

EBOOK

GESPONSERT VON:



DATACORE

EFFIZIENTERE NUTZUNG VON SPEICHERSYSTEMEN

Trevor Pott

INHALT DIESES E-BOOKS:

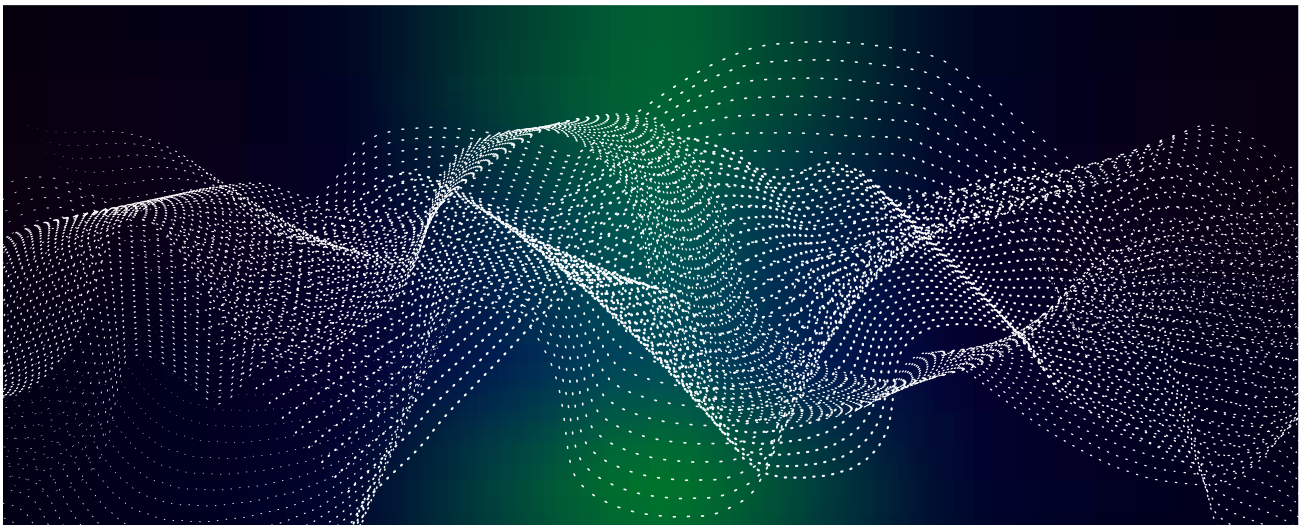
- 2** Data Fabrics und ihre Rolle bei der Aktualisierung von Speichersystemen
- 6** Die Zunahme hyperkonvergenter Infrastrukturen
- 9** EINE STRAGIE ZUR EFFIZIENTEREN NUTZUNG IHRES SPEICHERS

Wie Sie die 3 größten Herausforderungen bei der
Datenverwaltung mit **softwaredefiniertem Speicher** meistern



DATA FABRICS UND IHRE ROLLE BEI DER AKTUALISIERUNG VON SPEICHERSYSTEMEN

Data Fabrics ermöglichen Organisationen, einen Schritt weiter zu gehen und Daten über verschiedene Speichersysteme zu verteilen, sodass einheitliche Speichersysteme überflüssig werden.



Wie aktualisieren Sie Ihren Speicher? Oder genauer gefragt, was machen Sie mit Speichersystemen, deren Supportverträge auslaufen, die aber eigentlich noch reibungslos funktionieren? Empfiehlt es sich, Supportverträge zu verlängern und womöglich deutliche Preissteigerungen in Kauf zu nehmen – um Neukauf zu vermeiden – oder können Speichersysteme der Vorgängergeneration vielleicht auch ohne Support nutzbringend eingesetzt werden?

Je nachdem, an wen Sie diese Fragen richten, werden Sie unterschiedliche Antworten erhalten. Mitarbeiter von IT-Herstellern bieten für gewöhnlich recht nüchterne Antworten und begründen meist ausführlich, warum jegliche IT grundsätzlich niemals ohne gültigen Supportvertrag genutzt werden sollten. Außerdem scheint es sehr gute

Gründe dafür zu geben (warum eigentlich?), dass die Preise für Supportverträge für Speichersystemen nach dem zweiten oder dritten Jahr exorbitant ansteigen.

Offenbar geschieht es ausschließlich zum Wohle des Kunden, dass diese alle drei Jahre recht vehement dazu gedrängt werden, ihre Speicherinfrastruktur zu aktualisieren – auch wenn diese an sich noch einwandfrei funktioniert und ihren Zweck erfüllt.

In Unternehmen, die über ein entsprechendes Budget verfügen, mag dieser Ansatz unter Umständen sogar realistisch sein. In allen anderen stößt die Praxis, funktionierende Geräte wegzuerwerfen, die sich eigentlich noch in tadellosem Zustand befinden, und dann sechs- oder sogar siebenstelligen Beträge in ihren Austausch zu investieren, auf Unmut und Unverständnis.

Anbieterbindung bei Laufwerken

Jedes Speichersystem umfasst grundsätzlich zwei Komponenten: Eine davon ist der Speicher selbst – also Festplattenlaufwerke (rotierende Speichermedien) oder SSD-Laufwerke (mit Flash-Technologie). Die zweite ist die Steuereinheit, der sogenannte Controller, der regelt, welche Daten auf welchem Laufwerk gespeichert werden.

Die Laufwerke eines Speichersystems der Enterprise-Klasse unterscheiden sich dabei meist nur geringfügig von den Laufwerken eines günstigen Speichersystems (z.B. NAS) für Endverbraucher. Auf Unternehmensgeräte wird zwar für gewöhnlich eine Anbieter-spezifische Firmware aufgespielt, doch inwiefern sich mit dieser tatsächlich eine Leistungssteigerung oder höhere Zuverlässigkeit der Laufwerke erzielen lässt, konnte auch nach Jahrzehnten immer noch nicht eindeutig nachgewiesen werden.

Fest steht dagegen, dass Anbieter von Speichersystemen mithilfe dieser Firmware verhindern können, dass Unternehmen die Laufwerke ihrer Systeme/Arrays mit Standardlaufwerken ersetzen. Viele Speichersysteme, besonders Produkte der gehobenen Preisklasse, können Laufwerke ohne die Firmware ihres Herstellers nicht erkennen. Auch nicht, wenn es sich um hochwertige Laufwerke speziell für Rechenzentren handelt.

Wenn ein privater Nutzer Probleme mit einem Laufwerk feststellt oder zusätzlichen Speicherplatz benötigt, kann er ganz einfach und schnell über Amazon Ersatz bzw. zusätzliche Kapazitäten bestellen. Bei Speicherlösungen für Endverbraucher verdienen die Hersteller an den Controllern, nicht an den Laufwerken. Stattdessen werden hier Standard-Laufwerke ohne spezifische Firmware verwendet. Da stellt sich doch die Frage, warum dies nicht auch bei Enterprise-Systemen möglich ist?

Das Prinzip hinter der Speicheraktualisierung

Die Gründe dafür, dass Anbieter von Enterprise-Speichersystemen den regelmäßigen Ersatz der Arrays forcieren, sind vielfältig und komplex. Zunächst einmal müssen Speicher-Arrays der Enterprise-Klasse ausfallsicher sein. Denn während private Nutzer und kleine Unternehmen gewisse Fehlerraten und Ausfallzeiten tolerieren können, ist dies in großen Unternehmen meistens nicht möglich. Das führt unter anderem dazu, dass Anbieter sehr darauf achten,

dass nur Laufwerke verwendet werden, die umfassend auf Kompatibilität mit der Hard- und Software des eingesetzten Controllers getestet wurden.

