HYPERKONVERGENTE INFRASTRUKTUR: SACHBEZOGENE AUSWAHLKRITERIEN

WIE MODERNE UNTERNEHMEN DIE BESTE HYPERKONVERGENTE LÖSUNG FÜR IHREN BEDARF AUSWÄHLEN

Eine Perspektive zur Bewertung von Technologien

Von
Jon Toigo
Vorstandsvorsitzender des Data Management Institute
jtoigo@toigopartners.com

ZUSAMMENFASSUNG

Das neueste Schlagwort bei Speichersystemen lautet „Hyperkonvergenz“. Dabei hängt die genaue Definition von hyperkonvergenten Speichersystemen von den einzelnen Anbietern ab, deren Lösungen sich im Hinblick auf die Unterstützung mehrerer Hypervisoren, der jeweils anfallenden Art der Arbeitslasten sowie die Flexibilität bei Hardware-Komponenten und -Topologie deutlich unterscheiden.

Ganz unabhängig von der jeweiligen Begriffsdefinition der Anbieter ist jedoch die Tatsache, dass der Aufbau einer für das Unternehmen sinnvollen hyperkonvergenten Infrastruktur zwei wesentliche Anforderungen erfüllen muss: Zum einen muss eine Kombination aus Infrastrukturkomponenten und -diensten ausgewählt werden, die den Anforderungen der jeweiligen Arbeitslasten optimal entspricht. Zum anderen muss eine Entscheidung für ein hyperkonvergentes Modell getroffen werden, das bei sich ändernden Speicherbedürfnissen angepasst und skaliert werden kann, ohne das zur Verfügung stehende Budget zu sprengen.

EINFÜHRUNG

Die Entstehung von hyperkonvergenten Infrastrukturen wird vielfach als neueres Phänomen angesehen, doch in der Unternehmens-IT existiert das Konzept schon länger. Die heutige hyperkonvergente Infrastruktur bezieht sich auf die Vereinigung von Server-und Speicher-Technologie mithilfe eines Software-basierten Speicher-Controllers (manchmal auch als SDS bzw. Software-Defined Storage bezeichnet), der sich im Software-Stack des Servers befindet. Die entsprechende Software-Schicht dient als Speicher-Controller und ersetzt die gleichartige Funktionalität in Speichereinheiten. Dies ist im Grunde genommen nichts Neues. Bei den sehr frühen Mainframe-Architekturmodellen waren die Plattenspeichereinheiten nur mit einem Minimum an integrierter Speicher-Software ausgestattet. Die meisten Funktionen zur Speichersteuerung waren in eine Software-Anwendung integriert, die – anstatt auf DASD-Einheiten (Direct Access Storage Devices) – innerhalb des Hauptprozessors ablief.

Die Befürworter einer hyperkonvergenten Infrastruktur greifen diesen Gedanken wieder auf und verknüpfen ihn mit Konzepten einer Architektur aus verteilten Hochleistungsrechnern, wie etwa einem Cluster aus Serverknoten. Ziel ist die Erschaffung eines hyperkonvergenten Modells mit vereinfachter Skalierbarkeit (durch Hinzufügen weiterer Server-Knoten), einer Verbesserung der Verfügbarkeit (durch Datenreplikation ist ein Failover möglich) und einer Senkung der Infrastrukturkosten (durch Austausch schwer zu verwaltender SANs und monolithischer Speichereinheiten durch eher generische Speicherkomponenten, die direkt an die Server angeschlossen werden). Ein solcher Aufbau reicht jedoch noch nicht aus, um eine allgemeine Definition bzw. eine gemeinsame Architektur für eine zeitgemäße hyperkonvergente Infrastruktur zu ermöglichen. Stattdessen ist eine Matrix zu verwenden, um die Eigenschaften der unterschiedlichen auf dem Markt erhältlichen Lösungen darzustellen.

