbeitenden Laufwerken bestückt werden und hunderte oder tausende Bandkassetten aufnehmen (Abbildung 2).

Bei etwa 65000 Euro beginnt der Einstieg in die Enterprise-Klasse (ab 1000 Slots). Die Kosten pro GByte einer solchen Konfiguration erreichen dennoch – abhängig von der Anzahl der Bänder pro Laufwerk – nur ein Viertel bis ein Zwanzigstel dessen, was für den gleichen Speicherplatz auf Festplatten aufgewendet werden müsste [4]. Zudem lassen sich Bänder getrennt lagern und sind dadurch eher gegen Viren und Würmer gefeit als Platten. Deshalb werden Bandroboter wohl auch in absehbarer Zukunft das Rückgrat der Backup-Systeme bilden.

Billig versichert

Am unteren Preis-Ende rangieren Autoloader mit Magazinen ab sieben Kassetten und einem Laufwerk. Je nach Typ der verwendeten Bänder sind solche Wechsler bereits unter 3000 Euro zu haben (Abbildung 3), die 20-Slot-Klasse mit SDLT- oder LTO-Laufwerken beginnt erst jenseits von 6000 Euro. Häufig sind die Bänder mit Barcode-Labels versehen, die von der Library gelesen und an die Backup-Software weitergegeben werden. So ergibt sich eine durchaus auch für kleine Firmen lohnende Backup-Lösung, die weitgehend automatisiert arbeitet und einigen TByte Daten als Lebensversicherung dienen kann.

Alte aufs Abstellgleis

Neben den Bändern erobern sich mit fallenden Preisen auch Platten einen Platz im Nearline-Segment. Hier geht es um Disk-to-Disk-Backups oder Pufferspeicher für große Libraries, die sonst oft durch überlastete Netze zu zeitraubendem Stop-and-go-Betrieb gezwungen werden könnten. Es geht um ein altes

Abbildung 2: Eine der größten Tape-Libraries ist die Powderhorn 9310 von Storagetek. Auf bis zu 144000 Bandkassetten bringt sie unglaubliche 28,8 Petabytes unter. (Foto: Storagetek)

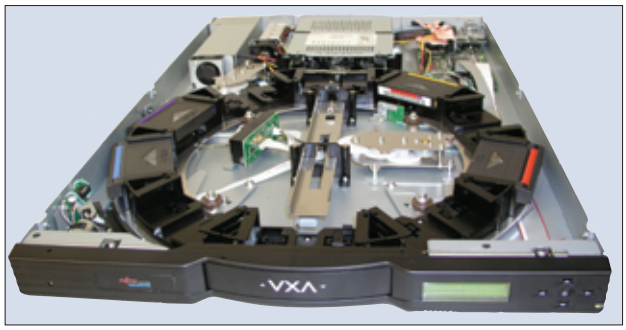

Abbildung 3: Blick ins Innenleben eines neuen Bandwechslers von Exabyte, der unter anderem von Fujitsu-Siemens gehandelt wird. Das nur eine Höheneinheit beanspruchende Gerät fasst unkomprimiert 800 GByte. (Foto: Fujitsu-Siemens)

Konzept, das gerade eine Wiedergeburt erlebt. Die Idee besteht darin, Informationen, die mit der Zeit an Bedeutung verlieren und auf die deshalb wesentlicher seltener zugegriffen wird, in der Speicherpyramide absinken zu lassen, bis sie schließlich auf billigen, langsamen Medien landen (oder gelöscht werden) und damit schnellen, teuren Online-Speicher wieder freigeben.

Ein bekanntes Produkt nach diesem Strickmuster ist Suns SAM-FS [5], zu dessen prominenten Anwendern unter anderem das Deutsche Zentrum für Luft- und Raumfahrt (DLR) gehört, das mit dieser Filesystem-basierten automatischen Archivierung seit vielen Jahren Satellitendaten speichert.