Außerdem gilt zu bedenken, dass Laufwerke für den privaten Gebrauch nicht einheitlich sind. Ein Beispiel hierzu: Die Gold-Festplatten von Western Digital werden vom Hersteller als Laufwerke der Enterprise-Klasse verkauft. Und private Nutzer, die auf der Suche nach einer erstklassigen Festplatte für ihren Server oder ihr NAS-System sind, können mit den WD Gold-Festplatten mit Sicherheit nichts falsch machen. Doch während alle Gold-Festplatten zwar derselben Produktreihe angehören, sind sie nicht automatisch auch identisch.

Das gilt sogar für Festplatten mit derselben Kapazität. Selbst bei den Gold-Festplatten mit 12 TB bestehen Unterschiede zwischen den einzelnen Produktionschargen und bei jeder sind andere Besonderheiten zu beachten. Zusätzlich entwickelt Western Digital das Design und die Firmware seiner Laufwerke kontinuierlich weiter. Das Resultat: Hinter dem Namen Western Digital Gold-Festplatte mit 12 TB können sich unter Umständen dutzende verschiedene Modelle verbergen, die alle denselben Namen tragen.

Wenn dies für Endverbraucher kein Problem darzustellen scheint, warum ist es nicht auch für Unternehmenslösungen möglich? Der Grund dafür liegt bei den Anbietern von Unternehmenssystemen, die den mit diesem Ansatz verbundenen Aufwand umgehen möchten. Und das zu Recht: Die Ausfallraten können extrem hoch sein. Vielleicht erinnern Sie sich noch an drei der bekanntesten Vorfälle mit hohen Ausfallraten? Da wären (in umgekehrter chronologischer Reihenfolge) die Probleme mit dem [Seagate ST3000DM001](#) (auch bekannt unter „Failcuda“), das OCZ-[Debakel](#) und die Probleme mit dem [IBM Deskstar 75GXP](#) (häufig auch „Deathstar“ genannt). Das OCZ-Debakel ist erwähnenswert, da es zeigt, dass es auch bei SSDs zu Ausfällen kommen kann, nicht nur bei rotierenden Festplatten.

Es gibt vermutlich noch weitere Modelle, die in dieser Liste nicht fehlen dürften, doch an Daten zu Ausfallraten zu gelangen ist nicht ganz einfach. Die meisten Anbieter (mit Ausnahme von Backblaze) hüten diese Daten äußerst streng und geben sie für gewöhnlich erst nach richtiger Anordnung preis.

Es ist also nahezu unmöglich, herauszufinden, welche Laufwerkmodelle tatsächlich fehleranfällig sein könnten. Da ist es verständlich, dass kein Anbieter riskieren möchte,

für Probleme mit dem Zugriff auf geschäftskritische Daten verantwortlich gemacht zu werden, die eigentlich daher rühren, dass fehleranfällige Laufwerke in Speichersystemen eingebaut wurden.

Ein weiterer Grund sind Bedenken bezüglich der Controller. Hinsichtlich der Hardware sind die meisten Controller für Enterprise-Systeme vergleichbar mit denen eines Standard x86-Servers. Folglich könnten die meisten problemlos mindestens zehn Jahre lang eingesetzt werden. Wenn an einen Controller, der auf Festplatten mit 1 TB bis 6 TB ausgelegt ist, nun aber mehrere 12-TB-Festplatten angeschlossen werden, kann dies seine Leistung durchaus beeinträchtigen.

Speichersysteme verfügen über RAM, der neben dem Caching auch für die Ausführung von Controller-Funktionen verwendet wird. Dazu zählen unter anderem die Deduplizierung, Komprimierung und Verschlüsselung. Bei Arrays der Enterprise-Klasse kann es sich bei diesem RAM zum Teil, oder auch vollständig, um nicht-flüchtigen RAM handeln (NVDIMMs), der sicherstellt, dass in den Cache geschriebene Daten auch bei einem Stromausfall nicht verloren gehen.

Die Installation größerer Laufwerke, als sie für das Array vorgesehen sind, könnte bedeuten, dass der Controller nicht mehr über genügend RAM verfügt. In Bezug auf Controller mit NVDIMMs empfehlen Anbieter die dreijährliche Aktualisierung unter anderem auch deshalb, weil die Lebensdauer der beiden Komponenten, die diese Ausfallsicherheit gewährleisten (Batterien und Flash-Speicher), sehr viel kürzer ist als die der anderen Komponenten des Controllers.

Data Fabrics

Aus Sicht der Kunden ist die Anbieterbindung bei Festplatten verständlicherweise ärgerlich. Sie empfinden diese Vorgehensweise als strategischen Schachzug, mit dem Anbieter sie dazu zwingen, mehr zu investieren als eigentlich notwendig wäre. Und damit haben sie auch nicht ganz unrecht. Irgendwann um 2009 wurden zahlreiche Startups gegründet, die neue Ansätze entwickelten, um aus der allgemeinen Unzufriedenheit mit dem Status quo Kapital zu schlagen.

Doch die etablierten Anbieter hatten damals durchaus gute Gründe, die Hardwarebindung und regelmäßige Speicheraktualisierungen zu forcieren – zumindest von einem bestimmten Standpunkt aus.

Speicheranbieter haben Ihre Systeme in der Annahme vertrieben, dass sie als einzige Speicherlösung für eine bestimmte Art von Workloads eingesetzt werden. Auch wenn davon ausgegangen wird, dass Kunden Backups durchführen, gilt dennoch, dass Speichersysteme höchste Ausfallsicherheit gewährleisten müssen. Sogar Menschenleben können davon abhängen. Dieser Ansatz stammt noch aus Zeiten vor der Entwicklung von Data Fabrics und wurde Generationen von Technikern, Storage-Administratoren und Speicherexperten eingebläut.

Mit Data Fabrics hat sich das jedoch geändert. Speichersysteme verfügen heutzutage über redundante Festplatten und sogar redundante Controller für den Fall, dass eine Komponente ausfallen sollte. Data Fabrics gehen einen Schritt weiter, indem sie die Verteilung von Daten auf mehrere Speichersysteme ermöglichen, sodass ganze Systeme redundant werden.

Die meisten Data Fabrics bieten jedoch weit mehr Funktionen als nur die Daten zu mehreren Systemen zu schreiben um die Daten beim Ausfall eines ganzen Systems zu schützen. So prüfen sie zum Beispiel fortlaufend die Leistung aller eingebundenen Speichersysteme und schreiben Daten dann – auf Basis der Speicherprofile der Anwendungen – automatisch in das am besten geeignete Array.

Erbringt ein Array nicht mehr die volle Leistung (weil es beispielsweise mit zu vielen Festplatten mit zu hoher Kapazität bestückt wurde), wird es automatisch nur noch für Daten mit geringen Performanceanforderungen (auch Cold Data genannt) verwendet. Das heißt, dass dort Datenblöcke gespeichert werden, auf die aller Wahrscheinlichkeit nach nicht häufig zugegriffen werden muss oder die zu Workloads gehören, die über ihr Speicherprofil als latenzunempfindlich eingestuft wurden.