	UNTERSTÜTZT ARBEITSVERTEILUNG MIT EINEM HYPERVISOR	UNTERSTÜTZT ARBEITSVERTEILUNG MIT MEHREREN HYPERVISOREN	UNTERSTÜTZT ARBEITSVERTEILUNG MIT MEHREREN HYPERVISOREN BZW. OHNE VIRTUALISIERUNG
MODELL MIT BESTIMMTER HARDWARE	Auf die virtualisierten Arbeitslasten eines Hypervisors zugeschnitten Spezifische Hardware für Server- und Speicher-installationen	Gemeinsam nutzbare Infrastruktur für Arbeitslasten mit mehreren Hypervisoren Spezifische Hardware für Server- und Speicher-installationen	Gemeinsam nutzbare Infrastruktur für Arbeitslasten mit mehreren Hypervisoren bzw. ohne Virtualisierung Spezifische Hardware für Server- und Speicher-installationen
MODELL MIT FLEXIBLER HARDWARE	Auf die virtualisierten Arbeitslasten eines Hypervisors zugeschnitten Flexible Hardware-Auswahl für Server- und Speicher bei vorgegebenem Modell für Speicherimplementierung	Gemeinsam nutzbare Infrastruktur für Arbeitslasten mit mehreren Hypervisoren Flexible Hardware-Auswahl für Server- und Speicher und Unterstützung für gemischte Implementierungs-modelle	Gemeinsam nutzbare Infrastruktur für Arbeitslasten mit mehreren Hypervisoren bzw. ohne Virtualisierung Flexible Hardware-Auswahl für Server- und Speicher und Unterstützung für gemischte Implementierungsmodelle

Wie der obigen Matrix zu entnehmen ist, lassen sich die Unterschiede bei den hyperkonvergenten Infrastrukturlösungen (auch als hyperkonvergenter Speicher bekannt, da die Technik ursprünglich dazu verwendet wurde, hinter den Anwendungs-Servern Speicherknoten für den Rechnerverbund anzulegen) an zwei Parametern festmachen:

1. **das Hardware-Modell hängt mit der Lösung zusammen und**

2. **die Lösung sieht vor, Datenspeicherung aus unterschiedlichen virtuellen und physikalischen Arbeitslasten zu unterstützen.**

Einige hyperkonvergente Speicherknoten sind vorgefertigt oder an eng gefasste Komponentenlisten gebunden, sodass der Knoten nur als Modell mit fester Hardware ausgeführt werden kann. Andere hingegen weisen mehr Flexibilität bezüglich der Hardware-Komponenten auf, die zur Bereitstellung des jeweiligen Speicherdienstes verwendet werden können. Der oben in der Matrix genannte zweite Parameter betrifft die jeweilige Unterstützung für die Arbeitsverteilung. Einige hyperkonvergente Speicherknoten sind auf einen spezifischen Server-Hypervisor zugeschnitten und können nur dann verwendet werden, wenn die Arbeitsverteilung durch diesen Hypervisor virtualisiert und gesteuert wird. Andere hyperkonvergente Speicherknoten hingegen können zusammen mit unterschiedlichen Hypervisoren verwendet werden, verfügen aber dennoch über eine gemeinsame Verwaltungskonsole für alle Instanziierungen der Knoten.

Das dritte Lösungsmodell bietet Speichermöglichkeiten für sowohl eine virtualisierte als auch für eine nicht virtualisierte Arbeitsverteilung. Dies ist insbesondere für Unternehmen wichtig, die mehrere Hypervisoren einsetzen und gleichzeitig nicht virtualisierte Anwendungen betreiben, wie etwa Hochleistungsdatenbanken zur Transaktionsabwicklung. Führende Analysten vermuten, dass diese Mischung aus virtualisierter und nicht virtualisierter Arbeitsverteilung in manchen Unternehmen bis mindestens zum Ende des Jahrzehnts zum Einsatz kommen wird. Im Rahmen von Planungen sollte dieser Punkt unbedingt berücksichtigt werden, da sich hieraus unmittelbare Auswirkungen auf die kurz- und langfristige technische Eignung der gewählten hyperkonvergenten Lösung wie auch deren Investitionserträge ergeben.

	UNTERSTÜTZT ARBEITSVERTEILUNG MIT EINEM HYPERVISOR	UNTERSTÜTZT ARBEITSVERTEILUNG MIT MEHREREN HYPERVISOREN	UNTERSTÜTZT ARBEITSVERTEILUNG MIT MEHREREN HYPERVISOREN BZW. OHNE VIRTUALISIERUNG
MODELL MIT BESTIMMTER HARDWARE	vmware EVO RAIL	NUTANIX	UNDER DEVELOPMENT
MODELL MIT FLEXIBLER HARDWARE	VMware Virtual SAN	StarWind SOFTWARE	DataCore SOFTWARE

In dieser Grafik, in der anstelle von Beschreibungen Produktnamen verwendet werden, sind einige Beispiele von Produkten aufgeführt, die in die Kategorien der Matrix passen. EVO:Rail von VMware ist ein Beispiel für ein hyperkonvergentes Speichermodell, bei dem ein einzelner Hypervisor in Verbindung mit einer vorgegebenen Hardware-Lösung für Knoten unterstützt wird. Nutanix hingegen verkauft sein entsprechendes Produkt als feste Hardware-Einheit, die jedoch auch zusammen mit mehreren Hypervisoren verwendet werden kann. Virtual SAN von VMware kann als hyperkonvergentes Modell angesehen werden, bei dem zwar nur die Arbeitsverteilung eines einzelnen Hypervisors unterstützt wird, das bei der in den Knoten eingesetzten Speicher-Hardware aber durchaus flexibel ist.