Der Daten Wanderleben

Das dahinter stehende Konzept heißt HSM (Hierarchical Storage Management), allerdings werden dabei hauptsächlich das Alter der Daten und die Zugriffshäufigkeit als Kriterien herangezogen. Die Reinkarnation nennt sich ILM (Information Lifecycle Management) und setzt sich das ehrgeizige Ziel, auch inhaltliche Aspekte zu berücksichtigen. So könnte beispielsweise eine Mail mit der Einladung zum Mittagessen und den Quartalszahlen im Anhang in einen Teil gesplittet werden, der am nächsten Tag im Papierkorb landet, und einen, der noch einige Monate für den schnellen Zugriff bereitgehalten und danach automatisch in ein Langzeitarchiv verschoben wird.

Verschärfte gesetzliche Bestimmungen für die Aufbewahrung digitaler Daten (Basel II) forcieren die Entwicklung zusätzlich. Heute gibt es bereits Komponenten, aber bis zu einer funktionierenden ILM-Gesamtlösung ist es wohl noch ein weiter Weg. Die anstehenden Probleme sind nicht nur technischer Natur. Die Software für die automatische inhaltliche Klassifizierung ist ebenso eine Herausforderung wie die Konzeption des nötigen Regelwerks oder die Durchsetzung einer unternehmensweiten Policy, die vorschreibt, welche Daten welchen Speicher in welcher Lebensphase belegen sollen.

Trotzdem rühren alle Branchengrößen im Storage-Markt bereits kräftig die Werbetrommel. EMC [6] beispielsweise hat sich mit dem Backup-Spezialisten Legato und dem Content-Management-Experten Documentum kürzlich zwei Firmen einverleibt, die den Weg zum ILM-Provider ebnen sollen. Aber auch Hitachi Data Systems (HDS), IBM, Storagetek, HP, Sun, Veritas und viele andere basteln an ILM-Komponenten.

Die Etage über Nearline in der Speicherpyramide ist für die Massenspeicher im Online-Zugriff reserviert. Hier hausen die Harddisks, die wie die Bänder ihr Fassungsvermögen über viele Jahre kontinuierlich mit hohen Wachstumsraten vergrößern konnten. Heute hat man fast 100 GBit pro Quadratzoll (6,4516 cm^2) erreicht. Das Spektrum reicht von eingebauten oder direkt an Server angeschlossenen Platten, JBODs (Just a bunch of disks) und Raid-Systemen – Sammelbegriff: DAS (Direct Attached Storage) – über Speicher mit Netzwerk-Anbindung (NAS, Network Attached Storage) bis zu eigenständigen Speichernetzen (SAN, Storage Area Network).

Allein, vereint, vernetzt

Auch hier ist die Kapazität nur eins unter vielen, sich teils auch noch widersprechender Auswahlkriterien. So geht es einerseits, wie überall, um die Kosten, andererseits um Verfügbarkeit, die zusätzliche Ausgaben für redundante Komponenten verursacht. Zudem ist die Skalierbarkeit einer Speicherlösung ausschlaggebend dafür, dass dem wachsenden Bedarf so effizient wie möglich begegnet werden kann. Schließlich muss das System einfach zu handhaben und zuverlässig zu überwachen sein.

In den Server eingebaute Platten sind wohl der einfachste und billigste Lö-

Fakten rund um Fibre Channel

Bei den Diskarrays der Oberklasse und im SAN hat sich Fibre Channel als Transportmedium durchgesetzt. Diese Technologie überwindet viele Grenzen der Vorgängerverfahren zur Speicheranbindung: Die Beschränkung auf 15 Geräte und einen Controller wie bei SCSI entfällt – eine Arbitrated Loop (siehe unten) koppelt bis zu 126 Knoten, eine Fabric theoretisch sogar bis zu 16 Millionen.

Diese Systeme können bis zu 10 Kilometer voneinander entfernt sein, bei SCSI sind dagegen maximal 25 Meter erlaubt (Differential SCSI). Die Transfergeschwindigkeit erreicht heute bis zu 2 GBit/s, wogegen es der schnellste SCSI-Ableger, Ultra 640 SCSI, nur auf etwas mehr als ein Viertel davon bringt. Daneben ist die serielle Übertragung im Fibre Channel wesentlich weniger störanfällig als bei SCSI, auf dessen parallelen Datenleitungen es mit zunehmender Kabellänge schnell zu Laufzeitdifferenzen kommt.

