

Optimal aufgebaut

von Wolfgang Weith

Der Einsatz von vSphere 8 verlangt einige Planung – unabhängig davon, ob in bestehenden Umgebungen ein Upgrade erfolgt oder eine Neuinstallation auf der grünen Wiese ansteht. Ausgehend von Workload-Domains liefert unser Artikel Hinweise zur Gestaltung von Cluster, Host, Storage und Netzwerk.

Sein seiner Einführung vor rund 20 Jahren hat sich das vSphere-Ökosystem kontinuierlich weiterentwickelt, dennoch bleiben Verfügbarkeit, Verwaltbarkeit, Performance und Sicherheit die vier grundlegenden Anforderungen an eine leistungsfähige vSphere-Architektur. Je nach Einsatzzweck und Exposition kann einer dieser Pfeiler natürlich dominieren – so liegt aktuell bei vielen Unternehmen der Fokus verstärkt auf Sicherheit – jedoch sollten IT-Verantwortliche die anderen beim Design der Umgebung nie vernachlässigen.

Grundlage des Designs

Für das Design einer vSphere-Umgebung stellte VMware in der Vergangenheit den Ansatz der "VMware Validated Designs" [1] bereit. Dieser ist zwar noch im Netz verfügbar, wird vom Hersteller aber nicht mehr offiziell unterstützt. Die logische Fortsetzung dieser Richtlinien lässt sich aus "VMware Cloud Foundation" (VCF) ableiten.

Dabei handelt es sich um ein integriertes Softwarepaket, das die Bereitstellung und Verwaltung einer kompletten Hybrid-Cloud-Infrastruktur ermöglicht. VCF soll hier nicht das eigentliche Thema

sein, sondern der architektonische Ansatz dahinter. Denn dieser stellt aktuell die beste Hilfestellung dar, um ein Best-Practices-Design aufzubauen. Dieses ermöglicht IT-Verantwortlichen, ein standardisierbares und damit für diverse Anwendungsfälle wiederholbares und anpassungsfähiges Architekturmuster und eine daraus resultierende Plattform zu bauen. Eine solche Umgebung ist in der Lage, verschiedenste Workloads mit deren Performance-, Sicherheits- und Skalierungsanforderungen abzubilden. VCF berücksichtigt dabei Industriestandards und VMwares Erfahrungen aus tausenden Kundeninstallationen.

VMware Cloud Foundation ist sehr restriktiv in Bezug auf das Design und auch die supporteten Architekturausprägungen. Jedoch lassen sich auch für Nicht-VCF-Umgebungen viele Best Practices, Designmuster und Architekturideen ableiten, die in jeder vSphere-Umgebung sinnvoll sind. In den folgenden Betrachtungen werden Sie jedoch von VCF kaum noch etwas lesen – stattdessen referenzieren wir auf vSphere mit seinen verschiedenen Designebenen:

- High-Level-Design
- RZ-Layout

- Workload-Domain und Cluster
- VM-Anforderungen
- Host-Design
- vSphere-8-Update

Mit Workload-Domains Umgebungen gestalten

Als oberstes Designziel definieren wir, dass wir Workloads mit den bestmöglichen SLAs betreiben möchten. Grundsätzlich gibt es in jeder vSphere-Umgebung zwei wesentliche Workload-Typen zu unterscheiden: die Managementkomponenten und die regulären Workloads, also Applikationen, die das Kerngeschäft unterstützen. Die physische oder logische Trennung des Managements von anderen Workloads reduziert das Risiko von Störungen oder Sicherheitsverletzungen. Ein unerwarteter Fehler oder eine Überlastung in den regulären Workloads kann somit die vSphere-Administration nicht beeinträchtigen, was zu einer besseren Stabilität und mehr Kontrolle führt. Aus dem VCF-Konzept leitet sich hier der Aufbau einer Management-Domain und zusätzlich einer oder mehrerer Workload-Domains als Best Practice ab. Eine solche Domain kann wiederum mehrere Cluster enthalten, ein Beispiel zeigt Bild 1.

Das Domain-Konzept aus VCF auf vSphere-Umgebungen zu übertragen ist sehr sinnvoll, denn eine Domain ist eine isolierte und skalierbare Einheit, die Ressourcen und Services für spezifische Workloads bereitstellt. Es handelt sich also um eine logische oder physische Abgrenzung innerhalb der VMware-Infrastruktur. In Bild 1 ist die Management-Domain physisch abgetrennt, um die Sicherheit der Management-Workloads zu erhöhen.

Eine Workload-Domain stellt ebenfalls eine physische und/oder logische Trennung dar, die auf unterschiedlichen Anforderungen basiert. Als Erstes ist hier die organisatorische Struktur zu betrachten, die Themen wie Mandantentrennung, Administratoren-, Nutzer- und Automationszugriffe et cetera beinhaltet. Dazu erhält jede Workload-Domain ihr eigenes vCenter und so sind Sie innerhalb einer Workload-Domain in der Lage, Ressourcen und Services optimal zu konfigurieren und zu verwalten, um die spezifischen Bedürfnisse Ihrer Workloads zu erfüllen.

Der zweite wichtige Punkt in der Workload-Domain sind Sicherheitsrichtlinien. So ist in Bild 1 beispielsweise die Netzwerk trennung durch eine unabhängige NSX-Installation realisiert. Workload-Domains stellen in der Regel auch Sicherheitsbereiche dar, das heißt, häufig kommt hier ein dediziertes NSX oder auch zusätzlich physisch getrennte Netzwerke (Airgapping) zum Einsatz. Eine physische Trennung ist auch heute häufig noch das Mittel der Wahl für eine DMZ.