„Frankenstorage“

Ein großer Vorteil von Data Fabrics ist, dass Sie quasi jegliches Speichersystem einbinden können. Wer möchte, kann sich also aus den [Systemen von Supermicro](#) eine White-Box-Speicherinfrastruktur zusammenstellen und zum Beispiel bis zu [90 3,5-Zoll-Festplattenlaufwerke](#) in ein 4-HE-Rack oder auch [48 NVMe-Festplatten](#) in einen 2-HE-Server packen.

Während die Systeme von Supermicro wohl das extremste Beispiel für White-Box-Lösungen darstellen, bieten die meisten Serverhersteller mittlerweile [Open-Compute-Lösungen an](#)

– und der neue Trend in diesem Bereich lautet Open-Compute-Storage. Wer nach einer möglichst günstigen White-Box-Lösung sucht, für den könnte der [Backblaze Storage Pod](#) das Richtige sein.

Das Grundprinzip hinter dem Data-Fabric-Modell ist, dass herkömmliche rotierende Festplattenlaufwerke als Massenspeicher eingesetzt werden und dieser dann durch leistungsstarke SSDs (und zwar immer häufiger NVMe-SSDs) ergänzt wird. Ist mehr Kapazität erforderlich, werden weitere Festplattenlaufwerke hinzugefügt; reicht die Leistung nicht mehr aus, werden zusätzliche SSDs bereitgestellt. Die Verteilung der Daten auf die einzelnen Systeme erfolgt in der Fabric automatisch.

Data Fabrics müssen aber nicht auf White-Box-Speicher basieren. Die meisten Fabrics ermöglichen die Einbindung beliebiger Speichersysteme. So können Sie hier auch vermeintlich veraltete Speichersysteme recyceln, deren Supportvertrag ausgelaufen ist. Fügen Sie sie zur Fabric hinzu, bestellen Sie bei Bedarf Ersatzfestplatten auf eBay und nutzen Sie das Array bis zum letzten Tag. Sogar Cloud-Speicher können Sie hinzufügen, um zusätzliche Kapazitäten bereitzustellen.

Und wenn Ihr veraltetes Array eines Tages tatsächlich ausfällt? Kein Problem. Denn die darauf gespeicherten Daten wurden bereits repliziert und an anderer Stelle in der Fabric gesichert. Sie müssen also keinerlei Ausfallzeiten oder Datenverlust befürchten. Vielmehr werden die

meisten Fabrics den Ausfall des Speichersystems erkennen und zusätzliche Sicherungskopien der betroffenen Daten erstellen.

Storage-Administratoren der alten Schule verspotten Data Fabrics häufig als „Frankenstorage“. Ein Ausdruck, den Anbieter von Fabrics nicht gerne hören, da sie befürchten, dass er das Modell in ein schlechtes Licht rückt. Dennoch ist der Begriff zutreffend. Data Fabrics geben Unternehmen die Möglichkeit, verschiedene Speicherlösungen von unterschiedlichen Anbietern flexibel zu einem großen Ganzen zusammenzufügen, das wesentlich mehr ist als die Summe seiner Teile. Genau wie Franksteins Monster werden auch Data Fabrics zutiefst missverstanden und sind Opfer einer FUD-Strategie.

Die Skepsis überrascht nicht, letztendlich stößt jede neue Innovation im Bereich der IT zunächst auf Widerstand. Doch Data Fabrics sind mittlerweile mehr als ein Forschungsprojekt. Auf dem Markt sind bereits verschiedene bewährte Lösungen unterschiedlicher Anbieter verfügbar und nach und nach zeigen sich auch erste Auswirkungen darauf, wie Unternehmen die Aktualisierung ihrer Speicherinfrastrukturen angehen. Wenn in Ihrem Unternehmen in nächster Zeit umfassende Hardware-Upgrades anstehen, lohnt es sich vielleicht, diese Technologie in Erwägung zu ziehen.

DIE ZUNAHME HYPERKONVERGENTER INFRASTRUKTUREN

Was das für Ihre Speicherstrategie bedeutet.

Viele bekannte Anbieter von hyperkonvergenten Infrastrukturen (HCI) gründeten ihre Unternehmen im Jahr 2009. Diese Technologie besteht also bereits seit knapp zehn Jahren. HCI ist bekanntlich eine Infrastruktur, die Speicher und Rechenleistung in einer einzigen Lösung vereint. Doch was genau bedeutet das? Wie passen sie zu existierenden, teilweise sehr komplexen, IT-Umgebungen? Haben hyperkonvergente Infrastrukturen ihr Versprechen gehalten?

Um das HCI-Modell zu verstehen, sollte man die Entwicklung der Speichersysteme in den letzten 10 Jahren kennen, zumindest die wichtigsten Schlagwörter. Als Hyperkonvergenz im Jahr 2009 aufkam, war eins davon in aller Munde: Software-Defined Storage, kurz SDS.

Software-Defined Storage ist der perfekte Begriff, der ähnlich nichtssagend ist wie „eine vom Menschen entwickelte Struktur“. Während der letztgenannte Begriff immerhin noch auf den Menschen als Entwickler eingrenzt, ist er immer noch so vage, dass er im Prinzip völlig bedeutungslos ist.

Heutzutage basiert nahezu jede Speicherlösung auf Software-Defined Storage. Selbst hinter einem Großteil der Daten in Papierform steht SDS. So ist es heute sogar möglich, Dokumente in Papierform mithilfe von Barcodes, QR-Codes oder NFC zu organisieren und ihre Verwaltung zu automatisieren.

Der Begriff SDS diente ursprünglich als Bezeichnung für Software, die Speicheranbieter auf die gleiche Weise 'standardisiert', wie VMware Server-Anbieter 'standardisiert' hat. Daraus wurde dann im Laufe der Zeit die inhaltslose Definition: „Jede Speicherlösung, die sich von einem herkömmlichen Speichersystem unterscheidet.“

Ein modernes Array unterscheidet sich mittlerweile grundlegend von den

Speichersystemen anno 2009. Da es aber immer noch ein Speichersystem ist, wird es eher selten als SDS bezeichnet. Eine hyperkonvergente Infrastruktur, die nur mit wenigen Knoten desselben Anbieters kompatibel ist, wird dagegen häufig ohne Weiteres als SDS betrachtet.

In anderen Worten: SDS ist ein Schlagwort, das Startups gezielt gewählt haben, um Verbrauchern durch die Blume zu kommunizieren, dass Anbieter wie EMC, NetApp, IBM und HP etc. zu teuer seien und es viel klüger sei, stattdessen ihren Speicher zu wählen. Mehr Inhalt sollte dieser Begriff tatsächlich nie haben.

Data Fabrics

Das Konzept, das ursprünglich hinter dem Begriff SDS stand, also die Speicherstandardisierung, ist auch heute noch relevant. Eine Zeit lang wurde auch der Begriff „Speichervirtualisierung“ verwendet. Damit sollte einerseits ein Bezug zur Servervirtualisierung seitens VMware hergestellt werden. Andererseits sollte er vermitteln, dass Daten genauso problemlos zwischen verschiedenen



GiroScience / Shutterstock.com

Speicherlösungen verschoben werden kann wie virtuelle Maschinen zwischen verschiedenen Hosts eines Clusters.

