



DER ULTIMATIVE LEITFADEN FÜR  
**VDI auf einer  
hyperkonvergenten  
Infrastruktur**

**NUTANIX™**

**Brian Suhr** [Autor] verfügt über mehr als zwei Jahrzehnte IT-Erfahrung in den Bereichen Design, Implementierung und Administration von Unternehmensinfrastrukturen. Er hat architektonisches und technisches Fachwissen in einer Vielzahl von Virtualisierungs-, Rechenzentrums- und Cloud-basierten Projekten eingebracht und dabei mit technischen High-Performance-Teams in global agierenden Umgebungen zusammengearbeitet. Als unabhängiger Autor der Blogs DataCenterZombie und VirtualizeTips konzentriert sich Brian auf die Erstellung von Inhalten, die sich um Virtualisierung, Automatisierung, Infrastruktur und Weiterentwicklung von Produkten und Services drehen, die der Technologie-Community zugutekommen. Folgen Sie Brian auf Twitter: [@bsuhr](#)

**Sachin Chheda** [Editor] ist Direktor für Lösungen und vertikales Marketing bei Nutanix. Er ist seit langem in der Informationstechnologie tätig, und hatte Ingenieur-, Management- und Marketingpositionen in den innovativsten Unternehmen der Branche inne. Bei diesen entwickelte er Produkte und Lösungen, die einige der größten und zukunftsweisenden Unternehmen antreiben. Folgen Sie Sachin auf Twitter: [@StorSC](#)

Offizieller Release „1.0“

Copyright 2017 Nutanix, Inc. Alle Rechte vorbehalten. Dieses Produkt ist durch US-amerikanische und internationale Urheberrechtsgesetze und Gesetze zum Schutz geistigen Eigentums geschützt. Nutanix ist eine Marke von Nutanix, Inc. in den Vereinigten Staaten und/oder anderen Ländern. Alle anderen hier genannten Marken und Namen sind möglicherweise Marken ihrer jeweiligen Inhaber.

Autor	2
Über dieses Buch	4
Einführung	5

ARCHITEKTURGRUNDSÄTZE

Einstiegspunkt	6
Skalierbarkeit	7
Leistung	8
Kapazität	9
Überwachung	10
Bausteine	12

INFRASTRUKTUR-ALTERNATIVEN

Build Your Own (BYO)	14
Konvergente Infrastruktur	16
Hyperkonvergente Infrastruktur	18
Speicherungsanforderungen	20

SPEICHERUNGSARTEN

Herkömmliche Architekturen	22
All Flash	23
Hybrid Flash	23

Größenberechnung	24
Design von Virtualisierungsclustern	28
Erste Schritte	30

ÜBER NUTANIX	32
--------------	----

## ÜBER DIESES BUCH

Dieses Buch befasst sich mit dem Infrastrukturdesign für VDI und End-User-Computing (EUC)-Umgebungen. Der Inhalt dieses Buches wurde aus dem auf die Infrastruktur ausgerichteten Kapitel des geplanten Buches mit dem Titel „Architecting and Designing End-User Computing Solutions“ übernommen.

## EINFÜHRUNG

Nach der Auswahl der richtigen Strategie und des richtigen Softwareanbieters für die Bereitstellung von EUC-Services und -Anwendungen stellt die Wahl der Infrastruktur die nächste große Herausforderung für Anwendungs- und Desktop-Virtualisierungsprojekte dar.

Die Rechen- und Speicherinfrastruktur bildet die Grundlage, auf der man Services aufbaut. Ähnlich wie bei Strom und Wasser verlassen wir uns darauf, und Strom und Wasser sollte nur fließen, wenn wir den Wasserhahn aufdrehen oder den Schalter betätigen.

Ohne eine dem Design zugrundeliegende stabile, hochverfügbare und leistungsstarke Infrastruktur wird die IT während der Implementierungs- und Betriebsphase Ihres EUC-Projekts vor eine Vielzahl von Herausforderungen gestellt. Dies verstärkt die Tatsache, dass die Infrastruktur sehr wichtig ist, aber einen großen Teil der Zeit der IT für die Infrastruktur zu verwenden ist wie ein schwarzes Loch. Architekten und Ingenieure müssen sich auf die Bereitstellung der EUC-Services und -Anwendungen konzentrieren und sollten nicht Zeit mit Reparaturen verschwenden.

Es gibt mehrere wichtige Faktoren, die im Rahmen des Entwicklungsprozesses der EUC-Infrastruktur berücksichtigt werden sollten. Wenn man diese Faktoren zusammen mit den Anforderungen des Unternehmens nutzt, ist man besser in der Lage, einzuschätzen, was die Architekturalternativen sind. Die folgenden Faktoren sollten bei der Bewertung von Architekturalternativen und Anbieteroptionen für EUC-Projekte berücksichtigt werden:

- Einstiegspunkt
- Skalierbarkeit
- Leistung
- Überwachung
- Kapazität

### EINSTIEGSPUNKT

Der Einstiegs- oder Ausgangspunkt für die Infrastruktur kann oft über Wohl oder Wehe eines Projekts entscheiden. Hier wird festgelegt, wie viel Infrastruktur und Kosten ein Unternehmen benötigt, um die Anwendungs-/Desktopvirtualisierung und Bereitstellung auf der Grundlage unterschiedlicher Ausgangsgrößen zu starten.

Wenn im Rahmen des Projekts geplant ist, nach vollständiger Implementierung 10.000 Benutzer zu erreichen und in der Anfangsphase mit 5.000 Benutzern gestartet wird, wird das Unternehmen von den Vorabkosten wahrscheinlich weniger schockiert sein. Der Grund hierfür ist, dass je nach Art der gewählten Infrastruktur die Kosten pro Benutzer möglicherweise erst dann sinnvoll sind, wenn Sie einige tausend Benutzer zur Verfügung haben.

Umgekehrtes gilt, wenn ein Unternehmen 10.000 Benutzer bereitstellt, aber nur mit 500 Benutzern beginnen und diese Anzahl über die Projektlaufzeit gleichmäßig erhöhen will. Dann wird das Unternehmen die Kosten für den ersten Infrastrukturausbau in dieser Größenordnung genauer untersuchen, anstatt einen größeren ersten Schritt zu unternehmen. Die Kosten pro Benutzer bei dieser Größe können mit zunehmender Skalierung der Umgebung konstant bleiben, oder sie können am Anfang aufgrund größerer Ausgaben für die Anlaufinfrastruktur wirklich ungünstig aussehen.

Während die Kosten pro Benutzer als Faktor zur Bestimmung Ihrer Infrastrukturkosten als unklar und fast irrelevant zur Bestimmung Ihrer Infrastrukturkosten angesehen werden können, wird man Sie dazu wahrscheinlich befragen, wenn Sie versuchen, das Projekt im Unternehmen zu verkaufen oder Ihre Infrastrukturauswahl gegenüber der Führung zu rechtfertigen. Wenn Sie sich für eine Alternative mit höheren Vorabkosten entscheiden, müssen Sie darauf vorbereitet sein, die Details zu erklären. Bewerten Sie Lösungen, von denen Sie glauben, dass sie besser für Ihre Umgebung geeignet sind. Andernfalls sollten Sie bereit sein, die Entscheidung darüber zu argumentieren, wie sich die Kosten entwickeln werden. Eine Beispiel dieser beiden Szenarien ist in Abbildung 1 dargestellt.

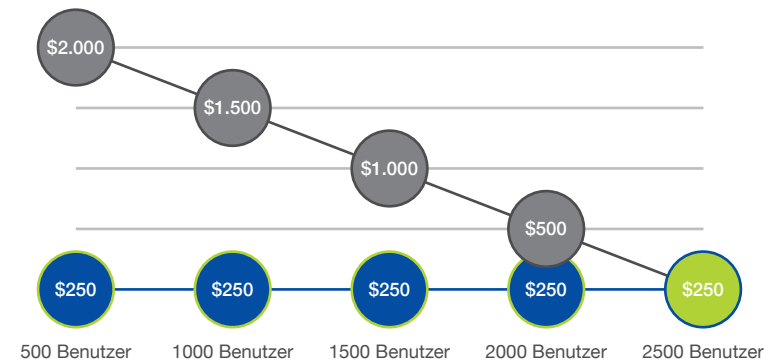


Abbildung 1  
Einstiegspunkte pro Desktop

### SKALIERBARKEIT

Die Skalierbarkeit der Architektur ist ein wichtiger Faktor bei der Beurteilung der Projektdurchführbarkeit. Ein Architekt muss die Ausgangsgrößenoptionen für die verschiedenen Alternativen verstehen; dies führt zurück zum Einstiegspunktthema, das gerade behandelt wurde. Erlaubt die Alternative problemlos, dass das Design bei Bedarf mit einer kleineren Größe beginnt, oder muss ein Unternehmen mehr Infrastruktur als notwendig erwerben, um die Ausgangsgröße eines Projekts zu erfüllen - und ist nicht in der Lage, alle Ressourcen zu nutzen, bis das Projekt hineinwächst?