Die dritte und letzte Überlegung gilt den Performanceanforderungen. Hierbei liefern die Workload-Domains Flexibilität und Skalierbarkeit, da sich unterschiedliche Domains für unterschiedliche Workload-Typen optimieren lassen. So stellen beispielsweise VDI-Umgebungen ganz andere Anforderungen als klassische Serveranwendungen. Eine weitergehende Unterteilung von Workload-Domains in unterschiedliche Cluster erlaubt zudem eine zusätzliche Verfeinerung oder die weitergehende Trennung von Applikationen. Je genauer Sie die Anforderungen

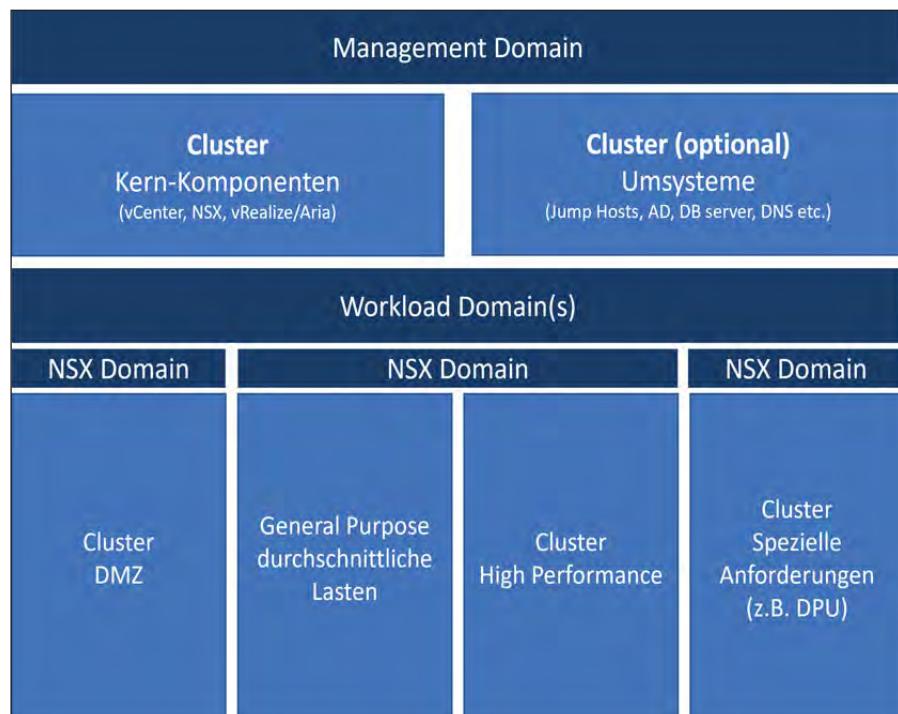


Bild 1: Der Aufbau einer vSphere-Umgebung mit physischer Trennung in Management- und mehrere Workload-Domains.

der abzubildenden Workloads (virtuelle Maschinen und Applikationen) kennen, desto besser lassen sich Workload-Domains und die zugehörigen Cluster auslegen, um den unterschiedlichen Anforderungen Rechnung zu tragen.

Verfügbarkeit der Workloads planen

In erster Näherung lässt sich sagen, dass in den meisten Umgebungen zwei Cluster-Typen sinnvoll sind: General-Purpose-Cluster für reguläre Anwendungen (Webserver, viele Applikationsserver) und High-Performance-Cluster für latenzsensitive Datenbanken, performancehungrige Applikationen oder besonders große virtuelle Maschinen. Darüber hinaus lässt sich diese Unterteilung natürlich beliebig weiter treiben – Cluster für Container, AI/ML mit GPU-Hardware, In-Memory-Datenbanken und so weiter.

Wir gehen später noch genauer auf das Cluster-Design ein, vorher ist jedoch noch die Frage nach der Resilienz der Gesamtumgebung von sehr hoher Bedeutung. Grundsätzlich können wir dabei zwei Architekturansätze unterscheiden, die auf der Anzahl der Rechenzentren basiert: Gibt es ein Rechenzentrum beziehungsweise Standort oder mehrere? Ein wich-

tiger Begriff aus dem Cloudkontext ist dabei die "Availability Zone" (AZ). Diese beschreibt in der Regel einen in sich geschlossenen Standort, wir können ihn also näherungsweise mit dem Begriff "Rechenzentrum" oder "Serverraum" gleichsetzen. AZs sind räumlich getrennt, aber üblicherweise innerhalb von Campus- oder Metro-Distanzen angesiedelt, um die Roundtrip-Latenz gering zu halten. Meist basieren diese Maximallatenzen auf den Anforderungen eines synchronen Storage-Zugriffs. Bei VMware vSAN liegt diese Latenz für eine gestreckte Installation beispielsweise bei 5 ms [2].

Für ein Maximum an Verfügbarkeit sind zwei Availability Zones der präferierte Ansatz. Selbst wenn erst eine AZ existiert oder aufgebaut wird, sollte deren Design idealerweise eine künftige Erweiterung auf eine zweite AZ berücksichtigen. Als Beispiel sei hier die Situation genannt, dass das Managementnetzwerk für ESXi und vCenter auf Layer 2 über beide AZs gestreckt werden muss (Bild 2).

Gemäß Best Practices sollten Sie die Management-Domain verlängern, sobald auch nur eine Workload-Domain gestreckt sein soll, damit sowohl Management als auch Workloads gleiche Verfüg-



Bild 2: Selbst wenn erst eine Availability Zone existiert oder aufgebaut wird, sollte deren Design idealerweise eine künftige zweite berücksichtigen, etwa zur Erweiterung des Managementnetzwerks.



Bild 3: Über zwei Availability Zones gestrekte Workload-Domains erfordern auch das Erweitern der Managementebene.



Bild 4: Ein Kombination von gestreckten und nicht-gestreckten Domains erlaubt verschiedene Verfügbarkeiten.

barkeit gewährleisten können (Bild 3). Der duale AZ-Ansatz gewährleistet dennoch hohe Flexibilität, denn nicht alles muss zwangsläufig gestreckt sein. Eine

Kombination aus gestreckten und nicht-gestreckten Domains (Bild 4) kann unterschiedliche Verfügbarkeits- und Kostenansätze adressieren.