Die Bezeichnung setzte sich aber nie durch, weil VMware selbst bessere Speichervirtualisierung anbot als die meisten Startups. So bietet vSphere beispielsweise mit Storage vMotion die Möglichkeit, Daten zwischen verschiedenen Speicherlösungen zu verschieben – sofern die Anwendungen in VMware virtuellen Maschinen (VMs) laufen. Zusätzlich können VMwares Virtual Volumes (VVOLs) die Verwaltung von LUNs vereinfachen. Eine Diskussion der VVOL Vor- und Nachteile würde den Rahmen dieses Dokuments sprengen.

Das ursprüngliche Konzept von SDS wurde nun in Form der Data Fabric erneut aufgegriffen. Eine Data Fabric ist eine verteilte Anwendung, die alle mit ihr verknüpften Speichersysteme zu einem einzigen Speicherpool zusammenfasst und auf Basis vordefinierter Parameter neu strukturiert und verwaltet.

In einer offenen Data Fabric kann dabei jede Art von Speicher genutzt werden – ob Cloud-Speicher, White-Box-Systeme, lokaler Serverspeicher, SANs, NAS-Systeme usw.. Dieser 'zusammengewürfelte' Speicher kann dann wiederum in verschiedenster Form, unterschiedlich je nach Data Fabric Anbieter, zugänglich gemacht werden, d. h. über iSCSI, Fibre-Channel, SMB, NFS, als Object Storage oder sogar als lokal an einen Knoten angeschlossenen Speicher. Eine offene Data-Fabric Infrastruktur ermöglicht es also, jede beliebige Art von Speicher einzubinden und den Speicher dann auf beliebigem Wege zur Verfügung zu stellen.

Während die meisten Data Fabric so flexibel sein könnten wie oben beschrieben, sieht das in der Realität oft anders aus. Nichts hindert einen Anbieter von Data Fabric daran, eine beliebige Speicherquelle zu integrieren oder Speicher in einem von ihm gewünschten Format auszugeben. Denn die größten Herausforderungen bei der Entwicklung von Data Fabric Software beziehen sich auf die Funktionen zur Verteilung der Datenblöcke sowie die Bereitstellung von hochwertigen Speicherdiensten, wie Datenspiegelung, Komprimierung, Deduplizierung, Snapshots, schnelles Klonen, Verschlüsselung usw.. Im Vergleich dazu ist es recht einfach, einen neuen Speichertyp oder ein neues Speicherformat zu unterstützen.

Anbieter von Data Fabric schränken die Funktionen ihrer Fabric jedoch aus denselben Gründen ein, aus denen Anbieter herkömmlicher Speichersysteme die Anbieterbindung bei Festplatten und die regelmäßigen,

dreijährlichen Aktualisierungen forcieren. Womit wir wieder bei hyperkonvergenten Infrastrukturen wären.

Der HCI-Plus-Ansatz

Es spielt durchaus eine Rolle, für welchen Anbieter Sie sich bei Ihrer Data Fabric Software entscheiden. Denn davon hängt ab, wie flexibel die Fabric sein wird. Und hier kommen hyperkonvergente Infrastrukturen erneut ins Spiel: Bei HCI handelt es sich im Prinzip um eine Data Fabric aus mehreren Servern, die der Fabric ihren lokalen Speicher zur Verfügung stellen und überflüssige Rechenkapazitäten für parallel laufende Anwendungen nutzen.

Obwohl es keinen erkennbaren Grund gibt weitere Speicherquellen in die Fabric zu integrieren, ist dies bei den meisten nicht möglich. Stattdessen ist die Auswahl von Server, die zu einem Cluster hinzugefügt werden können, sehr eingeschränkt, genau wie ihre Anzahl. Hinzu kommt, dass die Server selbst oft nur geringfügig an individuelle Anforderungen angepasst werden können.

Data Fabric mit mehr Flexibilität ermöglichen dagegen einen „HCI-Plus-Ansatz“. Das heißt, dass Kunden Server mit hohen Speicherkapazitäten nutzen können, um Cluster zu bilden, die mit einer HCI vergleichbar sind. Außerdem können sie ihre älteren Speichersysteme oder Cloud-Storage hinzufügen und für seltener benötigte Daten, Snapshots und als Archiv etc. verwenden.

Data Fabric in der Praxis

Die Frage, warum ein HCI-Plus-Ansatz für Data Fabric überhaupt in Erwägung gezogen wird, ist durchaus berechtigt. Denn was zunächst überzeugend klingt, wirkt realistisch betrachtet schnell die Frage auf, warum man nicht einfach beim herkömmlichen HCI-Ansatz bleibt und die Fabric vollständig aus White-Box-Servern bildet. Vielleicht ist es noch sinnvoll, Cloud-Storage als externen Speicherort für Snapshots oder selten benötigte Daten hinzuzufügen, aber ist alles andere wirklich notwendig?

Die Antwort auf diese Frage ist komplex. Für komplett neue IT-Umgebungen bietet der HCI-Plus-Ansatz tatsächlich wenig Vorteile. Herkömmliche HCI Umgebungen erfüllen diese Aufgabe, und selbst grundlegende Cloud-Replikations- oder Backup-Funktionen zeigen eher eine geringe Akzeptanz bei Kunden.

Allerdings ist der Ansatz auf grüner Wiese eher sehr selten der Fall. Die meisten Unternehmen verfügen über

bunt zusammengewürfelte unterschiedliche Infrastruktur-Komponenten und -Arten, oft anders für die jeweiligen Fachbereiche aufgebaut und somit auch mit nicht synchronen Aktualisierungszyklen. Hinzu kommt, dass Unternehmenszusammenschlüsse und Übernahmen häufig dazu führen, dass ein Unternehmen mehrere Rechenzentren betreibt, die verschieden aufgebaut sind und verwaltet werden.

Viele Unternehmen sind noch nicht bereit, sich bei geschäftskritischen Anwendungen auf eine hyperkonvergente Lösung – geschweige denn Data Fabrics – einzulassen. Aus ihrer Sicht reichen 10 Jahre nicht aus, um die Zuverlässigkeit der Technologie zu beweisen. Stattdessen ziehen sie ihren bisherigen kostspieligeren, aber bekannten Ansatz vor und werden diesen wohl auch noch über Jahre hinweg verfolgen.

Doch selbst in konservativen Unternehmen wird ständig nach Möglichkeiten zur Steigerung der Effizienz gesucht. So sind Administratoren bei nicht geschäftskritischen Anwendungen oder in Bereichen wie Archivierung und Entwicklung durchaus bereit, „neuen“ Technologien eine Chance zu geben. Hier finden Data Fabrics – oder genauer der HCI-Plus-Ansatz – tatsächlich auch den größten Anklang.

Das Problem mit klassischer HCI

Herkömmliche HCI-Lösungen bieten leider nicht besonders viel Flexibilität bei der Zusammenstellung eines Clusters. So ist das Verhältnis zwischen Speicherkapazität, der Anzahl von Prozessorkernen und RAM beispielsweise relativ klar vorgegeben. Hinzu kommt, dass die Anzahl der SSDs (und NVMe-SSDs) pro Server und damit auch die Speicherleistung der Knoten eingeschränkt ist.

Solange Sie beim Kauf des Clusters bereits genau wissen, wofür er eingesetzt werden soll, ist das wahrscheinlich kein Problem. In diesem Fall können Sie die Kapazitäten und Leistung des Clusters perfekt auf Ihre Anforderungen abstimmen. In der Realität ist es aber häufig so, dass sich Prioritäten und Nutzung im Laufe der Zeit ändern und ein Cluster im Laufe seines Lebenszyklus ganz unterschiedliche Aufgaben bewältigt.