Abgesehen davon, wie klein die Alternative starten kann, ist die Betrachtung der Skalierbarkeit der Alternative ebenso wichtig. Wenn der Wunsch besteht, mit 500 Anwendern zu beginnen und dennoch auf 10.000 skalieren zu können, wie wird die Alternative an beiden Enden dieses Spektrums aussehen? Wird das Unternehmen mit den Mindest- oder Maximalwerten zufrieden sein - oder mit beiden?

Das Thema Skalierbarkeit ist nicht nur eine Speicherdiskussion. Es betrifft auch Rechenleistung, Networking und möglicherweise auch andere Ebenen innerhalb des Designs. Wenn Anpassungen an der Konfiguration der Computing-Layer vorgenommen werden, um eine geringere Dichte an VMs pro Host-Server zu erreichen, wie wirkt sich dies auf verschiedene Designentscheidungen bei der Skalierung aus? Ein Beispiel wäre, wenn das anfängliche Host-Design mit 128 GB Speicher pro Host beginnt und die endgültige Auswahl 256 GB oder mehr beträgt. In diesem Fall muss sichergestellt sein, dass DIMMs der richtigen Größe verwendet werden, damit die Konfiguration in Zukunft skaliert werden kann. Wenn im Vorfeld falsche Entscheidungen getroffen werden, um Kosten zu sparen, wirkt sich dies aufgrund von Einschränkungen auf die Dichte aus oder kostet langfristig mehr aufgrund von DIMMs, die nicht mehr verwendet werden konnten.

Der Architekt sollte sich darauf konzentrieren, wie die Lösung klein anfangen und bis zum größten Punkt skaliert werden kann. Man kann jedoch auch nicht alle Punkte dazwischen ignorieren, denn je nachdem, wie man die Bereitstellung skaliert, kann es zwischen Start und Ziel viele Skalierungspunkte geben. Es ist ideal, nach einer Lösung zu suchen, die es dem Design ermöglicht, die Anzahl der durch das Projekt identifizierten Benutzerzahlen in Stufen leicht zu skalieren, ohne dabei die Bereitstellungszeit und die Kapazitäten zu überschreiten. Die ideale Skalierungsgröße für ein Projekt kann in Stufen von 100-200 Benutzern liegen. Wenn aber die gewählte Architekturalternative größer als diese Stufen skaliert, sollten Sie verstehen, wie sich dies auf die Kosten und die Bereitstellung auswirkt.

## LEISTUNG

Die EUC-Leistung, gemessen an der Endbenutzererfahrung, wird immer genau betrachtet. Ihre gewählte Architektur muss in der Lage sein, die Anforderungen in jeder Phase des Projekts zu erfüllen. Dies kann mit einigen Alternativen ein schwieriges Unterfangen sein. Wenn man eine Lösung auf ein Minimum herunterskaliert, um die Anforderungen in Bezug auf die Zahl der Anfangsanwender zu erfüllen, kann dies zu Leistungseinbußen führen, wenn die Lösung nicht linear skalierbar ist. Architekten möchten keine Kompromisse in der Architektur eingehen, um diesen kleinen Ausgangspunkt zu erreichen, der die gesamten maximalen Leistungsoptionen einer Lösung beeinträchtigen kann. Wenn Sie im Vorfeld Zeit damit verbringen, die richtige Entscheidung zu treffen, können Sie später Probleme vermeiden.

Ein Design einer EUC-Lösung wird typischerweise viele verschiedene Leistungsanforderungen haben. Wählen Sie eine Architekturalternative, die flexibel genug ist, um alle Leistungsanforderungen innerhalb einer einzigen Lösung zu erfüllen. Unabhängig davon, ob das Design mehrere Arten von EUC-Services bereitstellt oder sich nur auf Anwendungs- und Desktop-Virtualisierung konzentriert, müssen mehrere Leistungsanforderungen berücksichtigt werden. Das Wissen darüber, wie jede Alternative in der Lage sein wird, individuelle Leistungsanforderungen zu erfüllen oder nicht, wird Ihren Bewertungs- und Designprozess erheblich beeinflussen.

## KAPAZITÄT

Die Kapazitätsdiskussion ähnelt derjenigen der Leistung. Innerhalb der EUC-Designs gibt es eine Reihe von unterschiedlichen Kapazitätsanforderungen, die erfüllt werden müssen. Die Lösung erfordert den Einsatz von Server-VMs, Desktop-VMs, Anwendungen, Benutzerprofilen und Benutzerdaten für diese Art von Architektur. Jede Schicht innerhalb des Designs kann sehr unterschiedliche Kapazitätsanforderungen haben. Einige verwenden große Datenmengen, die sich in der Regel gut deduplizieren lassen. Andere Bereiche, wie Benutzerprofile und Daten, bestehen pro Benutzer aus kleineren Mengen an komprimierbaren Daten, die jedoch mit Tausenden von Anwendern multipliziert werden und am Ende einen großen Umfang aufweisen.

Ein größeres Problem in den vergangenen Jahren war der Einkauf von zu viel oder zu wenig Kapazität, während versucht wurde, die geforderten Leistungsniveaus zu erreichen. Schauen Sie sich während der Designphase die Architekturalternativen genau an, um zu sehen, wie sie die erforderliche Kapazität bereitstellen und gleichzeitig auch die Mindestleistungsanforderungen erfüllen. Die Alternative sollte nicht 2-3x oder mehr Kapazität bereitstellen, um die Anforderungen an die Speicherleistung zu erfüllen, oder eine signifikante zusätzliche Leistung zur Erfüllung der Kapazitätsanforderungen erbringen. Die ideale Lösung ist eine, die genügend Flexibilität bietet, um Leistung und Kapazität mit ähnlichen Raten zu skalieren, damit sich beide nicht zu weit voneinander entfernen.

In der Vergangenheit löste dieses Thema viele Debatten und Probleme aus. Viele Unternehmen gerieten in Leistungs- und Kapazitätsplanungsprobleme, weil sie die Kapazität schneller skaliert haben als die Leistung. Nur weil die Lösung 5 TB freien Speicherplatz hat, bedeutet das nicht, dass sie um weitere 500 Benutzer skalierbar ist. Dieses Szenario kann dazu führen, dass die Leistung stark beeinträchtigt wird. Administratoren und IT-Führungskräfte, die kein fundiertes Verständnis dafür haben, wie die Skalierbarkeit der Lösung aussieht, können in diese Falle geraten.

## ÜBERWACHUNG

Die Überwachung ist sehr wichtig und wird oft unterschätzt. Wenn es um die Überwachung der Infrastruktur in einer EUC-Umgebung geht, konzentrieren sich die Administratoren in der Regel auf den Leistungsaspekt. Sie brauchen die Fähigkeit zu verstehen, was normal ist und wann es ein aktives Problem gibt.

Die Handhabung des Monitorings sollte einfach sein, aber dennoch eine Vielzahl von Detailinformationen bieten. Dies ist bei vielen Herstellern nicht der Fall, daher sollte man sich genau ansehen, wie die Überwachungserfahrung bei jeder Alternative aussieht.

Eine weitere Anforderung ist die Möglichkeit der Performance-Überwachung auf der Ebene der virtuellen Maschine. Leider kann die Mehrheit der Infrastrukturanbieter diese Transparenz in der Virtualisierungsumgebung immer noch nicht bieten. Die Möglichkeit, sich schnell die Speicherschicht anzusehen und festzustellen, ob das Problem mit der Speicherleistung auf globaler Ebene besteht oder ob es auf einen Host, eine Gruppe von VMs oder nur eine einzelne VM isoliert ist, ist nicht länger nur eine Option.