Workload-Domains pro AZ anstelle von gestreckten Workload-Domains ergeben vor allem dann Sinn, wenn:

- Hochverfügbarkeit/Replikation auf Applikationsebene gewährleistet ist.
- Durchschnittliche Verfügbarkeit erforderlich ist.
- Test-/Entwicklungsumgebungen aufgesetzt werden.

Obwohl natürlich eine gestreckte Umgebung eine sehr hohe Verfügbarkeit gewährleistet, sollten Sie keinesfalls vergessen, dass ein Stretching auch doppelt so teuer ist, da Sie durch die zweite AZ zweimal so viel Hardware und Lizenzen benötigen, die nutzbare Kapazität aber dennoch nur der einer AZ entspricht. Cluster müssen Sie entsprechend über beide Availability Zones aufteilen und konfigurieren. Und natürlich muss die Speicherumgebung dieses – häufig auch als Metro-Storage-Clustering bezeichnet – ebenfalls unterstützen.

Produktiv können Workloads auf beiden Seiten einer gestreckten Umgebung laufen, bei einem Failover muss aber immer sichergestellt sein, dass Workloads auch auf der überlebenden Seite automatisch starten (können). Das heißt, es darf keine lokalen Abhängigkeiten wie CD- oder ISO-Mounts et cetera geben. Zudem müssen Sie die Regeln des VMware DRS (Distributed Resource Scheduling Cluster) wie Anti-Affinitäts- und Host-zu-VM-Richtlinien entsprechend konfigurieren und regelmäßig überprüfen.

Und was in der Praxis oft vergessen wird, ist die Tatsache, dass ein dualer AZ-Ansatz in der Regel nicht zwei, sondern drei Standorte beziehungsweise AZs benötigt. Denn für viele Komponenten wie zum Beispiel vSAN oder Datenbank-Clustering ist in einem 2-Standort Setup üblicherweise ein Quorum (oder ein Witness) erforderlich. Dieses sollte eben nicht in einer der beiden AZs, sondern idealerweise in einem dritten unabhängigen Standort laufen (Bild 5). Meist kann dieser deutlich kleiner ausfallen und auch deutlich höhere Latenzen verkraften, also ist hier mitunter auch die Platzierung in einer Public Cloud denkbar, sollte kein eigener dritter Standort vorhanden sein. Für VMware vSAN und

seine Witness-Appliance ist etwa eine Platzierung im Clouddienst "VMware Cloud on AWS" möglich [3].

Anforderungen des Cluster-Designs

Innerhalb einer Workload-Domain können ein bis mehrere Cluster existieren, wie Bild 6 veranschaulicht. Die Frage ist nun aber, wie viele Cluster Sie in einer Workload-Domain platzieren sollten. Das hängt von einigen Faktoren ab, die wir im Folgenden betrachten. Der erste Aspekt sind Sicherheitsanforderungen: Müssen Sie Workloads (VMs) physisch voneinander trennen, wird häufig der Cluster (inklusive der Storage-Anbindung) als Demarkationslinie gesehen, was logischerweise mehr Cluster erfordert. Ein Beispiel hierfür ist eine Compliance-Anforderung hinsichtlich PCI (Payment Card Industry), bei der Sie entsprechende VMs nicht mit denen anderer Nutzer oder Nicht-PCI-Workloads auf demselben Cluster und Storage mischen dürfen. Obwohl sich dies in einer virtualisierten Welt auch mittels sicherer logischer Trennung realisieren lässt, ist die physische Trennung noch ein häufig gesehenes Muster.

Die zweite Überlegung umfasst die Charakteristika Ihrer Workloads. Existieren verschiedene Performanceanforderungen und extrem unterschiedliche VM-Größen (zum Beispiel 16-GByte-VMs am einen und 6-TByte-VMs am anderen Ende der Größenskala), ergibt eine Trennung in separate Cluster durchaus Sinn im Hinblick auf die initiale Platzierung oder VMotion-Zeiten. Auch in Sachen Performanceanforderungen, die neben CPU und RAM auch das Netzwerk und den Storage sowie deren Latenzen umfassen, sollten Sie auf dedizierte Cluster setzen, um Performance-SLAs einzuhalten.

Einplanen sollten Sie zudem einen "Blast Radius" für Fehlkonfigurationen. Denn aus operativer Sicht müssen Sie früher oder später auch an Upgrades und Updates denken und dafür sind mehrere kleine statt eines großen Clusters hilfreich. Denn erstens ist im Fall eines Konfigurations- oder Treiberfehlers der Blast Radius auf einen von mehreren Clustern beschränkt und zweitens verwendet



Bild 5: Beispiel für ein sauberes Stretching inklusive der Quorum-Lokation.

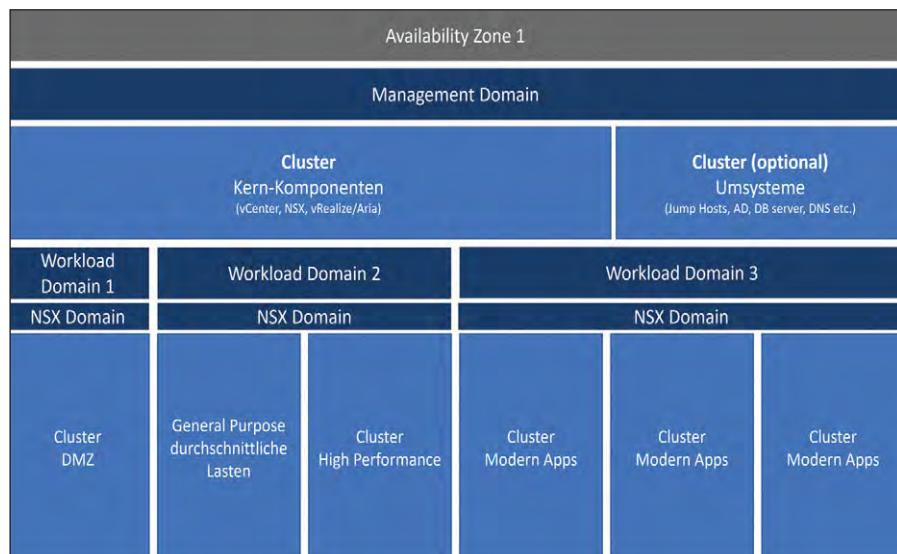


Bild 6: Ein Design mit mehreren Workload-Domains und einer unterschiedlichen Anzahl von Clustern.

vSphere 8 präferiert den vSphere Lifecycle Manager und Cluster-Images. Damit ist nicht mehr der einzelne Host, sondern der Cluster das kleinste Objekt für ein orchestriertes Update und mehrere kleinere Cluster bieten hier wiederum mehr Flexibilität.