Im Gegensatz dazu unterstützt die Lösung von StarWind Software, einem Drittanbieter von SDS-Produkten, mehrere Hypervisoren und unterschiedliche Instanziierungen, bietet aber zugleich eine gemeinsame Verwaltung für alle Instanzen. Auch die Lösung von DataCore Software unterstützt mehrere Hypervisoren und nicht virtualisierte Arbeitslasten. Zudem bietet sie eine gemeinsame Verwaltung und eine große Flexibilität bei der zu verwendenden Hardware.

Diese Matrix kann als Ausgangspunkt verwendet werden, um die unterschiedlichen Möglichkeiten beim Aufbau einer hyperkonvergenten Infrastruktur aufzuzeigen. Im ersten erforderlichen Schritt sollten die verfügbaren Produkte anhand der einzelnen Betriebsmodelle, der Unterstützung für Arbeitslasten und der Abhängigkeit, von bestimmter Hardware kategorisiert werden, um so die bestmögliche Lösung für die jeweilige Unternehmens-Anforderung herauszuarbeiten.

Im zweiten Schritt können die im vorliegenden Dokument genannten Schlüsselkriterien dazu verwendet werden, die unterschiedlichen Möglichkeiten zu bewerten.

SCHLÜSSELKRITERIEN ZUR AUSWAHL DER RICHTIGEN HYPERKONVERGENTEN INFRASTRUKTUR-TECHNOLOGIE

Es gibt zahlreiche Möglichkeiten, um die Auswahlkriterien für hyperkonvergente Speichersysteme festzulegen. Hierbei ist eine unternehmerisch durchdachte Auswahl stets die beste. Anders formuliert: die Technologie mit der größtmöglichen Kostensenkung, der höchsten Verfügbarkeit, den besten Performancemerkmalen und der optimalen Eignung für die angedachte Aufgabe.

Kostenaspekte

In den meisten IT-Abteilungen hat die Eindämmung von Kosten heutzutage oberste Priorität. Dies gilt insbesondere für Niederlassungen oder externe Büros, aber auch für kleinere bis mittlere Unternehmen, die nicht über die IT-Budgets großer IT- oder Cloud-Dienstleister verfügen.

Kosten ergeben sich aus Anschaffungsinvestitionen, Betriebsausgaben und Arbeitsaufwand. Eine effiziente IT-Abteilung schafft nur die zur Speicherung der Daten erforderliche Hardware an und plant hierbei Kapazitätsreserven ein, um Anforderungsspitzen abfedern zu können. Wenn Speicherkomponenten zusammenhängend und einheitlich verwaltet werden können, führt dies zu einem geringeren Personalbedarf für die Infrastruktur sowie zu niedrigeren Betriebskosten.

Zieht man eine hyperkonvergente Speicherinfrastruktur in Erwägung, sollten folgende Kosten berücksichtigt werden:

1. Hardware-Abhängigkeiten: Einige hyperkonvergente Server- bzw. Speicherlösungen bestehen aus vorgefertigten Einzelgeräten oder kompletten Hardware-Plattformen (von bestimmten Herstellern), die vom jeweiligen Hypervisor-Anbieter „zertifiziert“ sind. In beiden Fällen wird eine Hardware-Bindung geschaffen, die den traditionellen bzw. veralteten einheitlichen Speicherstrukturen ähnelt, welche die Anbieter eigentlich ersetzen wollen. Idealerweise sollte aber eine hyperkonvergente Lösung gesucht werden, bei der es keine solchen Hardware-Abhängigkeiten gibt. Tatsächlich bietet nur ein Hardware-unabhängiger Ansatz die Möglichkeit, einzelne Teile des Knotens zu unterschiedlichen Zeiten zu skalieren, und zwar abhängig vom Bedarf an bzw. der Verfügbarkeit von neuer Technologie. Die vordringlichste bzw. grundlegendste Frage lautet daher: „Welche Hardware benötige ich für eine funktionsfähige Lösung?“