Viele Administratoren lösen dieses Problem, indem sie deutlich mehr Kapazitäten und Leistung bereitstellen als tatsächlich erforderlich. Diese Vorgehensweise ist zwar gängig, steht aber dem eigentlichen Zweck der Einführung einer HCI entgegen. Denn der große Vorteil von HCI ist, dass Kunden ihre Cluster bei Bedarf gezielt erweitern und damit Kosten sparen können.

Anders als bei offenen Data Fabrics können Sie bei herkömmlicher HCI nicht einfach weiteren Speicher oder zusätzliche NVMe-Server für mehr Leistung hinzufügen. Bei einer HCI-Plus-Infrastruktur sieht das jedoch anders aus.

Hier profitieren Sie nicht nur von mehr Flexibilität bei der Zusammenstellung Ihrer IT Infrastruktur. Sie können außerdem auch sämtliche Speichersysteme recyceln, die Sie nicht mehr für geschäftskritische Anwendungen einsetzen möchten, und für nichtkritische Daten nutzen.

Buzzword-Bingo

Buzzwort oder nicht, Bezeichnungen sind wichtig. Nur wenige Anbieter, die sich als HCI-Anbieter ausgeben, verfolgen einen HCI-plus-Ansatz. Und diejenigen, die bereits flexiblere Data Fabrics anbieten, haben noch keine einheitliche Bezeichnung dafür gefunden. Das liegt unter anderem daran, dass die Spanne zwischen klassischen hyperkonvergenten Infrastrukturen und offenen Data Fabrics sehr groß ist.

Ohne eine einheitliche Bezeichnung war es bisher eher schwierig, Kunden die verschiedenen Ansätze der Anbieter nahezubringen und aufzuzeigen, wo die Unterschiede liegen. Diese Informationen benötigen Kunden aber, um die am besten geeignete Lösung für ihre Anforderungen zu finden. Das ist sicherlich einer der Gründe, warum der Speichermarkt heute derart widersprüchlich ist.

Beachten sollte man allerdings, dass HCI-Anbieter, die ihren Ansatz als neu und innovativ verkaufen, in Relation zu herkömmlichen Speichersystemen, dabei von einer Technologie sprechen, die fast 10 Jahre alt ist. Sie ist also alles andere als neu. Außerdem zeigt sich immer wieder, dass auch HCI-Anbieter ihre Kunden gerne von sich abhängig machen, Preise sehr hoch ansetzen und unnötige Einschränkungen festlegen. Das heißt, sie verfolgen genau den Ansatz, zu dem sie ursprünglich eine Alternative bieten wollten.

Offene Data Fabrics bringen uns nun endlich die Vorteile, die wir uns einst von der Hyperkonvergenz erhofft hatten. Bleibt nur noch abzuwarten, ob HCI Anbieter diesen Schritt mitgehen und zumindest den HCI-Plus-Ansatz annehmen werden. Vielleicht werden sie auch nach und nach einer neuen Generation von Anbietern weichen, die offene Data Fabrics etablieren, ihre eigenen Schlagwörter schaffen und den IT-Markt in eine neue Richtung lenken werden.

EINE STRATEGIE ZUR EFFIZIENTEREN NUTZUNG IHRES SPEICHERS

Wie Sie die 3 größten Herausforderungen bei der Datenverwaltung mit softwaredefinierten Speichern meistern

Es gibt große Herausforderungen in der IT. Die wesentlichen sind: unterbrechungsfreie Dienste zu bieten, den ständig wachsenden Kapazitätsanforderungen gerecht zu werden, hoch performante-IT bereitzustellen und dabei noch das Budget im Griff zu haben.

Das Kernproblem: DATEN

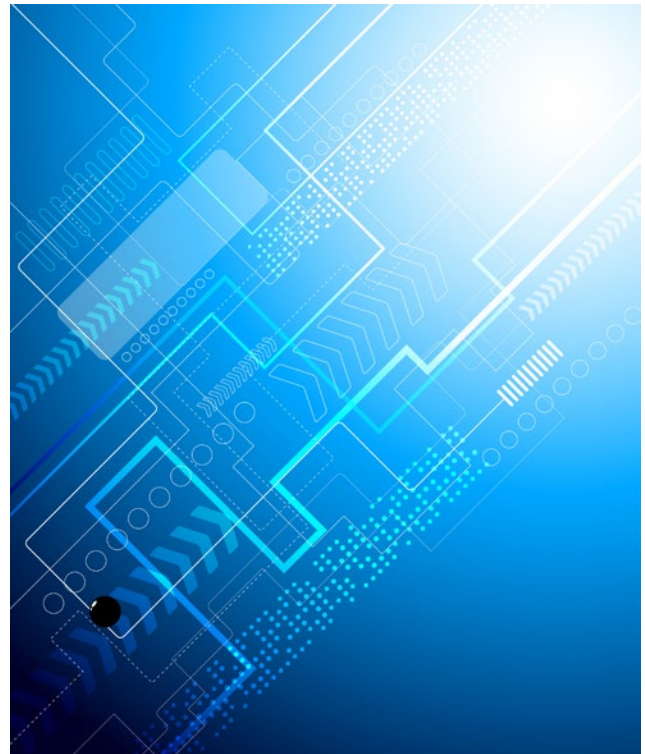
In Unternehmen dreht sich alles um Daten – sei es in Bezug auf die Einhaltung neuer Bestimmungen, die Bereitstellung neuer Anwendungen oder deren immensen Wachstums. Geschäftskritische Daten müssen permanent verfügbar und schnell abrufbar sein. Ist dies nicht der Fall, kann das weitreichende negative Konsequenzen für das jeweilige Unternehmen haben.

Zusätzlich verkomplizieren gesetzliche Bestimmungen die Verwaltung von Daten. Daraus folgt, dass Speichersysteme heute eine nie dagewesene Skalierbarkeit bieten müssen. Doch damit nicht genug, man muss auch mit knappen IT-Budgets zurechtkommen.

Es ist mehr als eine große Herausforderung, eine traditionelle Speicherlösung mit unternehmenstauglicher Datenverfügbarkeit, Skalierbarkeit und Performance bereitzustellen. Aber so ist noch nicht für die erforderliche Flexibilität gesorgt, um jederzeit möglichen Änderungen gerecht zu werden. Eine solche Flexibilität einzuführen bedeutet erheblichen Aufwand, sowohl finanziell als auch zeitlich, geschweige denn von dem späteren Betrieb.

Die Lösung: Software-Defined Storage

Zuverlässige Software definierte Speicherlösungen der Enterprise-Klasse schaffen Abhilfe. Sie bietet Unternehmen unterbrechungsfreien Datenzugriff, umfassende



Skalierbarkeit und beispiellose Anwendungsleistung. Die Software erzeugt eine transparente Virtualisierungsebene über verschiedene Speichersysteme hinweg, um die Verfügbarkeit, Skalierbarkeit und Leistung sämtlicher Speicherressourcen zu maximieren.

Sehen wir uns an, wie die richtige SDS Lösung alle Herausforderungen löst, die Speicherverwaltung in der gesamten Organisation vereinfacht und Ihnen den Überblick und die Kontrolle über Ihren Speicher zurückgibt.