Zwei weitere wichtige Aspekte beim Cluster-Design sind Kapazität und Kapazitätsverschnitt, die Auswirkungen auf die Effizienz der Umgebung haben. Die Grundidee hinter einem Cluster ist natürlich die Hochverfügbarkeit (realisiert über den vSphere-HA-Cluster). Dafür stellen Sie entweder dedizierte Failover-Hosts oder umgerechnet einen entsprechenden Prozentsatz der Clusterkapazität zur Verfügung.

Nehmen wir die typischen N+1- und N+2-Hochverfügbarkeitsansätze (also ein oder zwei Failover-Hosts beziehungsweise deren Kapazität) und betrachten verschiedene Clustergrößen mit jeweils N+1 und N+2

in der Tabelle 1 auf dieser Seite. Diese zeigt Ihnen, dass der Kapazitätsverschnitt durch Failover-Ressourcen bei kleinen Clustern besonders hoch ist. Im Bereich von 16 bis 32 Hosts liegen Sie mit ein oder zwei Failover-Hosts immer im Bereich um 90 Prozent der nutzbaren Kapazität. vSphere 8 erlaubt im Übrigen eine maximale Cluster-Größe von 96 Hosts [4]. Unabhängig von der Host-Größe (dazu später mehr) ergibt es bei größeren Clustern eventuell Sinn, die Redundanz auf N+3 oder N+4 zu erhöhen.

Wie die Tabelle zeigt, ist die nutzbare Kapazität in 64- und 96-Host-Clustern selbst mit N+4 immer noch jenseits der 90-Prozent-Marke. Die tatsächliche Cluster-Größe müssen Sie mit den oben genannten Aspekten von Fall zu Fall definieren. Häufig ergeben sich in der Praxis Cluster-Größen automatisch aus der Anzahl der Workloads und der zu Grunde liegenden Host-Größe.

vSphere-Clustergrößen und nutzbare Kapazitäten

Anzahl Hosts im Cluster	Abzug Failover-Hosts		Nutzbare Kapazität in Prozent	
	N+1	N+2	N+1	N+2
4	1	2	75,0	50,0
8	1	2	87,5	75,0
16	1	2	93,8	87,5
32	1	2	96,9	93,8
64	1	2	98,4	96,9
96	1	2	99,0	97,9
N+3		N+4	N+3	N+4
64	3	4	95,3	93,8
96	3	4	97,9	95,8

vSphere-Clustergrößen und nutzbare Kapazitäten

Daneben spielen jedoch weitere Faktoren beim Cluster-Design eine wichtige Rolle. Eine Grundregel lautet: Setzen Sie in einem Cluster immer gleiche oder gleichartige Hosts ein! Denken wir an eine Cloud, ist der Cluster die kleinste funktionale Einheit im Clouddesign und der einzelne Server die kleinste Austauscheinheit (FRU; Field Replaceable Unit). Im Fehlerfall wird also idealerweise der gesamte Server und nicht einzelne Komponenten ersetzt. Daher sollten Sie alle Server identisch ausstatten – nicht nur im Hinblick auf Netzwerkkarten, Smart-NICs, GPUs, Disks/NVMe und so weiter, sondern auch im Hinblick auf RAM und den Prozessor.

Denn eine saubere Verfügbarkeitskalkulation (N+1, N+2, ...) richtet sich entweder nach der standardisierten Host-Größe oder sollte in einem heterogenen Cluster den größten Host als Grundlage heranziehen. Neben einer komplizierten Failover-Kalkulation sind aber auch Schwierigkeiten beim Lastausgleich (DRS Cluster) in einem heterogen ausgestatteten Cluster prognostizierbar. Setzen Sie daher auf gleiche Host-Größen und -ausstattung.

Zuvor haben wir bereits den Lifecycle Manager und die Cluster-Images angeprochen. Ein einheitlicher Hardwaretyp im Cluster erleichtert auch das Lifecycle-Management, da Sie mit den Cluster-Images auch Treiber und Firmware ak-

tualisieren und dies selbstredend bei gleicher Hardwarebasis problemfrei funktioniert und damit den Betrieb Ihrer Umgebung massiv erleichtert.

Ein bekanntes Problem beim Hardweraustausch nach längerer Zeit ist, dass eventuell der Servertyp, aber nicht mehr die identische Hardware beim Hersteller verfügbar ist. Um diesen Fall abzudecken und einen neueren Server desselben Modells in einen bestehenden Cluster einzuführen, empfiehlt sich der EVC-Modus (Enhanced VMotion Compatibility). Diesen sollten Sie bereits beim Erstellen des Clusters einschalten. Das Pro und Contra von EVC können wir hier nicht ausführlich diskutieren – für Reparaturen und Austausch ist der Modus aber auf jeden Fall sinnvoll und auch nicht mit Nachteilen verbunden.

Ein weiterer Aspekt des Cluster-Sizing ist das Wachstum. In der Praxis bauen Sie ja keinen großen Cluster auf, auf dem kaum Workloads laufen, sondern der Bedarf nach Erweiterung der Umgebung ergibt sich meist im Laufe der Zeit. Bei Erweiterungen in kleinen Schritten gelten die Regeln aus dem vorhergehenden Abschnitt (Host-Größen, EVC-Modus). Daneben sollten Sie aber auch in der Gesamtumgebung dafür geplant haben mit freien IP-Adressen, Switchports und so weiter. Bei einem umfangreichen Ausbau ist der angesprochene Cloudansatz mit Erweiterung durch zusätzliche Cluster statt zusätzlicher Hosts für existierende Cluster durchaus sinnvoll. Cluster durchaus sinnvoll.