   Auch im Hinblick auf Skalierbarkeit sollten Hardware-Abhängigkeiten bedacht werden. Um es einmal deutlich zu sagen: Je eng gefasster eine Hardware-Spezifikation ausgelegt ist, desto geringer ist die Möglichkeit, einzelne Komponenten modular zu skalieren. Angesichts der üblichen sprung-

haften Entwicklung, die sich einerseits bei Computer-Komponenten (wie Prozessoren, Verbindungselementen und Netzwerken) und andererseits bei Speichertechnologien abspielt, können sich hieraus enorme Probleme ergeben. Die jeweiligen Zeiträume, in der die Technologien bestimmter Komponentenbereiche verbessert werden, unterscheiden sich teilweise erheblich. Hieraus ergibt sich häufig die Notwendigkeit, einzelne oder mehrere Komponenten auszutauschen, um mit aktuellen Neuerungen Schritt zu halten. Je größer also die Hardware-Abhängigkeit einer Geräteeinheit, desto geringer die Möglichkeiten, technische Innovationen nutzen zu können. Zudem kann die Bindung an eine eng gefasste Hardware-Liste die tatsächlichen Kosten der Lösung nach oben treiben.

2. Mehrfachknoten-Architektur: Einige namhafte Anbieter von Server-Hypervisoren bieten hyperkonvergente Speichermodelle an, die bereits in der Grundausstattung mindestens drei (oder sogar mehr) Cluster-Knoten benötigen. Ein Knoten benötigt üblicherweise einen physischen Server, eine Lizenz für die Speichersoftware, Cluster-Software (ob als Teil eines Hypervisor-Software-Pakets, des Betriebssystems oder einer spezialisierten Drittanbieter-Software), Flash-Speichergeräte sowie eine Speichereinheit oder ein JBOD-System. Gemäß einem kürzlich veröffentlichten Bericht[1] belaufen sich die Kosten pro Knoten für die hyperkonvergente Software eines führenden Hypervisor-Anbieters (mit der Bezeichnung „Virtual SAN“) auf 8.000 bis 11.000 Dollar für die Software-Lizenzen und auf 8.000 bis 14.000 Dollar für die Server- und Speicher-Hardware. Diese Zahlen sind dann mit der Anzahl der Knoten zu multiplizieren, die für den Aufbau einer hyperkonvergenten Infrastruktur erforderlich sind. Im Mindestfall sind dies drei Knoten, aber aus Gründen der Verfügbarkeit und der Leistung werden vier Knoten empfohlen. Im Gegensatz dazu erfordern die hyperkonvergenten Server-Speicher-Systeme einiger Drittanbieter in der Grundausstattung lediglich zwei physische Knoten und ermöglichen zudem die Nutzung weniger kostenintensiver Hardware (z. B. SATA- statt SAS-Festplatten).

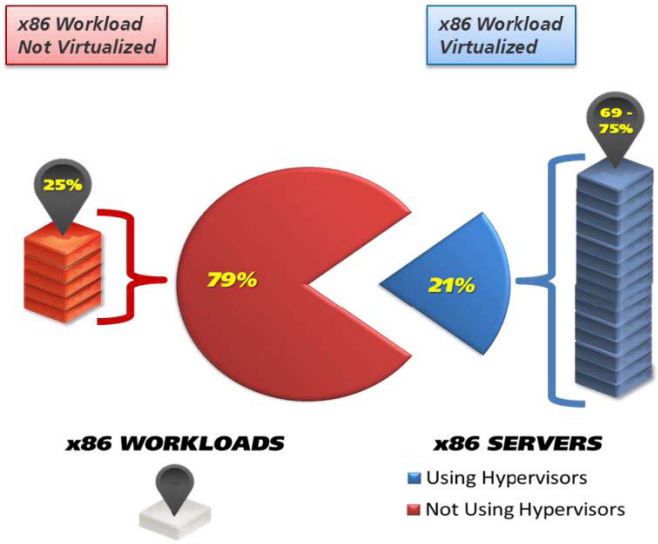
3. Verwaltbarkeit: Ginge es nach einigen führenden Anbietern von Server-Hypervisoren (bzw. nach den meisten Anbietern von hyperkonvergenten Geräten), würden zur Verwaltung aller angeschlossenen Ressourcen nur der vom Anbieter verkaufte Software-Stack sowie die spezialisierten Dienste des Anbieters verwendet. Für die Handhabung der Speichersysteme stellen die Anbieter von Hypervisoren Interfaces für Verwaltungsfunktionen zur Verfügung. Diese dienen zur Datenspiegelung und -replikation auf unterschiedlichen Speicherknoten, zum sog. „Thin Provisioning“ von Speicherressourcen, zur Deduplikation und Kompression sowie zur Durchführung sonstiger Funktionen auf Array-Controllern. Im Wesentlichen fassen sie die einst von den Speicher-Array-Anbietern als Unterscheidungsmerkmal angepriesenen Mehrwertdienste in einem zentralen, Software-basierten Controller zusammen, was vorgeblich der einfacheren Verwaltung von stetig skalierenden Infrastrukturen dient. Der Betrieb der einzelnen Dienste sollte allerdings überprüft werden, damit sichergestellt werden kann, dass die jeweilige Funktionalität in der entsprechenden Infrastruktur auch tatsächlich eingesetzt werden soll.