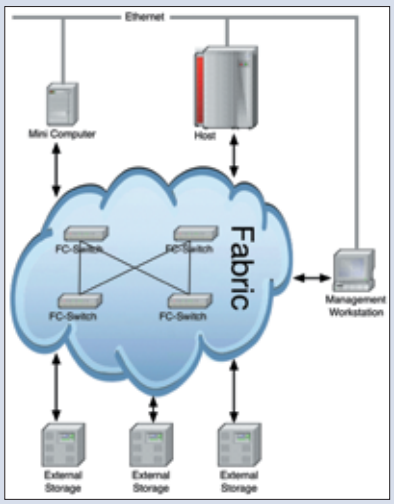

Abbildung 4: So funktioniert ein SAN: Server und Speicher verbinden sich über Switches in einem Hochgeschwindigkeitsnetz.

Fibre Channel kennt zwei Verbindungstypen: eine ring- oder sternförmige, kostengünstig realisierbare Bus-Topologie mit FC-Hubs (Fibre Channel Arbitrated Loop, FC-AL) sowie Fabric (FC-SW, Fibre Channel Switched Fabrik), das sind Speichernetze mit FC-Switches. Anstelle der Switches finden oft auch so genannte Directors Verwendung, die bei ähnlicher Funktionalität mehr Ports und erhöhte Ausfallsicherheit bieten.

Auch mit Blick auf die Medien ist Fibre Channel flexibel: Für kurze Distanzen reichen Kupferkabel, für längere müssen es jedoch die wesentlich teureren Multimode-, für sehr große Entfernungen sogar Single-Mode-Glasfaserkabel sein.

Ähnlich der MAC-Adresse im klassischen Ethernet verfügt jedes Fibre-Channel-Device über eine eindeutige Adresse, die WWN (World Wide Number). Zusätzlich haben auch die Ports – jedes Gerät kann über mehrere verfügen – eine globale Anschrift, die WWPN (World Wide Port Number). Dabei ist Fibre Channel selbst ein reiner generischer Übertragungsmechanismus, der andere Protokolle für den eigentlichen Zugriff auf den Speicher transportiert. In den meisten Fällen ist das SCSI, es nimmt aber auch Protokolle aus der Mainframewelt wie ESCON, ja sogar ATM und TCP/IP Huckepack.

sungsansatz und dabei vergleichsweise schnell. Doppelt ausgelegte Systemplatten gehören für einen Server ohnehin zum guten Ton, oft bieten Maschinen im Entry-Level- oder Midrange-Bereich aber noch weitere Steckplätze für zusätzliche Disks, die der Admin häufig sogar im Betrieb tauschen kann (Hot Swap).
Konfiguration und Monitoring gelingen mit Bordmitteln, die Infrastuktur für den Betrieb ist mit dem Server bereits bezahlt. Auf diese Weise lässt sich einfach eine Konfiguration mit einigen hundert GByte Volumen realisieren, die allerdings nur begrenzt ausbaubar ist und auch keinen besonderen Sicherheitsanforderungen genügt.

Klassiker

Einen Schritt weiter gehen direkt angeschlossene Diskarrays und Raids, die Klassiker unter den Storage-Systemen. Hier übernimmt immer spezialisierte Hardware das Disk-Management, Konfiguration und Monitoring werden meist von eigener Software unterstützt. In der Oberklasse bieten solche Programme auch spezielle Funktionen wie Snapshots, Replikationen, Array-to-Array-Kopien oder Serverless-Backup. Alle kritischen Hardwarekomponenten (etwa Netzteile) sind redundant ausgelegt. In der Regel lassen sich weitere Server anschließen, sodass Cluster oder Standby-Konfigurationen realisierbar sind, bei denen die Daten auch nach einem Servercrash verfügbar bleiben.
Häufig bieten die Hersteller auch so genanntes I/O-Multipathing an. Dabei werden parallele Datenpfade konfiguriert, zwischen denen beim Ausfall eines Controllers automatisch gewechselt wird. Solche Systeme sind modular aufgebaut und in anderen Größenordnungen erweiterbar als eine Lösung mit internen Platten. Die größten Arrays erreichen Kapazitäten über 150 TByte, zum Beispiel HP Storageworks, Hitachi Lightning, EMC Symmetrix, siehe **Abbildung 5**. Für kleinere Arrays gibt es ein breites Produktspektrum in allen Preis- und Größenklassen.
Der nächste Systemtyp löst die direkte Verbindung von Server und Array: NAS und SAN verfahren so. Der Speicher wird von der Rechenleistung entkoppelt, lässt sich vorteilhaft zentralisieren und steht damit prinzipiell allen Benutzern netzwerkweit zur Verfügung. Fast zwangsläufig ergibt sich so auch für die Datensicherung ein zentralisierter und damit günstigerer Ansatz.