Überlegungen zur Cluster-Konfiguration

In der Regel sind die Hardware, das Cluster-Sizing und die Verfügbarkeitsüberlegung nur die halbe Miete für einen stabilen Betrieb. Die Konfiguration der Cluster und der beteiligten Hosts spielt eine genauso wichtige Rolle. Hier kommt mit vSphere 8.0 Update 1 ein neues Werkzeug zum Tragen: Configuration Profiles. Diese erlauben die Host-Konfiguration auf Cluster-Ebene mittels eines Desired-State-Ansatzes, der vielen schon aus dem Kubernetes-Umfeld geläufig ist.

Die gewünschte Konfiguration (Desired State) können Sie allen vSphere-Hosts in einem Cluster mit einer einzigen Operation zuordnen, wozu Sie einen Referenz-Host als Vorlage zum Einsatz bringen. Der Vorteil gegenüber den bereits aus vorherigen vSphere-Versionen bekannten Host Profiles ist, dass Configuration Profiles auf Cluster-Ebene ansetzen und für einen homogenen Cluster sorgen. Diese Profile haben das JSON-Format, was erlaubt, Sie einfach zu erstellen, zu editieren und zu verwalten. Configuration Profiles können die alten Host-Profiles ersetzen und es ist möglich, bereits bestehende Cluster entsprechend umzustellen.

Aus Betriebssicht ist ein Konfigurationsmanagement auf Cluster-Ebene deutlich zielführender als auf individueller Host-Ebene. Der einzige Wermutstropfen ist, dass Sie den Cluster dafür auf Image-basiertes Lifecycle-Management umstellen und die Baseline-basierte Verwaltung aufgeben müssen.

Strom sparen im Cluster

Die Themen CO2-Fußabdruck, Nachhaltigkeit und Energiesparen sind inzwischen auch in den Rechenzentren angekommen. IT-Verantwortliche können den Energieverbrauch einerseits vorausschauend durch ein vernünftiges, bedarfsgerechtes Sizing der Server und der Cluster steuern, zum anderen kommen natürlich Konfigurationen und Hardwarefähigkeiten auf Host-Ebene wie zum Beispiel Dynamic Power Mode, Speed Stepping und so weiter dazu.

Aber auch im Cluster kann DPM (Distributed Power Management) durchaus zum Energiesparen beitragen, vor allem dort, wo Lasten zyklisch von "High" zu "Low" laufen. DPM setzt einen DRS-Cluster voraus und erlaubt dann, dessen Energieaufnahme dadurch zu reduzieren, dass Hosts je nach Ressourcennutzung aus- und angeschaltet werden. Da ein Host für das Wiederanfahren Zeit braucht, ist dies allerdings kein Weg, quasi in Echtzeit auf Laständerungen zu reagieren. Aber wo beispielsweise an einem Wochenende in einem Virtual-Desktop-Cluster kaum Nutzer arbeiten, führt DPM durchaus zu Einsparungen.

Sizing von CPU und RAM des ESXi-Hosts

Unabhängig von der Größe des Clusters und wie viele davon Sie einsetzen, die Grundlage ist der ESXi-Host. Wie schon erwähnt, hängt die Cluster-Größe auch vom Sizing des einzelnen Hosts ab. Kommen eher 1-U-Server, also Pizzaboxen, zum Einsatz, ist die Kapazität in Sachen CPU, RAM und Erweiterungs-Slots des einzelnen Servers überschaubar und der Cluster fällt wahrscheinlich entsprechend größer aus. Natürlich hängt auch das Server-Sizing vom Workload, also der entsprechenden VM ab, doch die Grundfrage ist sicherlich, ob Sie lieber einen großen oder einen kleinen Server einsetzen.

Hier spielen diverse Faktoren eine Rolle, die wir uns im Detail ansehen wollen. Entscheidend ist das Workload-Profil der VM, deren Größe unser erster Punkt ist. Nehmen wir zwei Extreme als Beispiel: eine große Oracle- oder SAP-HANA-Datenbank und einen Container-Host. Eine SAP-HANA-VM kann leicht eine Größe von 3 bis 6 TByte RAM und 192 bis 384 vCPUs erreichen. Das schreit natürlich nach einem ESXi-Host mit acht Sockets und 6 bis 12 TByte RAM – dennoch finden nur maximal zwei VMs auf dem Host Platz, zumal es kein Oversubscription geben darf.

Eine Container-Host-VM auf der anderen Seite mag 16, 32 oder 64 GByte RAM und 4 bis 16 vCPUs benötigen und Oversubscription der Ressourcen auf dem ESXi-Host stellt meist auch kein Problem dar.

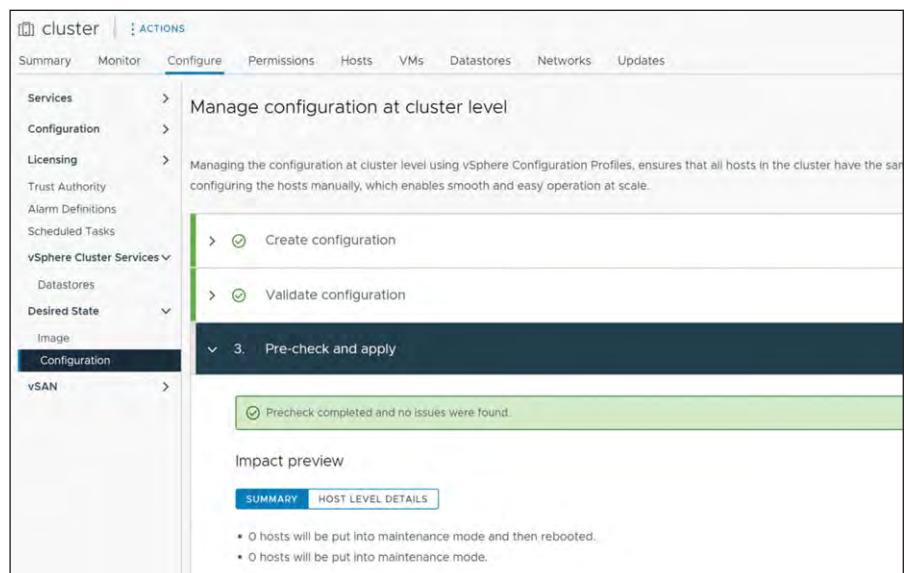


Bild 7: Configuration Profiles sorgen auf Cluster-Ebene für Homogenität.

