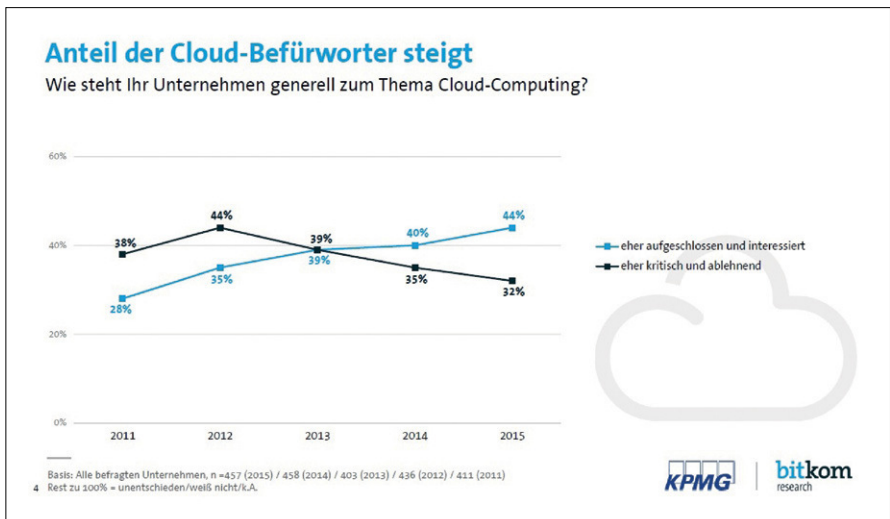


Unternehmens-IT in Cloud-Geschwindigkeit

Der Wettbewerb in der digitalen Wertschöpfung ist von einer hohen Schlagzahl geprägt. Gefragt sind Konzepte wie DevOps und Continuous Delivery, die ein Umdenken der IT-Abteilungen Richtung Cloud erfordern, mit der Intention: Cloud-Geschwindigkeit muss auch im Unternehmens-Rechenzentrum möglich sein.

In einer Umfrage von **Oxford Economics** (www.oxfordeconomics.com) unter 300 Geschäfts- und IT-Entscheidern gaben kürzlich fast zwei Drittel der Befragten an, dass in den letzten zwei Jahren vor allem der Druck gestiegen sei, schneller zu sein als die Konkurrenz. Als zweite Priorität nannten sie entsprechend eine schnellere Entwicklung von Produkten und Dienstleistungen.

Ein überraschendes Ergebnis der Umfrage: Das meistgenannte Hindernis war – noch vor Faktoren wie Regularien, Datensicherheit oder dem Mangel an qualifiziertem Personal – eine „rigide und inflexible Technologie-Infrastruktur“. Über die Hälfte der Befragten (54 Prozent) sahen ihre IT nur „einigermaßen“ gut gerüstet, schnelle Innovation zu unterstützen – vier Prozent trauten es ihr gar nicht zu.



➤ Der Anteil der Cloud-Befürworter steigt seit Jahren beständig an. (Quelle: Bitkom/KPMG)

Agilität à la Cloud

Dabei ist es heute mit agiler Softwareentwicklung und flexiblen IT-Infrastrukturen durchaus möglich, zügig auf sich ändernde Anforderungen zu reagieren. Namhafte Public-Cloud-Anbieter demonstrieren dies jeden Tag. Dauert es in einem herkömmlichen Unternehmens-Rechenzentrum mehrere Tage oder Wochen, bis die IT-Organisation einen neuen Server bereitstellt, so ist zum Beispiel eine neue AWS-Instanz nach wenigen Minuten verfügbar. Der Vergleich ist natürlich unfair. Die meisten Firmen müssen eine gewachsene IT-Landschaft betreiben und Hunderte oder Tausende von Anwendungen aktuell halten und das mit komplexen Genehmigungs-Prozessen und Applikationsabhängigkeiten.