Völlig losgelöst

Der Hauptunterschied zwischen NAS und SAN besteht darin, dass der Netzwerkspeicher bei NAS Filesysteme mit Hilfe eingeführter Sharing-Protokolle (NFS, CIFS) exportiert, während das Speichernetz (SAN) den Zugriff auf Devices gewährt, von deren Inhalt oder von dessen Organisation es keine Ahnung hat.
Eine NAS-Appliance ist dabei nichts anderes als ein Fileserver mit für diesen Einsatzzweck optimierter Anwendungs- und Systemsoftware. Das erhöht die Performance und vereinfacht Installation und Administration im Vergleich zur Server-Raid-Lösung. Auf der Habenseite verbucht der netzgebundene Speicher außerdem geringere Komplexität und weit niedrigere Kosten als für ein SAN zu veranschlagen sind. Ethernet, das Übertragungsmedium, ist ohnehin vorhanden, die Protokolle sind standardisiert und allen Clients verständlich, die meist Web-basierten Administrationstools überall einsetzbar.
An ihre Grenzen stößt diese Herangehensweise zum einen dann, wenn maximale Performance gefragt ist: Wegen des Protokoll-Overhead beim Filesharing wird die Bandbreite wesentlich schlechter ausgenutzt als bei einem SAN, ein überlastetes LAN bremst die Zugriffe zusätzlich. Außerdem braucht Ethernet für die Übertragung sehr viel mehr Interrupts als eine Übertragung via Fibre Channel. Zum anderen ist es mit der Skalierbarkeit vorbei, wenn die Möglichkeiten der jeweiligen NAS-Box ausgereizt sind, ein Netz dagegen lässt sich einfach um neue Teilnehmer erweitern.

Oberliga

Was Performance, Skalierbarkeit und die Möglichkeit des Speichermanagements angeht, bilden SANs heute die Königsklasse der Speichersysteme. Deren direkte Vernetzung via Fibre Channel

(siehe Kasten „Fakten rund um Fibre Channel“) ermöglicht nicht nur besonders kurze Zugriffszeiten, sondern bietet ebenso die Chance, das viel Bandbreite beanspruchende Backup unter eine zentrale Regie zu stellen und aus dem LAN zu verbannen (Serverless-Backup). Das benötigte Zeitfenster lässt sich damit oft um mehr als die Hälfte verkürzen. Auch Replikation oder Spiegelung von Diskarrays ist ohne Umweg durch den Flaschenhals Ethernet möglich. Redundante Datenpfade sorgen für hohe Ausfallsicherheit.

Profitabler Pool

Ein großer Speicherpool erlaubt eine wesentlich bessere Ressourcenauslastung als Insellösungen und ist zudem einfacher zu managen und kostengünstiger zu erweitern. Ein SAN kann Speicherkapazität dynamisch verwalten und bei Bedarf zuteilen. Das meint das Schlagwort Speichervirtualisierung.

Bei genauerem Hinsehen erweist sich dieses Feature allerdings als schlichte Notwendigkeit: Bereits in einem etwas größeren SAN mit redundanten Verbindungen könnte jeder Server so viele Devices erkennen, dass eine Administration auf der physischen Ebene unmöglich oder mindestens sehr fehlerträchtig wäre. Eine logische Schicht, die die Hardware in virtuellen Komponenten zusammenfasst, verbirgt einen Teil dieser Komplexität und macht eine größere Installation erst handhabbar.

Abbildung 5: Ein Speichersystem der Highend-Klasse ist das EMC Symmetrix DMX-2. (Foto: EMC)

Den zahlreichen Vorteilen stehen aber deutlich höhere Anschaffungskosten gegenüber. Für Installation und Administration ist spezielles Know-how erforderlich. Probleme gibt es ebenfalls immer noch mit der Interoperabilität: Geräte verschiedener Hersteller kommen nicht in jedem Fall miteinander klar. Auch für das Management eines SAN gibt es noch keine allgemeinverbindlichen Standards.