Von der kleinen "4 vCPU, 16 GByte RAM"-Variante passen bei einem konservativen 3:1-RAM-Oversubscription-Verhältnis auf unseren Server mit acht Sockets, 384 HyperThreads und 12 TByte RAM sage und schreibe 768 dieser VMs. Das ist schon sehr nahe an der maximal supporteten Obergrenze von 1024 VMs pro Host. Da vSphere 8 aber "nur" 8000 VMs pro Cluster unterstützt [4], ist ein 12-Node-Cluster bei N+2-Redundanz bereits das Ende der Fahnenstange.

Auch der Blast-Radius eines Servers im Fehlerfall ist ein entscheidendes Kriterium. Nehmen wir wieder das zweite Beispiel: Fällt dieser 12-TByte-Server mit 768 VMs aus, stehen 768 Applikationen oder Applikationsbestandteile still und die VMs müssen natürlich an anderer Stelle und dazu noch gleichzeitig neu gestartet werden – was eventuell auch zu Problemen im Storage führen kann (Stichwort "Bootstorm"). Für diese Workloads sind kleinere ESXi-Hosts also deutlich sinnvoller.

Der nächste zu betrachtende Aspekt ist die VM-Performance, die sich in der Regel in drei Teiltbereichen festmachen lässt und entsprechende Anforderungen an Host und umgebende Infrastruktur stellt. In Sachen CPU ist die erste Frage, die sich häufig stellt: "Wie viele Cores und welche Taktfrequenz?" Neben der Kostenfrage müssen Sie hier beachten, wie die Workloads aussehen. Profitiert die Applikation eher von einer schnelleren CPU – also

höherer Taktfrequenz – oder beherrscht sie sauberes Multithreading und profitiert von einer größeren vCPU-Zahl und damit mehr Cores.

Das Thema Cores ist auch direkt mit dem Thema Hyperthreading verbunden. Generell ist dies eine gute Idee und Sie sollten es anschalten, denn es erlaubt dem CPU-Scheduler im VMkernel des ESXi-Servers mehr Auswahl beim Verteilen der CPU-Requests der VMs und hält die Performance oben. Ist Hyperthreading auf dem Server aktiviert, zeigt er natürlich doppelt so viele logische CPUs wie Cores physikalisch vorhanden sind. Für das Sizing ist aber keinesfalls anzunehmen, dass Hyperthreads die Performance verdoppeln. Der Performancegewinn liegt erfahrungsgemäß zwischen fünf und maximal 25 Prozent und das ist wiederum abhängig vom Applikationstyp, NUMA Alignment, IO-Last et cetera.

Daneben müssen Sie auch den Arbeitsspeicher des Hosts planen. Das RAM ist für viele Applikationen ein entscheidender Faktor, wobei die Größe wichtiger ist als die Taktfrequenzen. Es sei jedoch angemerkt, dass je nach RAM-Ausbau und Mainboard-/Chipset-Typ (alle Slots, alle Channels oder nur zwei Drittel oder gar nur die Hälfte bestückt) durchaus Performanceunterschiede durch Heruntertaktungen des RAM auftreten können. Hier hilft die Dokumentation des Hardwareherstellers meist weiter.