Durch die Verwaltung der Speicherleistung auf VM-Ebene kann man einen ähnlichen Ansatz für die Verwaltung der CPU- und Speicherleistung einer VM auf Host-Ebene anwenden. Administratoren müssen wissen, ob eine VM vorübergehend zusätzliche Leistung in Anspruch nimmt oder ob sie ein regelmäßiger Verbraucher mit mehr Speicherleistung ist als typische Benutzer. Dadurch kann man verstehen, wann es eine Spitze gibt und wann man etwas tiefer gehen muss, um das Problem zu identifizieren.

Ein Baustein ist eine vordefinierte Gruppe von Infrastrukturen, die zu einer bestimmten Menge von Ressourcen oder Anzahl von Benutzern zugeordnet werden. Dieser Ansatz ist eine der besten Methoden, um das End-User-Computing auf das Infrastrukturdesign abzustimmen.

Mit diesem Ansatz kann man eine Architektur entwickeln, die ein vorhersagbares Kosten-, Leistungs- und Kapazitätsskalierungsmodell bietet. Wählen Sie bei der Bestimmung der Bausteingröße, welche Stufen Sie zur Skalierung der Anwender benötigen und wie die Infrastrukturauswahl die Wahlmöglichkeiten berücksichtigen kann. Zum Beispiel kann man Anwender in Schritten von 50 bis 100 Anwendern skalieren, doch die Infrastrukturauswahl skaliert in diesen kleinen Schritten nicht sonderlich gut. Dies kann dazu führen, dass das Design in größeren Schritten von 500 oder 1.000 Anwendern skaliert werden muss. Wenn die Wahl der Infrastruktur in großen Blöcken skaliert wird, kann man sich dafür entscheiden, sie so zu dimensionieren, dass sie ineinandergreift, oder einfach die Tatsache akzeptieren, dass die Infrastrukturkosten nicht in der gleichen Weise skaliert werden, wie es die Benutzer-Einsatzblöcke tun. Das bedeutet einfach, dass das Unternehmen die Infrastruktur in Blöcken von 1.000 Anwendern einkauft und nur in Gruppen von 50 bis 100 Anwendern bereitstellt.

Dadurch erscheinen die Kosten für virtuelle Desktops oder Anwender-Sessions teuer, wenn ein großer Block gekauft, jedoch eine geringe Anzahl von Anwendern implementiert wird. Dies gleicht sich aus, wenn die Organisation alle geplanten Anwender implementiert.

Architekturen im Bausteinstil sind in jedem Designprojekt hilfreich. EUC-Bereitstellungen haben jedoch immer gemeinsame Anwendergruppen und Anwendungsfälle, die ähnliche Merkmale aufweisen und in Gruppen bereitgestellt werden. Um mit dem Beispiel einer Blockgröße von 100 Anwendern fortzufahren, kann man durch das Verständnis der Ressourcenanforderungen von 100 Anwendern sicherstellen, dass der Infrastrukturblock in der Lage ist, alles bereitzustellen, was diese Anwender benötigen.

Wenn jeder Benutzer 15 IOPS im stationären Zustand und 30 GB Speicherkapazität sowie 2 GB Speicher und 200 MHz CPU benötigt, weiß der Architekt, dass die Bausteine 1.500 IOPS, 3 TB Kapazität, 200 GB Speicher und 20 GHz CPU bereitstellen müssen. Der Architekt kann die Bausteine so gestalten, dass sie zusätzliche Ressourcen enthalten, aber keiner von ihnen darf unter diesen Werten liegen. Wir wollen auch nicht verschwenderisch zu viele Extras in jeden Block aufnehmen, die wir nicht nutzen können.

Mit diesem Ansatz und der Granularität des Designs kann man die Umgebung nun in kleineren Gruppen von 50-100 Benutzern skalieren. Dies ermöglicht einen langsamen und stetigen Ansatz und bietet vorhersagbare Werte, mit denen Unternehmen hinsichtlich Bereitstellung, Leistung, Kapazität und Kosten planen können. Wenn Unternehmen schneller und in größeren Mengen skalieren wollen, platzieren sie einfach mehrere Bausteine auf einmal.

Schließlich hat sich der Bausteinansatz als attraktiv erwiesen, da die meisten Kunden ihre Implementierungen gerne mit kleineren Bereitstellungen beginnen und sich von dort aus nach oben skalieren. Das Modell „Klein beginnen und bezahlen, während man wächst“ ermöglicht Ihnen, im Voraus weniger Kapital zu investieren und Erfahrungen zu sammeln, während die Bereitstellung wächst. Der nächste Abschnitt behandelt die verschiedenen heute verfügbaren Arten von Infrastrukturarchitekturen und wie jede von ihnen den Bausteinansatz unterstützt oder nicht.

Derzeit gibt es drei primäre Architekturalternativen für die Anwendungs-/Desktop-Virtualisierung oder allgemein für EUC-Lösungen. Die Alternativen sind „Build Your Own“ (BYO), konvergente Infrastruktur (Converged Infrastructure oder CI) und hyperkonvergente Infrastruktur (HCI).

**BUILD YOUR OWN (BYO)**

Die BYO-Infrastrukturalternative ist wirklich genau das, was der Titel impliziert, der Architekt oder das Team wählt unabhängig voneinander Produkte aus, die ihm gefallen oder die er für die besten der Branche hält. Diese Alternative führt zu einer signifikanten Erhöhung der Vorausplanungs- und Recherchezeit, da das Team jedes Produkt einzeln bewerten muss und wie es zusammenarbeiten kann oder nicht.

Diese Alternative bietet auch die Möglichkeit, eine Referenzarchitektur auszuwählen und zu verfolgen, die ein Anbieter für die Art der zu entwickelnden Lösung veröffentlicht hat. Diese Referenzarchitekturen werden in der Regel von einem einzigen Anbieter veröffentlicht und konzentrieren sich auf sein Produkt. Diese Do-it-yourself (DIY)-Referenzarchitekturen können Zeit sparen und Risiken reduzieren, sind jedoch nicht immer für Ihre Designanforderungen, Anwendungsfälle und Ihre Umgebung geeignet.

Eine BYO-Alternative für ein EUC-basiertes Design enthält zumindest Rechen- und Speicherressourcen. Möglicherweise können Sie die vorhandene Netzwerkkonnektivität verwenden, so dass sie vielleicht nicht Bestandteil dieser Alternative ist. Abbildung 2 veranschaulicht ein einfaches Beispiel der Komponenten einer BYO-Alternative. Durch die Flexibilität bei der Skalierung sind die Kosten ziemlich vorhersagbar; die einzige Ausnahme betrifft den Speicher. Abhängig von der maximalen Größe Ihres Designs und der getroffenen Speicherwahl können Sie mehrere Speicher-Arrays oder -Geräte benötigen. Wenn Sie den Speicher skalieren und ein neues Array oder eine neue Appliance hinzufügen müssen, steigen die Kosten in diesen Bereichen.

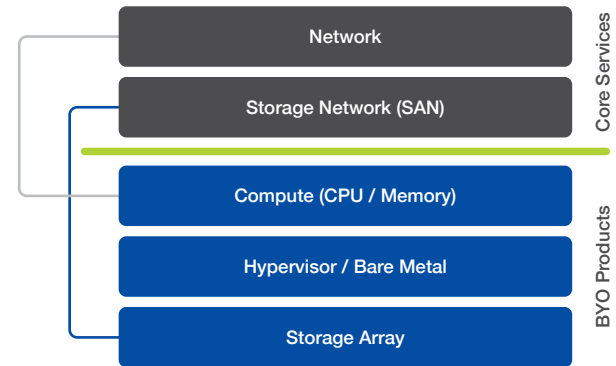


Abbildung 2  
Bring Your Own (BYO)-Infrastruktur

Jedes Mal, wenn Sie eine Reihe von Produkten desselben Anbieters oder mehrerer Anbieter ohne Vorkenntnisse zusammenstellen, besteht ein zusätzliches Risiko. Es wird ein gewisses Maß an Unsicherheit über die Leistung und Zuverlässigkeit der Lösung geben, bis die eigentliche Infrastruktur gekauft und im geplanten Objekt bereitgestellt wird.