Bemühungen darum sind aber bereits im Gang. Unter Schirmherrschaft der SNIA (Storage Networking Industry Association, [7]), einer 1997 gegründeten Organisation führender Hersteller wie EMC, Cisco, Dell, Sun, Veritas oder Storagetek, wird die Entwicklung eines einheitlichen Managementkonzepts unter dem Namen Bluefin vorangetrieben. Sein Kern ist eine objektorientierte Beschreibungssprache (CIM, Common Information Model), mit der die Abhängigkeiten und Eigenheiten jedes Geräts erfasst, in einem XML-Dialekt (XMLCIM) kodiert und an eine zentrale Broker-Instance (CIMON, CIM Object Broker) weitergeleitet werden, die diese Informationen in einem Repository verwaltet.

Die Broker-Instanz, bei der jedes ins Netz eingebaute Geräte anzumelden ist, bildet die Schnittstelle zu Management-Applikationen. Auf die Geräte greift sie aber nicht direkt zu, sondern bedient sich so genannter Object Provider, wodurch ältere Hardware leichter eingebunden werden kann, für die lediglich der Hersteller einen passenden Vermittler programmieren muss [8].

Alternativen

Außerdem gibt es noch eine Reihe weiterer Versuche, um Speicher netzwerkfähig zu machen. Die spezielle Linux-Lösung ENBD stellt ein anderer Artikel vor. Eine weitere Variante ist ISCSI, bei dem SCSI in IP-Pakete verpackt wird. Auch hier gibt es erste Ansätze unter Linux. Die beteiligte Industrie wirbt mit Kostenvorteilen gegenüber SAN, hat jedoch zurzeit noch Performance-Probleme: Das Verkapseln und Entpacken der SCSI-Kommandos kostet viel Zeit und Rechenleistung. Abhilfe sollen spezielle Ethernet-Adapter bringen, die mit Hilfe einer TOE (Transport Offload Engine) diese Aufgabe in eigene Hände nehmen und die CPU entlasten. Entsprechende ISCSI-Adapter sind auf dem Markt. Bis jetzt hat die Technik jedoch keine mit NAS oder SAN vergleichbare Verbreitung gefunden.

Im Soho-Markt gibt es ähnliche Ansätze. Hier trifft man auf Zwitterwesen aus WLAN-Router und Festplattenanschluss beziehungsweise Kreuzungen aus Ethernet-Switch und Harddisk, zum Beispiel Ximeta Netdisk [9].

Fazit

Eine Informationsflut droht – doch niemand ist dazu verurteilt, in ihr zu ertrinken. Dem Küstenschutz steht ein gut sortiertes Arsenal technischer Mittel zur Verfügung. Die Auswahl muss sich an den Anforderungen orientieren, die mit Blick auf entscheidende Kriterien wie Verfügbarkeit, Ausbaufähigkeit, Performance, Kapazität und natürlich Kosten so exakt wie möglich zu definieren sind. In diesem Licht erweist sich die gespiegelte Disk im gemieteten Rootserver vielleicht als ebenso optimale Lösung wie die NAS-Appliance als Ersatz für den Fileserver der Abteilung oder das SAN im Konzern-Rechenzentrum. ■

Infos

[1] Speicherwachstum: [http://speicherguide.de/magazin/aktuelles.aspβtheID=804]
[2] Berkely-Report: [http://www.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm]
[3] SAIT: [http://www.aittape.com/sait-tape-roadmap.html]
[4] Speicherkosten von Platte und Band: [http://www.horison.com/horison/topics/2004/06/]
[5] SAM-FS: [http://www.sun.com/products-n-solutions/hardware/docs/Software/Storage_Software/Sun_SAM-FS_and_Sun_SAM-QFS_Software/index.html]
[6] EMC: [http://germany.emc.com/index.jsp]
[7] SNIA: [http://www.snia.org/
[8] Bluefin: [http://www.snia.org/tech_activities/SMI/bluefin/]
[9] Net-Disk: [http://www.ximeta.de]