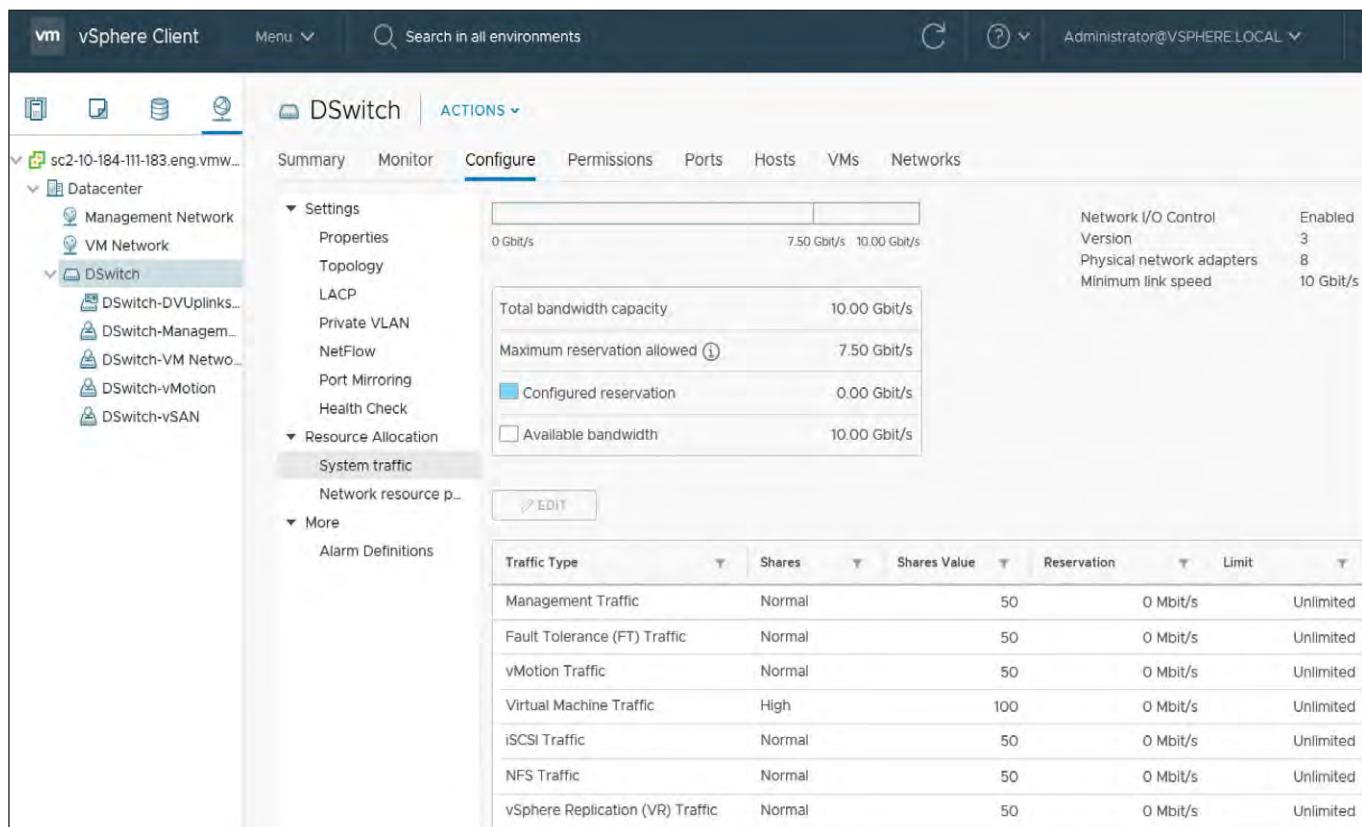


Bild 8: Die Kontrolle des Netzwerk-IO kann pro Traffic-Typ erfolgen.

Eine ausgewogene CPU-zu-Memory-Ratio ist sehr sinnvoll. Ein wertvoller Ansatz ist hier derjenige von Hyperscalern wie Amazon oder Google, nämlich die VMs in drei grobe Kategorien zu unterteilen: "Memory-lastig", "CPU-lastig" und "Ausgeglichen". Dementsprechend lassen sich Hosts dann aufbauen und natürlich in verschiedene Cluster aufteilen.

Auslegen der Netzanbindung

Spielt bei Ihnen allerdings Performance die größte Rolle und CPU-, Memory-, Storage- und Netzwerkperformance-Schwankungen sind nicht tolerierbar, bietet sich noch das Vehikel der Ressourcenreservierungen an. Damit lassen sich für CPU und RAM physische Ressourcen garantieren. Natürlich verschwenden Sie damit wertvolle Kapazität, jedoch bleibt das Applikationsverhalten so immer gleich und Sie müssen nicht mit RAM- oder CPU-Engpässen rechnen.

Für Netzwerkressourcen bedienen Sie sich hier der "Network-IO-Control"-Funktionalität (NIOC) und können pro Traffic-Typ ebenfalls Reservierungen und Limits anwenden, um beispielsweise VM Traffic nie unter ein bestimmtes Maß fallen zu

lassen [5]. Für eine noch granularere Zuordnung zur einzelnen VM, oder genauer gesagt zum einzelnen virtuellen Netzwerkadapter, bietet vSphere 8 die Funktionen des Vorgängers (NIOC v3) an, mit denen Sie pro virtuellem NIC die Bandbreite mittels Shares, Reservierungen und Limits steuern.

All das ergibt natürlich nur Sinn, wenn der Host über genug Netzwerkbandbreite verfügt. Hierzu bieten sich 25-, 40- oder 100-GBit-Adapter an. Meist reichen zwei 25-GBit-NICs völlig aus – selbst beim Einsatz von VMware vSAN und dem assoziierten Replikationsverkehr. Die Erfahrung zeigt, dass Sie zwei 100-GBit-Adapater nur selten ausreizen können, selbst in Kombination verschiedenster Traffic-Typen (VM, iSCSI/NFS, VMotion, Provisioning et cetera). Andererseits bieten zwei 100-GBit-Anschlüsse genug Bandbreite, um auf NIOC zu verzichten und operationellen Aufwand einzusparen. Also eine klassische Abwägung zwischen CAPEX und OPEX.

Storage-Planung

Beim Thema Storage müssen Sie auch wieder mehrere grundsätzliche Fragen

beantworten. Die erste davon ist, ob Sie diesen traditionell oder Policy-basiert aufbauen. Mit "traditionell" ist hier der Ansatz gemeint Datastores via iSCSI, FC oder FCoE, LUNs oder NFS-Shares als Datastore bereitzustellen, auf denen die VM-Daten liegen. In der Regel bestimmt das (externe) Array und die Art der Anbindung (LAN oder FibreChannel, geteilte oder dedizierte Infrastruktur und so weiter) die Performance sowie Verfügbarkeit auf LUN- oder Share-Ebene. Der Wechsel auf eine andere Performance- oder Verfügbarkeitsklasse ist in der Regel nur durch Verschieben auf einen anderen Datastore mittels Storage-vMotion möglich.

Policy-basierter (oder Software-defined) Storage ist für vSphere in Form von vSAN oder VMware vVols verfügbar. Hier können Sie Policy-Änderungen auf Objekt-Ebene wie etwa einer virtuellen Disk im laufenden Betrieb und ohne Verschieben der VM durchführen. Dies ist vor allem in hochautomatisierten Umgebungen sinnvoll beziehungsweise wenn zu erwarten ist, dass sich im VM-Lifecycle die Verfügbarkeits- oder Performanceanforderungen ändern.

Nun gilt es noch zu entscheiden, ob Sie auf HCI (Hyper Converged Infrastructure) oder externen Storage setzen wollen. Wir haben eben schon vSAN erwähnt, das zur Klasse des Hyper Converged Storage gehört. Das Grundprinzip lautet, internen Speicher des Servers (präferiert NVMe oder SSD) zu nutzen und über die Server im Cluster hinweg zu aggregieren und hochverfügbar zu gestalten. Der Vorteil liegt auf der operativen Seite, denn so konfigurieren und verwalten Sie den Speicher als Teil der vSphere-Plattform.