Herausforderung 1: Datenverfügbarkeit

Daten verfügbar zu halten bedeutet, dass der Zugriff auf Daten stets gesichert sein muss – auch bei schwerwiegenden Ausfällen innerhalb der Speicherarchitektur aufgrund von Hardwarefehlern, Standortausfällen, regionalen Ausnahmesituationen oder Benutzerfehlern. Für genau diese kontinuierliche Verfügbarkeit kann SDS mit den folgenden Features: synchrone Spiegelung, permanente Datenprotokollierung, mehrpfadige-Kommunikationskanäle sowie Replikation an Remote-Standorten zur Notfallwiederherstellung sorgen. Außerdem führt eine gute SDS Lösung Sie durch die Einrichtung Ihrer Speicherinfrastruktur mit allen Prozessen, um zu gewährleisten, dass Ihre Lösung vollständig ist.

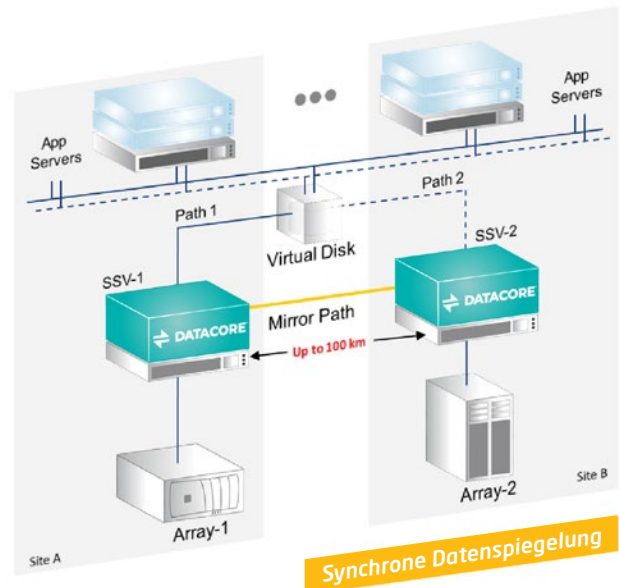
Synchrone Datenspiegelung

Der Schutz geschäftskritischer Daten zählt zu den wichtigsten Prioritäten jedes Unternehmens. Bevor wir uns ansehen, wie dies umgesetzt werden kann, betrachten wir den Unterschied zwischen Fehlertoleranz und Hochverfügbarkeit.

Fehlertoleranz ist die Eigenschaft eines Systems, auch bei Ausfällen einzelner Hardwarekomponenten eines Speichersystems, wie Festplattenlaufwerken oder Netzgeräten, weiterhin reibungslos zu funktionieren. Jedes IT-System sollte eine angemessene Fehlertoleranz aufweisen, um vor unvermeidlichen Hardwareausfällen geschützt zu sein. Doch Fehlertoleranz alleine ist nicht ausreichend, denn sie berücksichtigt die wichtigste Komponente nicht: Ihre DATEN.

Die passende SDS-Lösung bietet Schutz vor Datenverlust aufgrund von Ausfällen von Speichersystemen durch äußere Einflüsse (z. B. Stromausfälle, Brände, Ausfälle der Raumklimatisierung usw.). Hier würden Fehlertoleranzmaßnahmen alleine zu kurz greifen. Hohe Verfügbarkeit bedeutet Redundanz auf Datenebene zusätzlich zur Redundanz auf Geräteebe. So wird die Datenverfügbarkeit maximiert, und Ihre Anwendungen können unterbrechungsfrei ausgeführt werden.

Gute SDS-Lösungen gewährleisten hohe Verfügbarkeit durch synchrone Spiegelung, indem jeder Datenblock auf einen weiteren aktiven Knoten gespiegelt wird. Auf diesem Knoten werden redundante Kopien beibehalten. So ist gewährleistet, dass die Daten und Anwendungen selbst bei Ausfällen größerer Bereiche der Speicherinfrastruktur weiterhin verfügbar sind.



Bei der Auswahl einer hochverfügbaren Lösung sollten die folgenden vier Gesichtspunkte berücksichtigt werden:

- › **Umfassende Redundanz:** Die zugrunde liegenden Speichersysteme benötigen eine angemessene Fehlertoleranz (Redundanz auf Hardwareebene), und die Daten müssen hochverfügbar sein (Redundanz auf Datenebene).
- › **Autonomie der Systeme:** Die zugrunde liegenden Speichersysteme müssen unabhängig voneinander funktionsfähig sein und dürfen einander nicht berücksichtigen.
- › **Trennung der Systeme:** Die zugrunde liegenden Systeme können bis zu 100 km entfernt voneinander platziert werden, um zu vermeiden, dass Ereignisse an einem Standort die Bereitstellung von Speicherservices am anderen Standort beeinträchtigen.
- › **Asymmetrie der Systeme:** Die zugrunde liegenden Speichersysteme variieren hinsichtlich Hersteller und Modell, um zu vermeiden, dass Soft- oder Hardwareprobleme an einem Standort einen anderen Standort beeinträchtigen.

Kontinuierliche Datenprotokollierung

Ein weiteres Feature, mit dem der Schutz geschäftskritischer Daten gewährleistet werden kann, ist kontinuierliche Datenprotokollierung. Sie bietet eine sekundengenaue Rollback-Funktion. So lassen sich im Falle einer ungewollten Datenveränderung durch Malware oder Benutzerfehler die besten RPOs (Recovery Point Objective) und RTOs (Recovery Time Objective) erzielen. Kontinuierliche Datenprotokollierung

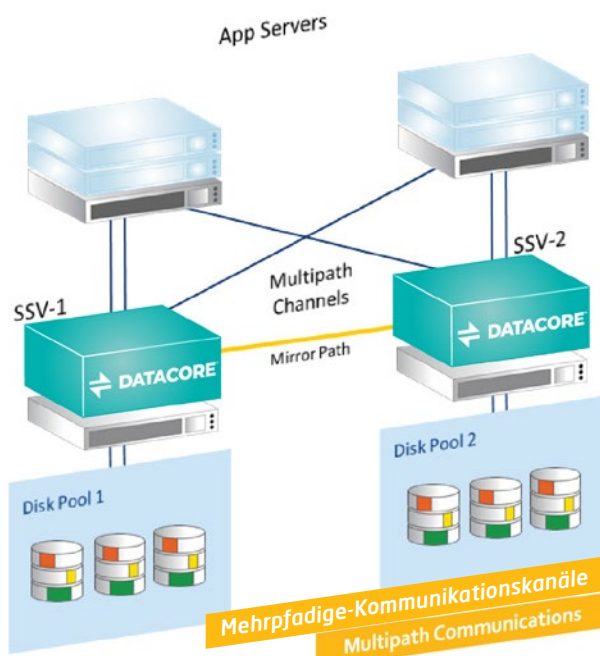
kann außerdem verwendet werden, um bei Volumes Rollbacks zu bestimmten Zeitpunkten auszuführen, falls ein Backup ausgelassen wurde. Kontinuierliche Datenprotokollierung empfiehlt sich besonders für Unternehmen, die strikte Richtlinien zur Datenaufbewahrung und -wiederherstellung erfüllen müssen.