   Nur weil etwa der Kompressionsdienst eines bestimmten Produktes eindrucksvoll funktioniert, muss dessen netzwerkweiter Replikationsdienst nicht automatisch der beste auf dem Markt erhältliche sein.Sie sollten sich ebenfalls bewusst machen, dass zwar alle Anbieter von hyperkonvergenter Software darin übereinstimmen, dass Speichersoftware-Dienste in einem Software-Stack außerhalb von Speichereinheiten implementiert werden müssen, viele der Anbieter aber gleichzeitig den Gedanken scheuen, das Kapazitätsmanagement ebenfalls außerhalb der Speichereinheiten zu installieren. Dies führt bei vielen hyperkonvergenten Lösungen zu einer nicht unerheblichen Begrenzung, da das Kapazitätsmanagement so zu einer separaten Verwaltungsmaßnahme wird, die auf jedem einzelnen Speichergerät durchzuführen ist und in vielen Fällen spezielle Tools bzw. Fachkenntnisse erfordert.

1 http://www.infoworld.com/article/2608637/data-center/data-center-review-vmware-virtual-san-turns-storage-inside-out.html

Sofern Speicherkapazität „virtualisiert“ oder in einem virtualisierten Ressourcen-Pool zusammengefasst wird, kann dessen Verwaltung alternativ auch zentralisiert und dadurch in einer verteilten Infrastruktur deutlich einfacher verfügbar gemacht werden, so wie es bei anderen Speicherdiensten der Fall ist.

Sofern eine Skalierung durchgeführt werden soll, ohne dass Kapazitäten gebündelt werden, könnte dies den gesamten Nutzen einer hyperkonvergenten Infrastruktur sogar gefährden. Ohne die Virtualisierung von Kapazitäten wird die Speicherverwaltung in einzelne Bereiche zerlegt. Dies gilt insbesondere dann, wenn mehrere Standorte (wie Niederlassungen oder externe Büros) von einer einzelnen Stelle aus verwaltet werden müssen oder wenn, wie in großen Unternehmen, mehrere virtuelle Server-Hypervisoren verwendet werden, die jeweils mit einem eigenen hyperkonvergenten Modell arbeiten. Die zuletzt genannte Problemstellung – also die Nutzung von Speicher-Ressourcenn in Umgebungen mit mehreren Hypervisoren – ist momentan vielleicht noch nicht vordringlich, wird es aber möglicherweise in einigen Jahren sein. Denn führende Analysten gehen davon aus, dass größere IT-Abteilungen bis 2016 75 % der Arbeitslasten virtualisiert haben werden. Allerdings werden die meisten IT-Abteilungen dann mehr als einen Hypervisor nutzen, wobei bis zu 25 % der Arbeitslasten noch immer ohne Virtualisierung und mit veralteter Speicher-Hardware betrieben werden.



By 2016, according to IDC and Gartner…

Wenn hinter unterschiedlichen Hypervisoren jeweils isolierter Speicher zum Einsatz kommt, hat dies Auswirkungen auf die Kosten für die Verwaltung, d. h. es können höhere Kosten verursacht werden. Ein solcher Fehler mag noch nicht offensichtlich sein, wenn man eine hyperkonvergente Lösung für einen bestimmten Standort oder einen Rechnerverbund in Erwägung zieht. Aber er wird definitiv dann in den Mittelpunkt gerückt, wenn die Netzwerkumgebung verwaltet und ausgebaut wird.

Fazit: Die Auswahl einer hyperkonvergenten Infrastrukturtechnologie, die eine Hardware unabhängige Verwaltung per Hypervisor unterstützt, ist der Schlüssel zur Reduzierung des zur Verwaltung der Infrastruktur erforderlichen Personals.

4. Nutzungseffizienz der Hardware: Natürlich ist es im Sinne einer Kostensenkung ebenfalls wichtig, sich für eine hyperkonvergente Infrastruktur zu entscheiden, bei der die Hardware optimal genutzt wird. So können zwar die meisten hyperkonvergenten Produkte auf dem Markt dynamischen RAM und Flash-Speicher nutzen, um Caches und Puffer zur Verbesserung der Anwendungsleistung anzulegen. Aber nicht alle verfügbaren Produkte nutzen Caches und Puffer wirklich effektiv oder ermöglichen es, auf die wachsende Vielfalt der auf dem Markt erhältlichen Drittprodukte zurückzugreifen. So eignet sich DRAM besser für Schreib-Caches als Flash-Speicher, was jedoch aus den Marketingunterlagen einiger Anbieter von hyperkonvergenter Infrastruktur nicht unbedingt hervorgeht.