Wenn man die Unbekannten und das zusätzliche Risiko akzeptieren kann, maximiert die BYO-Alternative die Flexibilität. Da man in der Lage ist, fast jede Hersteller- und Produktentscheidung zu treffen, wenn ein Zusammenspiel möglich ist, können Sie bei bestehenden Anbietern bleiben, mit denen Sie zufrieden sind, während Sie vielleicht in anderen Bereichen neue Lieferanten suchen.

Die BYO-Alternative ist in der Lage, die Rechen- und Speicherressourcen unabhängig voneinander zu skalieren. Die einzigen Grenzen für die Skalierungsmethode oder die maximale Größe sind Einschränkungen aufgrund der individuellen Produktauswahl. Da die Produkte separat gekauft werden, gibt es keine Mindestanforderungen oder festgelegte Mengen, in die die Produkte skaliert werden müssen. Dies ermöglicht Flexibilität bei dem Versuch, den bereits erwähnten Bausteinansatz zu berücksichtigen.



## KONVERGENTE INFRASTRUKTUR

Eine Alternative in Form einer konvergenten Infrastruktur (CI) ist eine Architektur, die um 2010 auf den Markt gebracht wurde. Konvergente Infrastrukturangebote bieten typischerweise die gleichen Produkte, die als Teil der BYO-Alternative ausgewählt werden können, und bündeln sie zu einer Produktlösung. Das bedeutet, dass ein CI-Anbieter Computer, Speicher und Netzwerke in sein Angebot aufnimmt. In der Regel enthalten die meisten CI-Angebote Produkte mehrerer Anbieter, die in ein einziges Angebot aufgenommen werden, oder ein Anbieter kann alle Komponenten eines CI-Angebots aus seiner eigenen Produktlinie anbieten. Abbildung 3 veranschaulicht ein einfaches Beispiel für eine Alternative in Form einer konvergenten Infrastruktur.

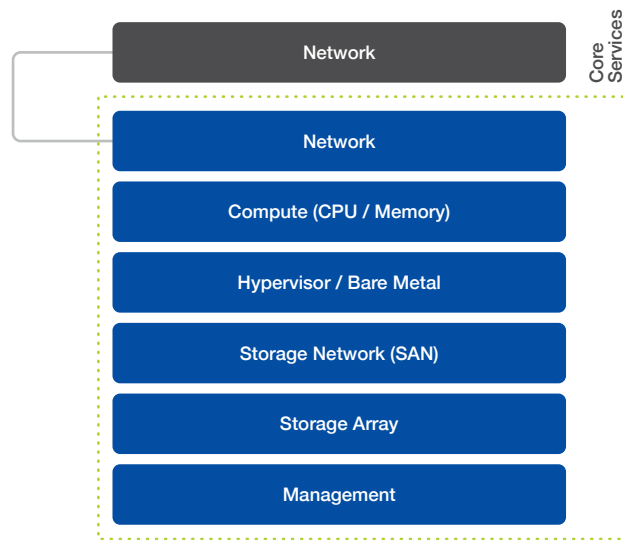


Abbildung 3  
Konvergente Infrastruktur

Ein Angebot für eine konvergente Infrastruktur ermöglicht Ihnen, bekannte Produkte zu erwerben, die in einer einzigen Lösung verpackt sind. Sie können sich dies wie eine Referenzarchitektur vorstellen, die als Produkt erworben werden kann. Je nach dem zu bewertenden CI-Produkt kann das Produkt eine zusätzliche Konvergenz bieten oder auch nicht im Vergleich zu einem separaten Kauf der Produkte im Rahmen einer BYO-Alternative.

In der Regel bieten die meisten CI-Anbieter und -Produkte die Möglichkeit, alle Infrastrukturateile in einer einzigen Produkt-SKU zu erwerben. Der CI-Anbieter sollte in der Lage sein, einen Single-Call-Support für die gesamte CI-Lösung anzubieten, was bedeutet, dass der CI-Anbieter alle Produkte innerhalb der Lösung unterstützen kann. Dies ist ein zusätzlicher Vorteil, da die Kunden so die Notwendigkeit vermeiden können, bei der Fehlersuche mit mehreren Anbietern zusammenzuarbeiten.

Bei den meisten CI-Angeboten gibt es innerhalb der Lösung eine begrenzte Anzahl von Produkten. Dies ermöglicht dem CI-Anbieter, alle Komponenten und Einzelteile vorab zu testen und zu validieren, um sicherzustellen, dass sie gemeinsam ordnungsgemäß funktionieren, wodurch ein Großteil des Risikos der BYO-Alternative beseitigt wird.

Auch nach mehreren Jahren des Verkaufs von CI-Produkten auf dem Markt wurde von den CI-Anbietern wenig getan, um das Management dieser Produkte zu vereinfachen. Mit CI-Angeboten, die die gleichen Produkte wie die BYO-Alternativen enthalten, wird man typischerweise beide Alternativen in ähnlicher und verstreuter Weise verwalten. Diese Alternative kann den Kauf und/oder einige der Produkte bündeln, aber sie führt in der Regel nicht zu einer Konvergenz des täglichen operativen Managements der Lösung.

Ein konvergentes Infrastrukturprodukt sollte in der Lage sein, die darin enthaltenen Ressourcen unabhängig voneinander zu skalieren. Dies würde bedeuten, dass Sie nur Rechenleistung hinzufügen können, obwohl es vielleicht Mindeststufen gibt, die man hineinskalieren kann. Die andere Ressource, die man in einem CI-Angebot erweitern könnte, ist der Speicher. Dies hängt jedoch stark von der Art der Speicherlösung ab, die als Teil des CI-Angebots ausgewählt wird. Ein konvergentes Infrastrukturprodukt hat eine maximale Größe, d.h. es gibt eine Begrenzung hinsichtlich der Anzahl der Server, die die Lösung unterstützen kann, und eine Speicherbegrenzung basierend auf dem enthaltenen Speicher-Array.

Die Skalierungsgrenzen eines CI-Angebots sind in der Regel recht groß, aber eines Tages wird das Maximum erreicht, wenn man die Ressourcen innerhalb des CI-Produkts skaliert. Um das Design an diesem Punkt weiter skalieren zu können, muss man ein zusätzliches CI-Produkt erwerben. Dies führt abhängig von der maximalen Größe Ihres Designs zu großen Spitzen bei den Infrastrukturkosten an verschiedenen Punkten des Skalierungsprozesses.

## HYPERKONVERGENTE INFRASTRUKTUR

Die hyperkonvergente Architektur kam etwa ein Jahr nach CI auf den Markt. Echte hyperkonvergente Architekturen werden durch die Zusammenführung von Rechenressourcen, Speicherressourcen und Managementebene in einem einzigen Produkt erreicht. Es ist möglich, eine hyperkonvergente Lösung in einer BYO- oder Referenzarchitekturmethode zu implementieren; um jedoch wirklich hyperkonvergent zu sein, muss das Produkt die Hardware-Appliance beinhalten.

Durch die Integration einer Hardware-Appliance als Teil des Produkts kann der Anbieter nun das Management der Infrastruktur zusammen mit den anderen im Produkt zusammengeführten Ressourcen mit einbeziehen. Abbildung 4 veranschaulicht ein einfaches Beispiel für eine hyperkonvergente Infrastrukturalternative.

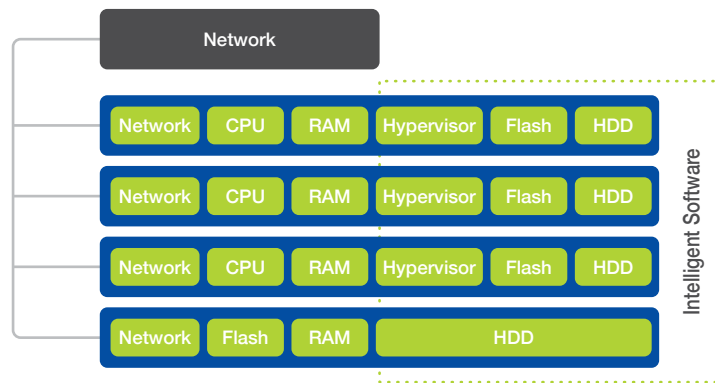


Abbildung 4  
Hyperkonvergente Infrastruktur

Ein wirklich hyperkonvergentes Produkt bietet eine Reihe von Vorteilen, die andere Referenzarchitekturen nicht aufweisen können:

**Einfache Installation** - Bei führenden HCI-Produkten sollte die Installation von Nodes durch einen hochautomatisierten Prozess innerhalb von Minuten und Stunden anstatt von Tagen oder Wochen möglich sein.