Mehrfadige-Kommunikationskanäle

Datenverfügbarkeit bedeutet nicht nur das Sichern der Daten mit synchroner Spiegelung und kontinuierliche Datenprotokollierung, sondern auch das Gewährleisten ständiger Datenverfügbarkeit über mehrfadige-Kommunikationskanäle.

Falls ein Kommunikationskanal aus irgendwelchen Gründen ausfällt, müssen die Daten dennoch technisch zugreifbar sein. Allerdings besteht aufgrund des Ausfalls eine entsprechende Beeinträchtigung der gesamten Bandbreite für Datenzugriffe. Angesichts der hohen Anforderungen moderner Anwendungen sind hier signifikante Auswirkungen auf Performance und Nutzerzufriedenheit zu befürchten.

Dies verhindert man durch eine eigenständige, unabhängige softwaredefinierte Speicherabstrahierung. So behalten Sie die vollständige Kontrolle über alle Anpassungen und Änderungen Ihrer Speicherarchitektur und Sie gewährleisten Ihre Service-Level unter normalen wie unter außergewöhnlichen Betriebsbedingungen

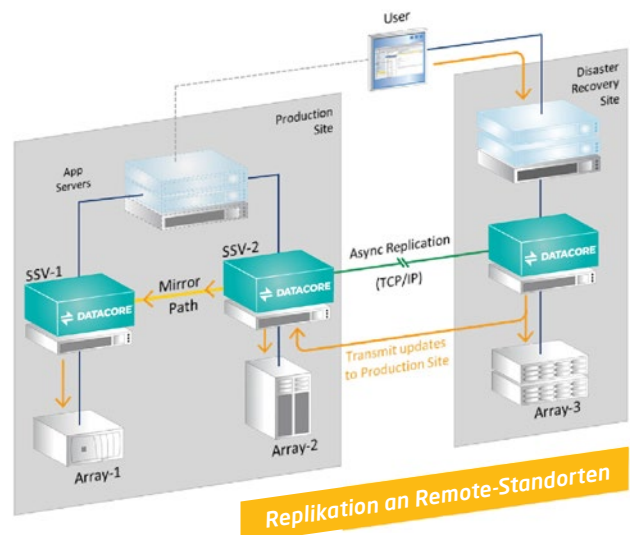


Replikation zu Remote-Standorten

In Sachen Datenverfügbarkeit besteht die letzte Verteidigungslinie in der Replikation zu Remote-Standorten, falls ganze geografische Regionen von einer Katastrophe beeinträchtigt sind. Die Vorstellung, einen kompletten Standort mithilfe von Backups wiederherzustellen, ist eher unrealistisch. Das würde viel zu viel Zeit in Anspruch nehmen. Durch asynchrone Replikation in Kombination mit innovativen Wiederherstellungsmechanismen wird ein potenzielles IT-Desaster zu einem handhabbaren und sogar testbaren Vorgang. Dies kann in kritischen Geschäftsphasen das Zünglein an der Waage zwischen Trendwende und Insolvenz sein.

Bei der asynchronen Replikation werden die Daten an einen Remote-Standort kopiert und sind umgehend verfügbar, falls der Produktionsstandort ausfällt. Asynchrone Replikation bietet höchste Datenintegrität und stellt die flexibelste Notfallwiederherstellungslösung für Betriebssysteme, geschäftskritische Anwendungen und Daten dar.

Kein Notfallwiederherstellungsplan ist vollständig ohne die Möglichkeit, den Failover-Vorgang zu testen und die Datenintegrität am Remote-Standort zu gewährleisten. Idealerweise können Sie das Failover-Verfahren zwischen Standorten beliebig oft testen, ohne die Produktionsumgebung zu beeinträchtigen oder den Replikationsvorgang am Remote-Standort zu unterbrechen. Sie aktivieren dazu einfach den Testmodus- für den Bereich, der überprüft werden sollen. Sobald Ihr Test am Remote-Standorts abgeschlossen ist, deaktivieren Sie den Testmodus



wieder. Der Remote-Standort sollte dann automatisch wieder den normalen Betriebsmodus gehen.

Daten an einem Remote-Standort zu replizieren ist ein etabliertes Vorgehen und dürfte kein Problem darstellen. Aber was geschieht, wenn nach einem erfolgreichen Failover ein anschließender Failback erforderlich wird? Dieser Fall ist deutlich komplizierter. Ausgereifte SDS Lösungen kehren die Richtung der asynchronen Replikation nach einem Failover- oder Failback-Vorgang automatisch um. So ist gewährleistet, dass die andere Seite alle Aktualisierungen erhält, sobald sie wieder voll funktionsfähig ist. Es sollte keine vollständige Neusynchronisierung aller Daten erforderlich sein, um den Ursprungsstandort wieder in Betrieb zu nehmen. So kommt es beim Übergang nicht zu unnötigen Ausfallzeiten.

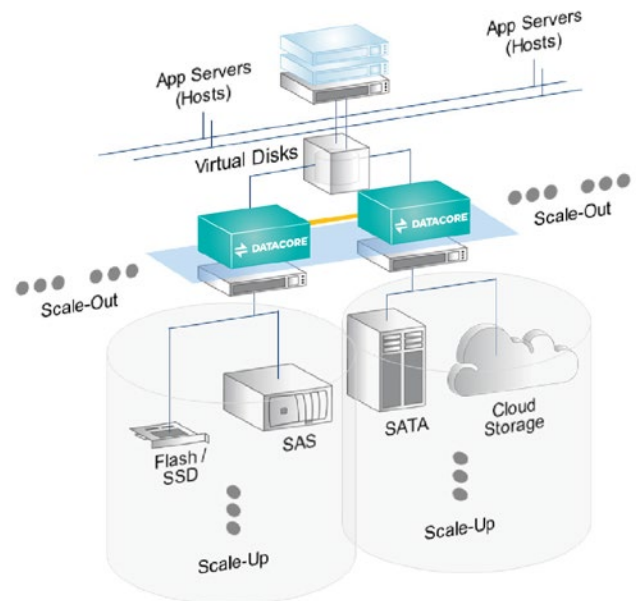
Herausforderung 2: Kapazitäts(aus)nutzung und Skalierung

Weiterhin sorgt eine gute SDS Lösung dafür, dass Ihre Daten effizient gespeichert und Ihre Speicherressourcen somit optimal genutzt werden. Dies wird u.a. durch Thin Provisioning erreicht: Nur bei Schreibvorgängen wird Speicherplatz dynamisch zugewiesen. Im Ergebnis kann die Nutzung der Kapazitäten um den Faktor 4 verbessert werden. Thin Provisioning eliminiert zudem die Notwendigkeit logische Volumen ständig anzupassen, da man einfach große logische Volumen gegenüber Anwendungen präsentieren kann, ohne den Speicherplatz zu belegen. Sehr gute SDS Lösungen erkennen dabei Leerdatenblöcke, die durch die Migration oder Löschung von Daten entstehen können, und machen diese wieder zugänglich.

Auf diesem Weg können Sie Ihre Speicherstruktur außerdem rasch und sicher nachskalieren, damit sie jederzeit Ihrem tatsächlichen Speicherbedarf entspricht. Ob Sie weitere Knoten hinzufügen oder die Kapazität der vorhandenen Knoten erhöhen möchten – beides sollte rasch und ohne Beeinträchtigung der Produktionsumgebung möglich sein.