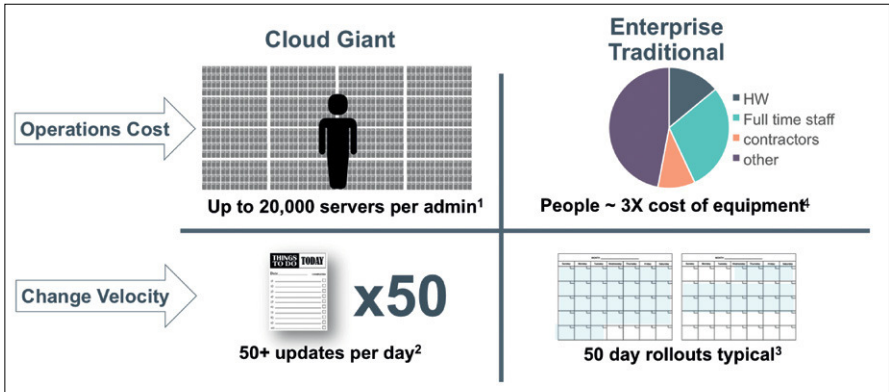
Cloud-Unternehmen wie Amazon Web Services (AWS), Facebook oder Google hingegen nutzen relativ wenige, speziell auf ihre Anforderungen zugeschnittene Programme, die sie meist selbst entwickelt haben. Als Hardwarebasis dienen Whitebox-Ware mit proprietärem Code (etwa bei Facebook) oder selbstgebaute Server (wie bei Google).

50 Updates pro Tag

Manche Online-Anbieter stemmen dabei rund 50 Updates ihrer Plattform pro Tag. Grundlage für dieses hohe Tempo sind DevOps-Praktiken mit kontinuierlicher Aktualisierung des Codes (Continuous Delivery) und automatisierter Softwareverteilung. Bei der E-Commerce-Plattform Etsy etwa hat jeder Entwickler seine eigene KVM-Umgebung, die zentrale Kontrolle erfolgt über das Orchestrierungswerkzeug Chef.

Im Fall der großen Cloud-Anbieter gesellt sich zu DevOps, Automation und Self-Service der Einsatz von Container-Technologie für die schnelle hardwarenahe Provisonierung von Applikationsservices, ohne den Umweg über ein Virtualisierungs-Layer gehen zu müssen. Da die Cloud-Größen in dieser Liga jeweils Millionen von Servern betreiben, erzielen sie zudem erhebliche Skaleneffekte bei Beschaffung und Betrieb.

Der von Oxford Economics festgestellte Zwang zur Beschleunigung treibt deshalb zunächst den Umsatz mit Public-Cloud-Diensten in die Höhe. Doch ohne interne IT geht es nicht – sei es aufgrund von Datensicherheitsbedenken, Regularien, Vorgaben der Geschäftsleitung, oder weil sich die IT-Organisation nicht selbst wegrationalisieren will und deshalb den Fehdehandschuh von Amazon und Co. aufnimmt. Doch wie erreicht sie mit hauseigener Infrastruktur eine ähnliche Schnelligkeit und Agilität wie die Cloud-Giganten?



› Die Cloud-Giganten agieren deutlich schneller und effizienter als durchschnittliche IT-Abteilungen: Sie kommen mit einem Administrator für bis zu 20.000 Server aus – und das selbst bei täglich über 50 Updates. (Bild: HPE)

Eiswürfel statt Legosteine

Das Problem: Die Unternehmens-IT ist – wie die Oxford-Economics-Umfrage belegt – nicht auf DevOps und die kontinuierliche Provisionierung neuer IT-Services ausgelegt. Vielmehr ist sie historisch gewachsen und oft technisch und organisatorisch unterteilt in Server, Storage und Netzwerk. Zum Aufbau einer neuen Infrastrukturmgebung ist damit weiterhin die umständliche Konfiguration einer Vielzahl von Komponenten erforderlich.