**Einfache Skalierbarkeit** - Das Produkt sollte leicht nach oben oder unten skalierbar sein. Das Hinzufügen neuer Nodes in die Umgebung sollte einfach und schnell über die Management-Schnittstelle erfolgen.

**Modernes Management** - Eine moderne Management-Schnittstelle muss sich auf die virtuelle Maschine (VM) als Managementpunkt konzentrieren. Ein Administrator muss in der Lage sein zu verstehen, wie VMs funktionieren, wie viele Ressourcen jede VM verbraucht, ob es Ereignisse oder Fehler gibt, und er muss die Möglichkeit haben, Berichte auf der Grundlage von VMs einfach zu erstellen.

**Erweiterbarkeit** - Sie müssen in der Lage sein, die Infrastruktur einfach mit anderen Teilen der Lösung zu integrieren und programmgesteuert zu kontrollieren. Dies erfordert, dass das HCI-Produkt eine API und möglicherweise eine andere Methode wie beispielsweise PowerShell-Befehle anbietet. Mit einer API können Sie die Kommunikation und Steuerung zwischen den Produkten automatisieren, um so den Aufwand weiter zu reduzieren und die Genauigkeit der Umgebung zu erhöhen.

Die Leistung wurde bewusst nicht in die Liste der Vorteile der HCI aufgenommen, da jeder erwartet, dass eine moderne hybride oder flashbasierte Lösung gut funktioniert. Bei HCI geht es darum, eine Infrastrukturschicht zu schaffen, die einfach und effizient ist. Damit können Teams aufhören, Zeit mit dem „Drehen von Schaltern“ zu verbringen und einen zusätzlichen Mehrwert für das Unternehmen auf der Automatisierungs- oder Anwendungsebene schaffen.

Es gibt bei jedem EUC-Design eine Reihe von verschiedenen Anforderungen an die Speicherressourcen. Das Design muss für serverbasierte VMs, Benutzerdaten und virtuelle Desktop-Infrastrukturen (VDI, auch als Benutzer-VMs bezeichnet) ausgelegt sein. Die damit verbundenen Speicheranforderungen sind die anspruchsvollsten innerhalb der Umgebung und auch diejenigen, die die meisten Projekte zum Scheitern bringen oder zu schlechten Erfahrungen führen.

Aus diesem Grund wird sich das Thema Speicherung dieses E-Books auf die Bedürfnisse des VDI-Service der Lösung konzentrieren. Die Anforderungen jedes virtuellen Desktops können oft klein und unbedeutend erscheinen, wenn man sie jedoch mit zunehmendem Speicherbedarf zu großen Gruppen zusammenfasst, können die Leistungsanforderungen den Speicherplatz leicht überfordern, wenn dieser nicht ordnungsgemäß für diese Anforderungen dimensioniert wurde.

Wenn jeder virtuelle Desktop durchschnittlich 15 IOPS mit angemessener Latenz ausführt und man 2.000 gleichzeitige Anwender erwartet, sind das 30.000 IOPS. Diese Zahl ist ziemlich groß und könnte das durchschnittliche Speicher-Array überfordern. Aber man kann die Speicherlösung nicht einfach so gestalten, dass sie dem durchschnittlichen I/O der Umgebung entspricht, das Design muss Spitzenwerte berücksichtigen, dazu gehören auch Desktop-Boots und massenhaft Anwender-Logins.

Ein virtueller Desktop-Workload unterscheidet sich stark von anderen Arten von Workloads, die innerhalb eines durchschnittlichen Unternehmensrechenzentrums ausgeführt werden, da virtuelle Desktops in ihrer I/O-Natur sehr zu Spitzen neigen. Wenn Sie beispielsweise eine Anwendung wie Outlook zum ersten Mal in einer Sitzung öffnen, können

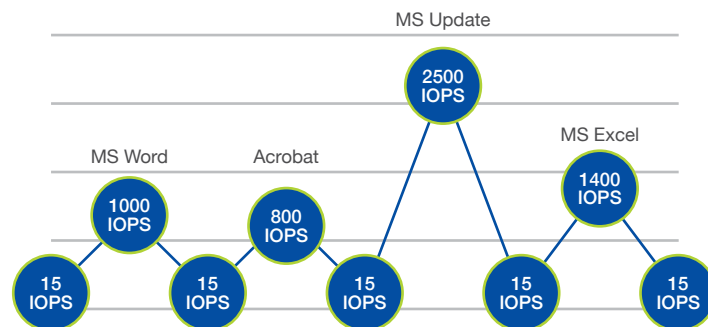


Abbildung 5  
VDI IOPS

Sie bis zu 1.000 IOPS für diese eine Benutzersitzung generieren. Das liegt weit über dem Durchschnitt von 15 IOPS, dem zuvor diskutierten Wert. Ein Beispiel für verschiedene Auswirkungen von IOPS von Anwendungen ist in Abbildung 5 dargestellt.

Andere Bereitstellungen und operative Elemente, wie Patches und Umgebungsaktualisierungen, können ebenfalls zu enormen Spitzenwerten bei IOPS führen und die Leistung beeinträchtigen, wenn sie nicht berücksichtigt und entsprechend geplant werden. Wenn man weitere 50 virtuelle Desktops bereitstellt, kann diese Aktion zu einem erheblichen Anstieg der I/Os führen. Aus diesen Gründen muss man die Instandhaltungsmaßnahmen in der Speicherarchitektur für Spitzen-IOPS berücksichtigen.

Es gibt eine Reihe von Möglichkeiten, VDI-Lösungen mit vollständigen Klonen oder gemeinsam genutzten Images zu erstellen, die sich sowohl in Bezug auf die Kapazität als auch auf die Leistung jeweils unterschiedlich auf die Speicheranforderungen auswirken können. Da vollständige Klone zusätzliche Kapazität und Speicher verbrauchen, spielt die Deduplizierung eine wichtige Rolle. Vollständige Klone müssen auch unabhängig voneinander gepatcht werden, was die I/O während dieser Operationen erhöht.

Der Shared-Image-Ansatz, den Citrix mit MCS oder PVS und VMware mit verlinkten Klonen anbietet, stellt verschiedene I/O-Herausforderungen dar. Von Natur aus benötigen diese Shared-Image-Ansätze weniger Speicherkapazität, da das übergeordnete Image gemeinsam genutzt wird und jeder virtuelle Desktop nur weniger Platz für seine einzigartigen Daten benötigt. Das Shared Image hat andere Leistungsanforderungen als die typische VM. Dieses Image wird nun von Hunderten oder Tausenden von virtuellen Desktops verwendet und muss in der Lage sein, große Mengen an IOPS zu generieren, um Situationen wie Boot-Engpässe zu bewältigen. Wenn das Shared Image einen Engpass darstellt, werden alle virtuellen Desktops, die es verwenden, negativ beeinflusst und die Anwendererfahrung wird beeinträchtigt.

Unter Berücksichtigung dieser Überlegungen zu Spitzenlasten und verschiedenen Arten von Anwendungs- und Desktop-Virtualisierungsarchitekturen muss man eine Speicherlösung auswählen und konzipieren, die in der Lage ist, die Anforderungen der Umgebung an Boot-Spitzen, Login und stationären Zustand zu erfüllen. Um die Speicheranforderungen des Designs zu verstehen, sollte man eine Bewertung der vorhandenen physischen PC-Umgebung durchführen. Im Rahmen dieser Einschätzung werden die tatsächlichen Leistungs- und Kapazitätsdaten aus der Benutzerbasis erfasst, so dass man diese auf die Designberechnungen anwenden kann.