Dabei bieten beide Formen der Skalierung – ob horizontal oder vertikal – Vor- und Nachteile. Je nach Serverhardware der SDS Instanz kann sich eine vertikale Skalierung empfehlen. Dabei werden dem System zusätzliche Speicherkapazitäten (sowohl Server-intern als auch extern), Netzwerkkomponenten, Prozessoren, weiterer High-Speed-Cache etc. hinzugefügt. Wenn die IOPS-Leistung nicht mehr ausreicht und alle Hardwarekapazitäten bereits erschöpft sind, können zusätzliche Knoten hinzugefügt werden, um

die Last der Anwendungen zu stemmen. Für welchen Weg Sie sich auch entscheiden, eine Betriebsunterbrechung sollte dafür in keinem Fall in Kauf genommen werden.



Herausforderung 3: Speicher-Performance

Speichersysteme müssen extrem leistungsfähig sein, um mit den üblichen Datenzuwachsrate und der zunehmenden Zahl der Anwendungen Schritt halten zu können. Mit SDS lassen sich über zwei Hauptmethoden Performance-Zugewinne erzielen: über Hochgeschwindigkeits-Caching und über Automated Storage Tiering.

Hochgeschwindigkeits-Caching

Die erste Methode macht sich Hochgeschwindigkeits-RAM als Cache zunutze. Die SDS Lösung nutzt die Prozessoren, den Arbeitsspeicher und die I/O-Ressourcen eines Knotens, um komplexe Multithread-Cache-Algorithmen auf allen verwalteten Speichergeräten auszuführen. Die Software sollte dazu mehrere Terabyte (TB) RAM pro Knoten für SAN-weite Mega-Caches reservieren können.

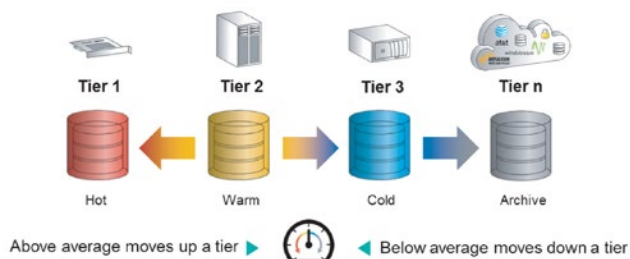
Hochgeschwindigkeits-Caching ist für die Anwendungsleistung maßgebend, da die Geschwindigkeit von DRAM die der schnellsten Flash-Technologien übertrifft und durch ihre Nähe zur CPU zudem Latenzen minimiert werden. DRAM ist immer noch die schnellste Komponente der Speicherarchitektur. Ein Nebeneffekt von DRAM als Cache ist die verlängerte Lebensdauer der nachfolgenden

Speicherkomponenten, da viele kurzfristige Datenänderungen 'abgepuffert' werden.

Automated Storage Tiering

Die zweite Hauptmethode zur Speicherbeschleunigung besteht in Automated Storage Tiering, also der automatischen Zuweisung von Daten zu unterschiedlich schnellen Speicherklassen. Hierbei werden Daten in Echtzeit der Speicherklasse zugewiesen, die der jeweiligen Zugriffsfrequenz entsprechen. In Kombination mit Hochgeschwindigkeits-Speicher wie Flash lässt sich die allgemeine Anwendungsleistung erheblich steigern.

Automated Storage Tiering ermöglicht zudem die nahtlose Integration von Flash und anderen Hochgeschwindigkeits-Speicher in die gesamte Speicherarchitektur: Fügen Sie den Flash-Speicher einfach dem SDS-Speicherpool hinzu, in dem sich auch die 'normalen' Festplattenlaufwerke befinden – fertig. Die Daten werden automatisch zwischen den Speichertechnologien verschoben und die Anwendungen mit dem höchsten Performance-Bedarf erhalten den schnellsten Speicher.



Ein elementarer Vorteil von Storage Tiering ist, dass keine reine All-Flash-Architektur erforderlich ist, um übergreifend Flash-Performance zu erzielen. Üblicherweise sollten 8–15 % des gesamten Speicherpools aus Hochgeschwindigkeits-Speicher bestehen. Automated Storage Tiering sorgt dafür, dass heiße Daten (auf die häufig zugegriffen wird) auf den schnellsten Speicher und kalte Daten (auf die selten zugegriffen wird) auf langsameren und kostengünstigeren Speicher gespeichert werden.

Branchenstatistiken zeigen, dass Unternehmen durchschnittlich nur bei 15 bis 20 % ihrer Daten von Flash-Speicher profitieren. Dies liegt daran, dass es sich beim Rest weitestgehend um kalte Daten handelt. So schreibt ein IT-Analyst in seinem Bericht:

„Zugriffsmuster, Wert und Nutzungseigenschaften der in Speichersystemen gesicherten Daten variieren stark und sind von Geschäftszyklen und organisatorischen Arbeitsabläufen abhängig. Durch diese große Variabilität ist es nicht möglich, Daten wirtschaftlich und effizient auf dem gleichen Speichertyp bzw. in der gleichen Speicherklasse, im gleichen

Format oder auf den gleichen Medien zu speichern.“

Von einer guten SDS Lösung werden bis zu 15 verschiedene Speicherklassen unterstützt, auch Cloudspeicher für Archivdaten, auf die nur selten zugegriffen wird. Werden fortschrittlichere Speichertechnologien verfügbar, können bestehende Speicherklassen nach Bedarf angepasst und neue Ebenen hinzugefügt werden, um die Diversifikation der Architektur voranzutreiben.

Das Ergebnis: überragende Anwendungsleistung bei geringen Gesamtbetriebskosten für die komplette Speicherarchitektur.

Fazit

Unternehmen haben einerseits hohe Ansprüche an die Investitionsrentabilität ihrer IT-Infrastruktur. Andererseits besteht jedoch Bedarf nach extrem leistungsfähigen, zuverlässigen Speichersystemen der Enterprise-Klasse – und diese sind teuer. Software-Defined Storage vereint diese beiden vermeintlich gegensätzlichen Ansprüche.

SDS ist eine eigenständige, unabhängige softwarebasierte Speichervirtualisierungslösung. Sie gibt Ihnen nicht nur die volle Kontrolle über Ihre Speicherarchitektur, sondern bietet Ihnen auch die robusteste Speicherfunktionalität für Unternehmen, die überhaupt verfügbar ist.

Bedenken Sie die Investitions- und Betriebsvorteile einer zweckmäßigen Lösung, die mit Speichergeräten aller Marken und Modelle operiert, die Anwendungsleistung potenziert und Ihnen die modernsten Datenschutzmethoden bietet. Und wenn sich die Umstände im Lauf der Zeit ändern, können Sie Ihre Ressourcen unterbrechungsfrei erweitern, um dem neuen Bedarf gerecht zu werden.

ALLES ZUVOR BESCHRIEBENE UND EINIGES SINNVOLLE MEHR LEISTET DATACORE SDS!

Trevor Pott ist ein angesehener IT-Experte, der seit mehr als 20 Jahren in der Branche tätig ist und Unternehmen unterschiedlicher Größen in seiner Funktion als IT-Experte, Berater, Consultant und Autor zur Seite steht. Pott arbeitet als Systemadministrator bei eGeek Consulting Ltd. und ist ein produktiver Tech Writer und gefragter Anbieterberater, der sowohl die Technologie als auch ihre Anwender kennt.

Weitere Informationen zu DataCore SDS finden Sie unter <https://www.datacore.de/products/software-defined-storage/>.