Dabei hat die IT-Branche längst Alternativen zu den Silos entwickelt, zunächst mit konvergenten, später mit sogenannten hyperkonvergenten Systemen. Konvergente Systeme bündeln die Compute-, Storage- und Netzwerkressourcen zum Komplettsystem und erleichtern damit Provisionierung und Skalierung nach Art eines Legobaukastens. Ihnen folgten hyperkonvergente Systeme, die gemäß dem „Software-Defined“-Ansatz Funktionen wie den Storage Controller in die virtuellen Maschinen und damit auf Standardhardware verlagern. Ein derart virtualisierter Legobaukasten verschlankt Managementabläufe und erleichtert es so etwa, schnell eine Vielzahl virtualisierter Desktops aufzusetzen. Doch zugleich bleibt die IT immer an die vorgegebene Kombination von Hardware und Virtualisierungs-Software gebunden.

Um den Vergleich weiterzutreiben: Heutige IT-Infrastrukturen gleichen Baukästen, aus deren Bausteinen man in stunden- oder tagelanger Arbeit einen LKW oder Transporter assemblieren kann. Die Flexibilität, die im Zeitalter der agilen Softwareentwicklung gefragt ist, erfordert dagegen einen Pool von Server-, Storage- und Netzwerkressourcen, die zustandslos („stateless“) – ohne vorgegebene Be-

stimmung, Software-Image oder Netzwerk-Konfiguration – auf Abruf für den Betrieb neuer Workloads bereitstehen. Ein solcher Ressourcenpool muss es erlauben, die Kapazitätsbausteine automatisch und in Echtzeit zu kombinieren und zu konfigurieren – einschließlich einer ebenfalls automatisierten Skalierung, wie man das von der Public Cloud her kennt.

Benötigt wird also statt eines Baukastens vielmehr eine neue Art von Eiswürfelmaschine: Der Anwender drückt auf den Knopf und erhält seinen Eiswürfel nach Wunsch in Form eines 40-Tonnens, Transporters, Kleinwagens oder sportlichen Coupés. Je nach Bedarf ist der Eiswürfel ebenso rasch wieder eingeschmolzen und in eine andere Form gegossen – erst das ist „Hardware as a Service“.

Templates für den IaaS-Stack

Neben der Hardwarevielfalt ist die nicht minder große Fülle an Managementwerkzeugen ein weiteres Hindernis auf dem Weg zum Tempo der Public Cloud. Denn für jede Produkt- oder Hardware-Gattung gibt es eine separate Verwaltungslösung – Microsoft System Center, VMware vCenter, OpenNMS und so weiter – mit jeweils eigenen Fachleuten und Prozessen.

Zur Beschleunigung der Abläufe ist zwar der Einsatz von Templates verbreitet, doch die sind meist auf ein einzelnes Silo beschränkt. Die (hyper-)konvergenten Systeme vereinfachen zwar derartige Abläufe, aber immer nur im Rahmen konkreter Einsatzfälle, zum Beispiel zur Provisionierung virtueller Desktops. Zudem kommen sie häufig nicht anstelle von, sondern zusätzlich zu etablierten Umgebungen zum Einsatz. Die Komplexität bleibt so weitgehend erhalten.

Die Intelligenz, die in den Management-Tools und in den Köpfen der Mitarbeiter steckt, muss deshalb in die Infrastruktur selbst hineinwandern: Für echtes Infrastructure as a Service müssen die Templates den gesamten IaaS-Stack vom Server über Storage und Netzwerk bis hin zu Betriebssystem, Virtualisierungslösung und/oder Containertechnologie abdecken.