Ein letzter Gedanke zu den Speicheranforderungen im Zusammenhang mit Anwendungs- und Desktop-Virtualisierung ist, dass Desktop-Workloads nicht nur sehr unvorhersagbar auf der I/O-Seite sind, sondern auch sehr schreibintensiv. Im Gegensatz zu vielen Server-Workloads, die hauptsächlich Daten lesen und an Anwender verteilen, verbringen Desktops in der Regel mehr Zeit mit dem Schreiben auf die Festplatte. Schreiben ist für das Speicher-Array intensiver als Lesen. Ein typischer Server-Workload kann zu 80% aus Lesen und zu 20% aus Schreiben bestehen, während bei einem stationären virtuellen Desktop-Workload das Gegenteil der Fall sein könnte. Bei der Bewertung Ihrer Speicherauswahl sollten Sie genau darauf achten, wie die Speicherlösung puffert und schreibt, anstatt sich nur auf das Versprechen zu konzentrieren, dass der Speicher „exzellente Arbeit“ beim Zwischenspeichern von häufig gelesenen Blöcken im Rahmen der Bewältigung von Boot-Engpässen leistet.

## **SPEICHERUNGSARTEN**

Es gibt eine Reihe von verschiedenen Arten der Speicherung. Die heute verfügbaren primären Speicheralternativen sind traditionelle mehrstufige Speicher-Arrays, Hybrid-Flash-Arrays und All-Flash-Arrays. Jede Alternative verfolgt einen anderen Ansatz, um Leistung und Kapazität für die Workloads bereitzustellen. Innerhalb jeder Alternative verfolgen die Anbieter unterschiedliche Ansätze beim Aufbau ihrer Angebote, weshalb im Folgenden eine kurze Erläuterung der einzelnen Ansätze aufgeführt wird.

**Traditionelle mehrstufige Architekturen** - Dies sind die älteren Unternehmens-Arrays, die in den letzten 10-20 Jahren für serverbasierte Workloads verwendet wurden. Es handelt sich typischerweise um duale Controller-basierte Architekturen, die in den letzten zehn Jahren so modifiziert wurden, dass mehrere Stufen von Leistungs- und Kapazitätsfestplatten in die Architektur integriert werden konnten. Es werden verschiedene Festplattenstufen bereitgestellt, um die Kapazität und Leistungsanforderungen von verteilten Workloads zu bedienen. Es gibt zwei Optionen in diesem Ansatz: Sie können die Lösung auf Leistung auslegen, indem Sie dedizierte Pools von High-Performance-Festplatten für einen Workload erstellen, dies kann jedoch sehr teuer und einschränkend sein. Die andere Möglichkeit besteht darin, zu versuchen, die Vorteile des dieser Architektur hinzugefügten Tierings zu nutzen, um das Array anzuweisen, Datenblöcke basierend auf der Nachfrage zu erweitern oder aufzubauen. Das Problem bei diesem Auto-Tiering ist, dass es oft zu viel Zeit braucht, um diese Entscheidungen für VDI-Workloads zu treffen.

**All-Flash** - All-Flash-Speicher-Arrays bestehen vollständig aus Flash-basierten Speichern. Es gibt viele verschiedene Arten von Flash, die innerhalb dieser Speicher-Arrays verwendet werden können. Moderne All-Flash-Arrays wurden entwickelt, um die Vorteile der Flash-Speicher zu nutzen, was bedeutet, dass das Betriebssystem und das Dateisystem mit Blick auf Flash entwickelt wurden. Einige Produkte haben ein traditionelles Array-Design übernommen und die rotierenden Disks einfach durch Flash ersetzt. Dies ist zwar immer noch schneller als die ältere Option, das Endprodukt wurde jedoch nicht für diesen Zweck entwickelt.

All-Flash-Speicher-Arrays sind sehr schnell, mit nur einer Performance-Stufe im Produkt. Um sicherzustellen, dass das Array auch die für das Design erforderliche Kapazität zu einem erschwinglichen Preis bereitstellen kann, sollten Sie nach Arrays suchen, die Deduplizierung und Komprimierung bieten. Obwohl fast jedes moderne All-Flash-Array einfacher zu verwalten ist als ihre Vorgänger, bieten sie nicht immer die gleiche einfache Verwaltung und Management pro VM wie viele der Hybrid-Flash-Angebote.

**Hybrid-Flash** - Hybrid-Speicher-Arrays sind moderne Architekturen, die entwickelt wurden, um eine Kombination aus Flash-Laufwerken und rotierenden Festplatten effizient zu nutzen. Anbieter haben unterschiedliche Architekturansätze gewählt, wie sie Kapazität und Leistung in ihren Arrays nutzen, aber die Endergebnisse sind ähnlich. Sie alle sind in der Lage, eine beeindruckende Leistung bei einer geringeren Größe an Flash zu bieten, während sie gleichzeitig eine große Kapazität aufweisen, weil Daten auf großen rotierenden Festplatten im Array gespeichert werden. Ideale Alternativen der hybriden Speicherarchitektur nutzen integrierte Intelligenz, um Daten je nach Bedarf automatisch über Flash- und Festplattenlaufwerke hinweg zu verwalten, wodurch manuelle Anpassungen und potenzielle Leistungsengpässe entfallen.

Die Architekturen, die sich am besten für ein modernes VDI-Design eignen, sind Hybrid- und All-Flash-Speicherarchitekturen. Diese Architekturen sind in der Lage, die für VDI-Umgebungen erforderliche Performance zu erbringen und bieten typischerweise auch die bereits erwähnten modernen Managementenerfahrungen. VDI-Workloads sind von Natur aus sehr unberechenbar, und wenn Ihre Speicherlösung warten muss, um Speicherentscheidungen zu treffen oder Blöcke auf eine Caching-Ebene zu verschieben, wird sich der Leistungsbedarf schon lange vorher erledigt haben und die Erfahrung dadurch negativ beeinflusst.

Es gibt verschiedene Denkansätze, um die Größe der Rechenebene des Designs zu bestimmen. Der erste ist der Scale-Up-Ansatz, der wenige große Hosts zur Bereitstellung von Ressourcen verwendet, während der Scale-Out-Ansatz mehrere kleine Hosts zur Ressourcen-Bereitstellung verwendet. Die bevorzugte Methode liegt irgendwo zwischen diesen beiden Ansätzen; es ist eine, die 2 Socket-Hosts nutzt und sie so dicht wie möglich macht, ohne die im Rahmen des Designs festgelegten Konsolidierungsraten zu verletzen. Dieses Buch konzentriert sich auf die Ermittlung der Größe der Rechenressourcen für den VDI-Workload.

Es gibt drei primäre Berechnungen, auf die man sich bei der Dimensionierung der Rechenressourcen im Design konzentrieren sollte. Diese sind die Größe des physikalischen Speichers in jedem Host, die Höhe der CPU-Taktfrequenz und die Anzahl der CPU-Kerne und deren CPU-Verhältnis. In erster Linie sollte man bei einem VDI-Design niemals den Speicher überbeanspruchen. Ein Verstoß gegen diese Regel hat nur einen sehr geringen Wert und führt nur zu Performanceeinbußen in der Umgebung.

Die Berechnung der CPU-Taktfrequenz hängt stark von den Details ab, die in der vorherigen Desktop-Bewertung erfasst wurden. Berichte aus der Bewertung liefern die CPU-Leistung, die die Benutzersitzungen im Durchschnitt und in der Spitze verbraucht haben. Diese Informationen werden zusammen mit den Speicherdetails aus der Bewertung verwendet, um die Berechnungen durchzuführen.

Weitere Empfehlungen für Host- und Virtualisierungs-Cluster sind, dass die Hostauslastung niemals 80% übersteigen darf und dass Ihr Cluster immer für N+1 ausgelegt sein sollte. Die 80%ige Hostauslastung gilt nicht nur für Anwendungs-/Desktop-Virtualisierungs implementierungen, sondern ist eine Empfehlung, die für jeden Workload gilt, der auf einem Hypervisor läuft. Wenn Sie Ihre Hosts über die 80%-Marke hinaus auslasten, haben Sie nur sehr wenig Platz für Spitzen und haben möglicherweise, abhängig von der Größe Ihres Clusters, auch nicht genügend Ressourcen-Overhead im Falle eines Host-Fehlers. Das zweite Element der Konfiguration von N+1 in Ihrer Clustergröße soll sicherstellen, dass es genügend Ressourcen in Ihrem Cluster gibt, um einen einzelnen Hostausfall zu verkraften, dass alle VMs weiterhin ausgeführt und ausgefallene VMs ohne Probleme neu gestartet werden können. Ein Ausfall eines einzelnen Hosts ist die häufigste Stufe der Belastbarkeit; es gibt eine kleine Gruppe von Kunden, die N+2 benötigen, um höhere SLA-Anforderungen zu erfüllen.