Oftmals wird Flash-Speicher für Zwecke empfohlen, für die er ungeeignet ist. Auch kommt es häufig vor, dass der Anbieter eines hyperkonvergenten Produkts die Möglichkeiten des Anwenders einschränkt, preisgünstige Komponenten zu nutzen oder zwecks Beschleunigung der Anwendungsleistung die besten auf dem Markt erhältlichen Technologien einzusetzen – und zwar zugunsten von Produkten, die vom Anbieter zertifiziert wurden. Solche bekannten Probleme mit veralteten Speicherkomponenten werden von den Befürwortern hyperkonvergenter Lösungen gerne verteufelt.

Hinzu kommt, dass es bei einigen Anbietern hyperkonvergenter Produkte zwingend erforderlich ist, dass „veralteter" Speicher durch neue, zertifizierte Komponenten ersetzt wird, was wiederum die Kosten für Aufbau und Betrieb der hyperkonvergenten Infrastruktur nach oben treiben kann. Die meisten Unternehmen schaffen sich Speichersysteme mit der Zielsetzung an, nach fünf bis sieben Jahren Investitionserträge (ROI) zu realisieren. Die Vorgabe, veraltete Speicherkomponenten mit hohem Aufwand zu ersetzen, bei denen die erwarteten Investitionserträge (ROI) noch nicht erzielt wurden, kann erhebliche Investitionskosten mit sich bringen, die in jedem Fall berücksichtigt werden müssen.

Verfügbarkeitskriterien

Verfügbarkeit ist ein weiteres maßgebliches Kriterium in Bezug auf die Anschaffung von hyperkonvergenter Technologie. Angesichts der Tatsache, dass Daten die Lebensader eines jeden Unternehmens darstellen, ist vollkommen klar, dass diese 24 x 7 zugänglich sein müssen. Die folgenden Richtlinien sollten daher offensichtlich sein:

5. Zwar sind hochverfügbare Cluster mit Funktionen zur Datenspiegelung das Markenzeichen einer hyperkonvergenten Infrastruktur, aber... Die von der in Frage kommenden hyperkonvergenten Software bereitgestellten Funktionen für das Clustern bzw. die Spiegelung von Daten müssen absolut stabil funktionieren. Zudem muss die Software dafür sorgen, dass die Daten in Speicher-Caches und -Puffern gespiegelt werden und auf SSD-Festplatten oder Magnetplatten gespeichert werden. Idealerweise sollte die Software nicht voraussetzen, dass jeder Knoten identisch ausgestattet ist. Ferner sollte, und zwar aus Kostengründen, die Anzahl der für die Spiegelung eines hochverfügbaren Clusters erforderlichen Hardware-Knoten in der Grundausstattung nicht mehr als zwei betragen.

6. Hochverfügbarkeit (Daten-Spiegelung) muss ohne Unterbrechung von laufenden Anwendungen getestet und verifiziert werden können. Zwar gibt es die Daten-Spiegelung als Vorrichtung zum Schutz von Daten schon seit den ersten Tagen von Netzwerken, doch wurden gespiegelte Daten nur selten getestet, da während des Testens Anwendungen gestoppt und im Anschluss daran die Arbeitslasten und Replikationsprozesse erneut synchronisiert werden mussten. Die von Ihnen favorisierte Lösung sollte wesentlich einfacher gestrickt sein und kurzfristige Überprüfungen von Spiegelungsprozessen ermöglichen. Sie sollten ebenfalls darüber nachdenken, ob das Failover gespiegelter Daten am sinnvollsten automatisch durchgeführt wird oder zunächst manuell bestätigt werden sollte. Stellen Sie sich „geclusterten Failover" wie eine Alarmanlage für Ihr Haus vor: Selbst die beste Anlage sorgt manchmal für einen Fehlalarm.

7. Verfügbarkeit ist keine Einbahnstraße: Nach dem Auftreten eines Failover muss eine entsprechende Funktionalität auch dafür sorgen, dass ein Fail Back möglich ist. Dies hat üblicherweise zur Folge, dass gespiegelte Schreibzugriffe so lange gepuffert werden, bis die Verbindungen wieder hergestellt worden sind – eine wichtige aber oftmals nicht überprüfte Anforderung.