Der Nachteil ist, dass die SSDs oder NVMes im Server Platz finden müssen und Sie für vSAN eine eigene Lizenz benötigen. Da die Replikation der Storage-Objekte und der Zugriff über das vorhandene Ethernet stattfindet, müssen Sie die Bandbreite einkalkulieren und diesen Traffic sollten Sie – wie alle anderen Arten auch – unbedingt über ein eigenes VLAN sauber separieren.

Kommen wir zum traditionellen Storage, der mit zentralisierten Arrays nach wie vor sehr populär ist. Viele Umgebungen nutzen bekannte Technologien wie File- oder Block-Speicher schon sehr lange, die Admins verfügen über eine entsprechende Erfahrung. Vor allem FibreChannel (FC) steht nach wie vor hoch im Kurs, wenn es um Performance und Verfügbarkeit geht. Die Tatsache, dass FC über eine dedizierte Infrastruktur und nicht das gemeinsame Ethernet läuft, gibt vor allem bei kritischen Applikationen zusätzliche Sicherheit.

Der Nachteil liegt im zusätzlichen Verwaltungsaufwand, denn Sie müssen eine separate Infrastruktur sowie Software für Konfiguration und Replikation betreiben. Auch tiefes Know-how ist erforderlich, denn ein Problem im Array oder ein Konfigurationsfehler lässt eventuell einen ganzen vSphere-Cluster ausfallen.

Verbunden mit traditionellem Storage ist natürlich auch die Frage nach dem Protokoll. Zur Wahl stehen IP basierte Protokolle wie iSCSI und NFS oder das erwähnte FibreChannel. Die Wahl, ob iSCSI oder NFS, ist häufig eine Frage

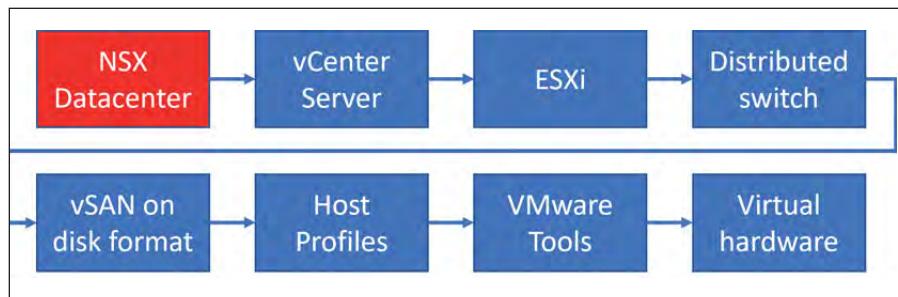


Bild 9: Die richtige Reihenfolge für das vSphere-8-Upgrade.

der persönlichen Erfahrungen, doch dem Irrglauben, dass iSCSI schneller ist als NFS, müssen wir an dieser Stelle widersprechen. Generell müssen Sie allerdings beachten, dass iSCSI im Kontext einiger Applikationen wie etwa SAP HANA nicht supportet ist. Hier sollten Sie sich vorher beim Softwarehersteller erkundigen.

Grüne Wiese vs Upgrade

Natürlich sind die hier erwähnten Prinzipien am besten in einer Greenfield-Umgebung, also einer komplett neuen Installation, umzusetzen. Jedoch schadet es nichts, wenn Sie den ein oder anderen Aspekt wie Workload-Domains, Cluster-Aufteilung oder die Umstellung auf Cluster Configuration Profiles auch im Rahmen eines Upgrades überdenken und anpassen.

Der wichtigste Aspekt bei Neuinstallation und Upgrade ist, dass die verwendete Hardware mit vSphere 8 U1 kompatibel ist, was Sie über die VMware Hardware Compatibility List [6] einfach prüfen können. Für ein Upgrade von vSphere 6.7 oder 7.x sieht es etwas anders aus [7], denn hierbei ist die Reihenfolge der Komponenten im Upgrade festgelegt und folgt der Reihenfolge in Bild 9.

Im Rahmen der vSphere+-Lizenzierung findet sich für das Update noch ein verstecktes Highlight. Ist vSphere+ (also die cloudverbundene Variante der Virtualisierungsplattform) aktiv und ein Cloud-Proxy in der lokalen Umgebung installiert, dann können Sie mit vSphere+ und vSphere 8 U1 lokal das vCenter-Update mittels eines einfachen und stabilen "One-Click"-Vorgangs durchführen.

Hier noch ein wichtiger Hinweis: Sollte in der bestehenden Umgebung NSX-T

beziehungsweise NSX Datacenter in Version 3.x im Einsatz sein, so müssen Sie NSX zunächst auf die Version 4.x (präferiert 4.1.x) migrieren, bevor Sie mit dem vSphere-8-Update starten können. Denn es besteht keine Kompatibilität zu NSX unter Version 4.

Fazit

Sie sollten beim Design Ihrer vSphere-8-Umgebung aktuellen Standards folgen. Der Aufbau von Domains ist aus Sicherheits-, Verwaltbarkeits- und Performanceaspekten sinnvoll – nicht nur in VMware Cloud Foundation, sondern in jeder vSphere-Infrastruktur. Die Aufteilung der Umgebung in Domains und Cluster erlaubt eine saubere Workload-Zuordnung, steigert die Effizienz und sorgt auch für Energieeinsparungen. Das Update 1 bringt hier zusätzliche, gute Ergänzungen wie Cluster Configuration Profiles ins Spiel und trägt damit zu mehr Effizienz und zur betrieblichen Vereinfachung bei. (jp) IT

Link-Codes

- [1] **VMware Validated Designs**
(nicht mehr aktuell und offiziell supported)
js147
- [2] **Leitfaden vSAN Stretched Cluster**
n9p11
- [3] **vSAN Witness Appliance in VMC on AWS**
n9p12
- [4] **Maximalkonfiguration von vSphere 8**
n9p13
- [5] **Distributed Power Management**
n9p14
- [6] **VMware Hardware Compatibility List**
js121
- [7] **Leitfaden zum vSphere-8-Upgrade**
n9p15