Das letzte Element zum Thema „Berechnung der Dimensionierung“ ist die CPU-Rate. Diese konzentriert sich auf die Anzahl der virtuellen CPUs gegenüber physikalischen CPUs. Dieses Verhältnis ist sehr wichtig, denn wenn es zu hoch ist, wird ein Punkt erreicht, an dem ein CPU-Scheduling-Problem auftritt, das die Leistung und die Benutzerfreundlichkeit dramatisch beeinträchtigt. Wenn ein CPU-Scheduling-Problem auf vSphere-Hosts auftritt, erhöht sich die CPU-Bereitschaftszeit, und dies zeigt an, dass der Scheduler Probleme hat, alle geplanten vCPUs auf pCPUs zu bekommen. Das bedeutet, dass die vCPU warten muss, obwohl sie bereit ist. Das CPU-Verhältnis ist für die verschiedenen Arten von Workloads, die auf VMware-Clustern virtualisiert werden, sehr unterschiedlich. Typischerweise haben Server- und Datenbank-Workloads ein viel geringeres Verhältnis, während VDI-Workloads ein höheres Verhältnis aufweisen können.

Die Verwendung von vCPUs ist keine lineare Berechnung, was bedeutet, dass man einen Host mit einer höheren Konsolidierungsquote aufbauen kann, wenn alle VMs nur eine einzige vCPU haben. Wenn viele VMs über 2 oder mehr vCPUs verfügen, wirkt sich dies auf die Berechnungen aus. Es ist nicht einfach nur eine Division 2, damit sich doppelt so viele vCPUs ergeben. Abbildung 6 stellt einen Bereich dar, der sich im Rahmen realer Kundenimplementierungen bewährt hat. Hersteller, die synthetische Tests durchführen, können höhere Werte aufweisen. Diese sollten mit Vorsicht betrachtet werden, da sie nicht immer für reale Designs gelten.

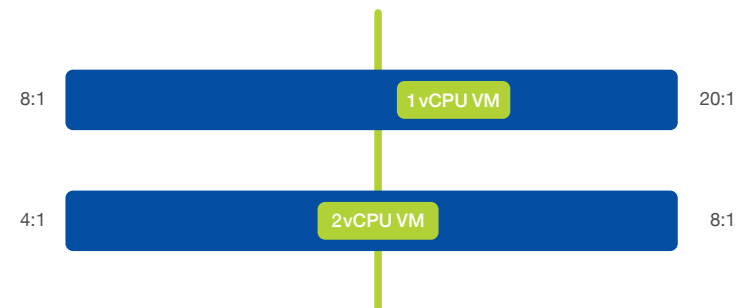


Abbildung 6

Die VDI-Konsolidierung wird stark durch das Verhältnis von vCPU bestimmt, mit dem Ihre virtuellen Desktops konfiguriert werden. Das Diagramm stellt einen Bereich dar, der sich erfahrungsgemäß als sicher erwiesen hat.

Der Arbeitsbereich, in dem virtuelle Desktops mit Single-vCPU normalerweise operieren, liegt zwischen 8:1 und 20:1. Dies ist ein großer Bereich, und wo man in diesem Bereich landet, wird durch verschiedene Entscheidungen bestimmt. Eine davon betrifft die Größe der Hosts, die Anzahl der VMs pro Host und wie der Kunde mit dieser Zahl zufrieden ist. Ein Beispiel wäre ein Dual-Socket-Host mit zwei 18-Core CPUs. Dieser könnte nach oben hin mehr als 700 VMs unterbringen, vorausgesetzt, Sie haben die richtige Menge an Speicher und genügend Taktfrequenz zur Verfügung. Normalerweise würde eine so große Anzahl von VMs auf einem einzigen Host die meisten Kunden beunruhigen. Es gibt also zwei Möglichkeiten, die in diesem Szenario zu treffen sind, und zwar erstens die Wahl einer niedrigeren Dichte, die man künstlich begrenzt.

Wenn man das untere Ende des Verhältnisses wählt, würden sich netto 288 VMs auf demselben Host ergeben. Die zweite Option wäre, CPUs mit weniger Kernen zu wählen, aber ein Verhältnis irgendwo in der Mitte zu wählen. Wenn man 12 Kern-CPU's wählt und ein Verhältnis von 12:1 verwendet, würde das 288 VMs netto bedeuten. Dieser Entscheidung liegt in der Regel eine Kombination aus Kundenfeedback, Architektenempfehlungen und Infrastrukturkosten zugrunde. Durch die Wahl verschiedener physikalischer CPU-Konfigurationen können erhebliche Kosteneinsparungen erzielt werden.

Die Berechnungen für einen virtuellen Dual-vCPU-Desktop sind ähnlich, außer dass es sich jetzt um die doppelte Anzahl von vCPUs handelt. Der Betriebsbereich liegt hier zwischen 4:1 und 8:1. Einige Anbieter versprechen mehr, aber diese Empfehlungen ergeben sich aus echten Implementierungen von Kunden. Man sollte die gleichen Entscheidungsmerkmale wie im vorherigen Beispiel verwenden, nur mit einem anderen CPU-Verhältnis-Bereich.

Was Sie außerdem beachten sollten, ist, dass sich die Freiheit ergibt, die Konsolidierungsdichte nach oben zu skalieren, wenn die Umgebung weiterhin innerhalb der Toleranzen arbeitet und Sie ein CPU-Verhältnis in der Mitte dieser Bereiche wählen. Zu beachten ist, dass es heute keinen Ort gibt, diese CPU-Verhältnisse als Einstellung in anderen Tools zu konfigurieren. Dies sind Attribute, die im Design angegeben werden müssen und zu Datenpunkten werden, die man bei der Verwaltung und Skalierung der Umgebung berücksichtigen muss. Genauso wie Speicher und Taktfrequenz muss das CPU-Verhältnis in die Entscheidung einbezogen werden, weitere VMs zu einem Cluster hinzuzufügen und wann man einem Cluster einen weiteren Host hinzufügt, um mehr Ressourcen bereitzustellen.

Man kann das CPU-Verhältnis durch manuelle Berechnungen verwalten, indem man Daten sammelt. Einige Administratoren verwenden ein PowerShell-Skript, das Daten erhebt und das Verhältnis als Ausgabe des Skripts darstellt. Ein Skript kann täglich als geplanter Job ausgeführt werden, um sicherzustellen, dass man das Verhältnis nicht verletzt und sich auf einem der Cluster eine Bedrohung ergibt.

Die RAM- oder Speicher-Bus-Frequenz ist ebenfalls mit der Berechnungsgröße verknüpft. Die Faustregel bei der Dimensionierung von Speicher ist, die höchste Dichte mit der schnellsten Busgeschwindigkeit im Rahmen des Budgets zu erreichen. Die Herausforderung, die sich oft mit dem Speicher ergibt, besteht darin, dass ein langsamerer Speicher zu ungenutzten CPU-Zyklen führen kann, die darauf warten, dass Lese-/Schreibvorgänge im RAM abgeschlossen werden.

Es gibt eine Reihe von Gründen, verschiedene Virtualisierungscluster in einem EUC-Design zu erstellen. Die Entscheidung, verschiedene Cluster zu nutzen, wird in der Regel von unterschiedlichen Workloads und Clustergrößen bestimmt. Diese Thematik wird in diesem E-Book nicht ausführlich behandelt, hier sind jedoch einige Empfehlungen, die auf den Themen aufbauen, die an anderer Stelle im Buch insgesamt und online angesprochen werden.

In erster Linie ist es beim Aufbau eines VDI-Designs mit mehr als einigen hundert Anwendern unerlässlich, die Virtualisierungsmanagement-Infrastruktur vom VDI-Workload zu trennen. Das bedeutet, dass alle Management-Server, VDI-Broker, Dateiserver, Anwendungs-Management-Server und alle anderen Funktionen, die keine virtuellen Desktops sind, auf einem anderen Cluster laufen sollten. Ob der Managementcluster ein einziger, dem EUC-Design gewidmeter Cluster sein muss, hängt davon ab, wie groß die Umgebung sein wird. Wenn das Design kleiner ist, kann man Management-VMs in einem bestehenden Server-Virtualisierungs-Cluster ausführen.