8. Lokale Hochverfügbarkeit ist keinesfalls höher anzusiedeln als Desaster Recovery. Hochverfügbare Spiegelungstechnologie für den lokalen Failover sollte durch „Stretch Clustering" bzw. Metrocluster

ergänzt werden, die eine räumlich entfernte Datenreplikation zwischen zwei an unterschiedlichen Orten befindlichen Knoten ermöglichen. Eine tatsächlich verfügbare hyperkonvergente Lösung sollte über „Stretch Clustering“ mit netzwerkweiter synchroner und asynchroner Replikation verfügen. Vorzugsweise sollte diese Lösung auch keine identischen Knotenkonfigurationen voraussetzen.

Passend zu den Einsatzkriterien

Zwei weitere Kriterien, die in jedem Ratgeber für Käufer von hyperkonvergenter Server- und Speicher-Technologie enthalten sein sollten, sind Aufgabe und Performance. Hierbei spielen mindestens zwei Faktoren eine wesentliche Rolle:

9. Die Beschleunigung von DRAM (und ggf. Flash-Speicher) muss für solche Server unterstützt werden, auf denen mehrere virtuelle Maschinen ausgeführt werden. Die Verwendung speicherbasierter Caches und Puffer ermöglicht die Beschleunigung der Anwendungsleistung, selbst wenn die eigentlichen Ursachen einer niedrigen Anwendungsleistung nicht bei den I/O‘s des Speichersystems liegen. Idealerweise wird Flash-Technik unterstützt – was aber keine zwingende Voraussetzung ist.

10. Die „Gesamteignung / Tauglichkeit“ ist ebenfalls von Wichtigkeit. Der Begriff bezieht sich darauf, wie gut die Technologie in die Umgebung bzw. die Szenerie passt, in der sie eingeführt wird. Es beginnt also bereits bei der Stellfläche der hyperkonvergenten Plattform und den Anforderungen an Lärmschutz, Stromversorgung, Heizung und Klima: All diese Faktoren können bei der Ausstattung von Niederlassungen oder externen Büros, in denen sich kein Geräteschrank oder kein Rechenzentrum befindet, von großer Wichtigkeit sein. „Eignung“ bezieht sich auch auf die Lernkurve, die für Verwaltung und Betrieb der Hardware und des Software-Stacks erforderlich ist sowie auf die Verfügbarkeit von Fachkräften vor Ort.

    Und zu guter Letzt bezieht sie sich auch auf die Kompatibilität der Speicher-Infrastruktur mit unterschiedlichen Hypervisoren oder mit nicht virtualisierten Anwendungen und deren Speicheranforderungen. Je nach Auslastung, Arbeitslast und Umgebungsbedingungen kann entweder eines oder aber ein anderes hyperkonvergentes Modell besser für die entsprechende Situation geeignet sein. Es lohnt sich auf jeden Fall, eine Liste mit den sich aufgrund der örtlichen Gegebenheiten, der Arbeitsauslastung und den Anwendern ergebenden Einschränkungen zu erstellen, bevor alternative Produkte in Erwägung gezogen werden.

Auch wenn diese zehn Kriterien möglicherweise nicht vollständig sind, können Sie mit ihrer Hilfe das Feld der Möglichkeiten dahingehend aussortieren, dass Sie nur noch solche Optionen in Betracht ziehen, die Ihrem Unternehmen langfristig zugute kommen. Wenn ich persönlich Geld zu investieren hätte, würde ich eine hyperkonvergente Infrastrukturlösung mit folgenden Eigenschaften bevorzugen: Das System ist unabhängig von bestimmter Hardware oder bestimmten Hypervisoren. Das System unterstützt DRAM und Flash-Geräte beliebiger Anbieter.

Die Festplatten-Infrastruktur ermöglicht sowohl DAS- als auch traditionelle SAN-Speicher- systeme (sodass ich mit den letztgenannten meine erwarteten Investitionserträge (ROI) in vollem Umfang realisieren kann). Und idealerweise würde die gewählte Infrastrukturlösung die gesamte Speicherkapazität virtualisieren, sodass ich über ein einziges Software-Interface die Kapazitätszuweisung und die spezifischen Speicherdienste für verschiedene Standorte und für heterogene Speichersysteme verwalten kann.

ZUSAMMENFASSUNG: PRÜFLISTE ZUR BEWERTUNG VON MÖGLICHKEITEN

Nutzen Sie die nachstehende Prüfliste zur Bewertung von hyperkonvergenten Speichermöglichkeiten:

1. Hardware-Abhängigkeiten

    a) Welche Hardware benötige ich für eine funktionsfähige Lösung?
    b) Können die Komponenten einzeln aufgerüstet werden?
    c) Bei welchem Anbieter/welchen Anbietern kann ich die Knoten, die Ausrüstung bzw. die Komponenten für meine Lösung erwerben?

2. Mehrfachknoten-Architektur

    a) Welche Mindestanzahl an Knoten garantiert eine hohe Verfügbarkeit?
    b) Welche Anzahl an Knoten wird in Bezug auf Hochverfügbarkeit und Performance empfohlen?
    c) Wie hoch sind die Gesamtkosten für einen Knoten (Hard- und Software)?

3. Verwaltbarkeit

    a) Welche Speicherdienste sollen bei der Lösung zum Einsatz kommen?
    b) Sind die Speicherdienste die besten auf dem Markt erhältlichen oder von mittlerer Qualität?
    c) Beinhaltet der Speicherdienst ein Kapazitätsmanagement?
    d) Wie wird bei dieser Lösung Speicherkapazität unabhängig von der Rechenleistung skaliert?
    e) Funktioniert die Lösung mit anderen Hypervisoren sowie mit nicht virtualisierten Umgebungen?

4. Nutzungseffizienz der Hardware

    a) Benötigt die Lösung Flash-Technik? Falls ja, wie wird diese benutzt?
    b) Kommen bei der Lösung DRAM und Flash in geeigneter Weise zum Einsatz?
    c) Ist es zur Umsetzung der Lösung erforderlich, dass Sie Ihre bereits vorhandenen Speichersysteme ausrangieren (also mit hohem Aufwand vollständig ersetzen)?

5. Hochverfügbare Cluster

    a) Ist die Cluster-Funktionalität der Knoten ein wesentlicher Bestandteil der Lösung? Oder wird diese über das Betriebssystem, den Hypervisor oder eine externe Cluster-Software bereitgestellt?
    b) Wie wird die Spiegelung von Daten konfiguriert und verwaltet?
    c) Müssen alle Knoten innerhalb des Clusters identische Konfigurationen aufweisen?
    d) Werden Speicher-Caches und -Puffer zwischen Knoten gespiegelt?

6. Test und Validierung der gespiegelte Daten

    a) Wie werden gespiegelte Daten bei Einsatz dieser Lösung getestet?
    b) Besteht die Möglichkeit, gespiegelte Daten zu überprüfen, bevor der Failover zwischen den Knoten stattfindet?
    c) Läuft die Failover eines Knotens automatisch ab oder sind manuelle Eingriffe erforderlich?

7. Fail-Back

    a) Wie werden gespiegelte Daten bei Einsatz dieser Lösung getestet?
    b) Besteht die Möglichkeit, gespiegelte Daten zu überprüfen, bevor der Failover zwischen den Knoten stattfindet?

8. Metrocluster

    a) Unterstützt die Lösung verteilte Cluster bzw. Metrocluster im gesamten WAN bzw. MAN?
    b) Beinhaltet die Lösung die Möglichkeit zur synchronen und asynchronen Spiegelung?
    c) Müssen lokale und remote Knoten eine identische Hardware aufweisen?

9. Unterstützung für DRAM

    a) Wird dynamischer RAM zum Puffern von Schreibzugriffen unterstützt?
    b) Wie wird ein DRAM-Puffer vor Datenverlust geschützt?

10. Allgemeine Eignung

    a) Ist die hyperkonvergente Lösung im Hinblick auf Energieverbrauch, Kühlungsanforderungen, Lärmschutzanforderungen und den verfügbaren physischen Platz für die Umgebung geeignet, in der sie zum Einsatz kommen soll?
    b) Hat das Personal die entsprechenden Fachkenntnisse, um die Lösung einzusetzen und zu verwalten?
    c) Sofern Schulungen erforderlich sind, werden diese vom Anbieter durchgeführt? Zu welchen Kosten?
    d) Unterstützt die Lösung die bei Ihnen vorkommende Arbeitslasten bzw. die vorhandenen Hypervisoren ?

Letzten Endes unterscheiden sich die Anforderungen für die optimale Auswahl eines hyperkonvergenten IT-Systems von Unternehmen zu Unternehmen. Im Allgemeinen besteht eine optimale Auswahl aus einem System, bei dem Server-Hypervisor(en) und Speicher-Hardware „unabhängig“ sind. Darüber hinaus müssen die entsprechenden Planungsbeauftragten die Alternativen anhand der finanziellen Vorgaben und den Anforderungen an Verfügbarkeit und Leistung eingrenzen, um so das am besten geeignete System zu bestimmen.