Es ist möglich, diese virtuellen Desktop-Cluster so zu skalieren, dass sie eine Größe von 16-32 Hosts erreichen. Dieser Bereich ermöglicht die Bildung eines größeren Ressourcenpools für die Verwendung von VMs und veranlasst die meisten Kunden, einen Cluster zu implementieren, der größer als ihre typische Dimensionierung ist. Neuere Hypervisor-Updates ermöglichen Cluster mit bis zu 64 Hosts, aber es wird einige Zeit dauern, bis sich viele Architekten und Kunden mit solchen Größenordnungen wohlfühlen. Wenn die Umgebung so groß ist, dass die Anzahl der Hosts diese Bereiche überschreiten würde, wäre mehr als ein VDI-Cluster erforderlich.

Ein weiterer Grund, warum man neben der Umgebungsgröße auch mehrere Virtualisierungscluster entwirft, sind unterschiedliche Workloads. Innerhalb der VDI-Cluster gibt es unterschiedliche Workloads. Wenn es eine beträchtliche Anzahl von virtuellen Desktops mit 1 vCPU und 2 vCPU gibt, sollte man für jede Art einen eigenen Cluster entwerfen. Abbildung 7 veranschaulicht einen Multi-Cluster-Designansatz. Dies ermöglicht eine unterschiedliche Steuerung des CPU-Verhältnisses in jedem Cluster und ein Design für einfachere Administration. Wenn man die verschiedenen CPU-Konfigurationen zusammenführen würde, gäbe es ein neues gemischtes Verhältnis, das man berechnen müsste und das nur Verwirrung stiften würde.

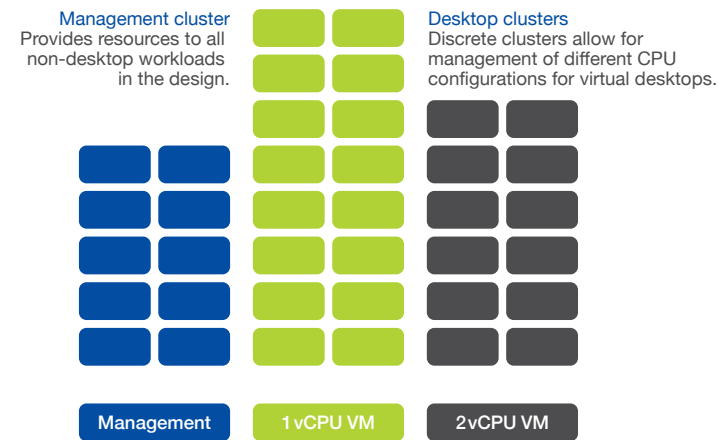


Abbildung 7  
Management- und Desktop-Cluster

Wenn Sie so weit gelesen haben, hoffen wir, dass Sie von den Möglichkeiten der Desktop- und Anwendungsvirtualisierung zumindest fasziniert sind. Wenn Sie und Ihr Unternehmen bereit sind, zu lernen, wie man Desktop- und Anwendungsvirtualisierung erfolgreich einsetzt, möchte Nutanix Ihnen helfen. Die unsichtbare Infrastruktur von Nutanix kann die Reise zur Virtualisierung dank ihrer preisgekrönten Web-Scale-Architektur - der besten VDI-Plattform - erheblich vereinfachen.

### WENN ES ZEIT FÜR VERÄNDERUNGEN IST

Es wird Sie wahrscheinlich nicht überraschen, dass Nutanix viel darüber nachgedacht hat, die besten Möglichkeiten für eine erfolgreiche Bereitstellung von Desktop- und Anwendungsvirtualisierung zu ermitteln.

#### Verstehen Sie Ihre aktuelle Umgebung:

Der Prozess beginnt mit einem vollständigen Verständnis Ihrer aktuellen End-User-Umgebung, dazu gehören:

- End-User-Metriken: Erfassen Sie End-User-Profile und verwandte Faktoren wie verwendete Anwendungen, End-User-Zugangsgeräte, Standort und Konnektivität.
- Netzwerkdienste und infrastrukturenspezifische Metriken: Sammeln Sie geeignete Informationen über verschiedene End-User-Services wie Datei-Services, Authentifizierung und Zugriffskontrolle sowie Firewalls/Load Balancing. Erfassen Sie auch Performance-Metriken, Latenzen, Durchsatz etc.
- Ordnen Sie alles den Verantwortlichen zu: Verantwortlichkeit ist ein wichtiger Erfolgsfaktor.

#### Dimensionierung der neuen Umgebung:

Mit den oben genannten Informationen können Sie Ihre neue Umgebung exakt dimensionieren. Nutanix Sizer macht diese Aufgabe einfach. Die folgenden Richtlinien sollten sie dennoch beachten:

- Achten Sie stets auf eine hohe Verfügbarkeit der wichtigsten Server und Desktops.
- Auf der Grundlage dieser Überlegungen kann eine zusätzliche Infrastruktur und/oder können zusätzliche Cluster erforderlich sein:
  - Business: SLAs, Lizenzierung, Sicherheit, Budget, Richtlinien
  - Planung des Übergangs: Befolgen Sie die Best Practices und Richtlinien von Nutanix und der Branche für P2V-Migrationen und achten Sie besonders auf die Erstellung von Golden Images, die zur Erstellung anderer Desktops verwendet werden. Wenn Sie eine bestehende Implementierung migrieren, empfehlen wir, nach Möglichkeit Partner- oder native Tools von Nutanix zu verwenden.

Natürlich kann Nutanix Global Services Ihnen bei einem dieser Schritte oder allen helfen, um Sie auf den Weg zu einem größeren Infrastrukturerfolg zu bringen. Durch unsere Global Services-Organisation bietet Nutanix die branchenweit einzige Lösung, um das Risiko einer falschen Dimensionierung der Infrastruktur für Desktop-Virtualisierungsprojekte zu vermeiden.

Im Rahmen des VDI-Assurance-Programms stellt Nutanix sicher, dass Ihre virtuellen Desktops immer die Rechenleistung (virtuelle CPU und Arbeitsspeicher) und Speicherressourcen (Performance und Kapazität) erhalten, die sie benötigen, um die VDI-Anforderungen der Endbenutzer zu erfüllen. Bestimmen Sie einfach Art und Anzahl der VDI-Anwender in Ihrer Umgebung und übertragen Sie das Risiko der Dimensionierung der Infrastrukturanforderungen mit VDI Assurance auf Nutanix.

Um mehr über die unsichtbare Infrastruktur für Unternehmensanwendungen zu erfahren, kontaktieren Sie uns unter [info@nutanix.com](mailto:info@nutanix.com), folgen Sie uns auf [Twitter @nutanix](https://twitter.com/nutanix) oder senden Sie uns eine Anfrage an [www.nutanix.com/demo](http://www.nutanix.com/demo), um Ihr eigenes maßgeschneidertes Briefing und eine Demonstration zu erstellen, damit Sie sehen können, wie validierte und zertifizierte Lösungen von Nutanix Ihrem Unternehmen helfen können, das Beste aus seinen Unternehmensanwendungen herauszuholen.

Bleiben Sie mit Experten und Kunden von Nutanix in der Nutanix Next-Online-Community ([next.nutanix.com](http://next.nutanix.com)) in Kontakt.



Nutanix bietet eine unsichtbare Infrastruktur für die nächste Generation von Enterprise Computing und stärkt die IT, damit sie sich auf die Anwendungen und Services konzentrieren kann, die das Geschäft vorantreiben. Die Software-gesteuerte Xtreme Computing-Plattform des Unternehmens vereint Rechenleistung, Virtualisierung und Storage nativ in einer einzigen Lösung, um die Einfachheit im Rechenzentrum zu erhöhen. Mit Nutanix profitieren Kunden von vorhersagbarer Performance, linearer Skalierbarkeit und cloud-ähnlichem Infrastrukturverbrauch. Erfahren sie mehr auf [www.nutanix.de](http://www.nutanix.de) oder folgen Sie uns auf Twitter @NutanixGermany.

**NUTANIX™**

Tel. +49 89 25552898

[info@nutanix.com](mailto:info@nutanix.com) | [www.nutanix.de](http://www.nutanix.de) | [@NutanixGermany](https://twitter.com/NutanixGermany)