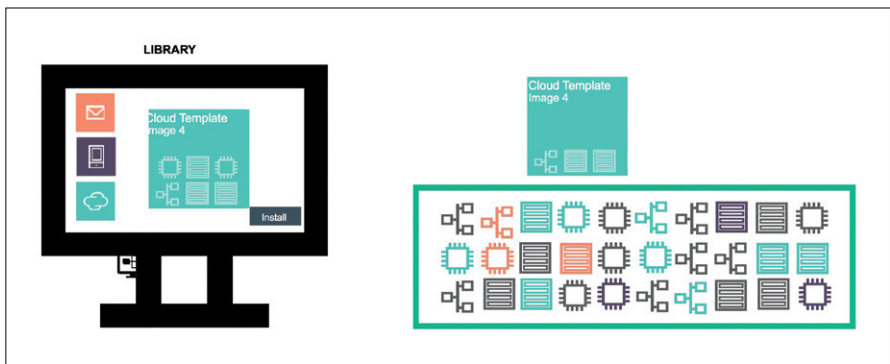
Wichtig ist dabei: Das Zusammenspiel der Ressourcen für Provisionierung und Skalierung muss Automatismen innerhalb der Infrastruktur überlassen bleiben. Diese muss selbsttätig erkennen, ob eine gewünschte Konfiguration widersinnig oder technisch unmöglich ist. So muss sie zum Beispiel dem Anwender bei der Erweiterung eines Virtualisierungs-Clusters die Auswahl der nötigen und passenden Ressourcen abnehmen und eine optimale Konfiguration anbieten. Dazu braucht man eine programmierbare IT-Umgebung, die Know-how über die Möglichkeiten und Grenzen der Programmierbarkeit bereits ab Werk mitbringt und den Anwender bei der Bereitstellung und Anpassung von IT-Ressourcen unterstützt.

Software-Defined-Infrastruktur

Um eine minutenschnelle Bereitstellung wie in der Public Cloud zu erreichen, muss man somit aufgabenspezifische Hardware und punktuelle Verwaltungswerkzeuge ersetzen durch eine programmierbare Infrastruktur. Für dieses Konzept etabliert sich derzeit der Begriff „Composable Infrastructure“. Er umschreibt eine „komponierbare“ – also beliebig zusammenstellbare – IT-Infrastruktur.

Wichtig ist es hier, die Komplexität der Programmierung auf einem Minimum zu halten. Auch traditionelle IT-Umgebungen lassen sich per Programmierung automatisieren, doch jedes Silo erfordert dedizierte Management-Werkzeuge – und jedes dieser Werkzeuge bringt seine eigene Programmierschnittstelle (API) mit, jeweils mit eigenem API-Format, Datenformat und Fehlercodeformat. Programmierbarkeit bedeutet dann das Aneinanderreihen all dieser Schnittstellen zu einer Kette mit einer Fülle von API-, Daten- und Error-Code-Formaten.

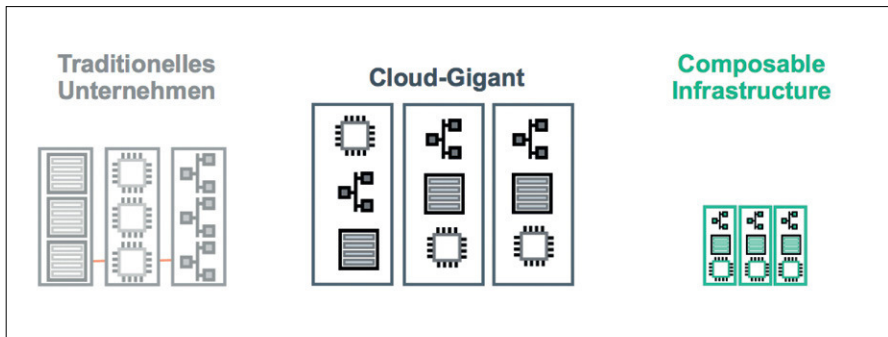
Um Server in drei Minuten bereitstellen zu können, muss eine programmierbare Infrastruktur diese Komplexität vor dem Anwender verbergen. Das Mittel der Wahl ist eine High-Level-API, die unterschiedlichste Sprachen (Java, Python, PowerShell, Go etc.) unterstützt. Sie bündelt alle Zugriffe auf darunter liegende APIs und gibt dem Anwender ein einheitliches API-, Daten- und Fehlercodeformat an die Hand. So kann dieser die Infrastruktur im Idealfall mit einer Zeile Code ansteuern, die dann zum Beispiel besagt: „Gib mir einen physischen Server mit zwei CPUs, 8 GByte RAM, 1 TByte Speicherplatz und Windows Server 2016.“



› Eine einheitliche High-Level-Schnittstelle muss alle benötigten Infrastrukturressourcen gemeinsam zugänglich machen. (Bild: HPE)

Cloud-Tempo statt IT der zwei Geschwindigkeiten

Die Cloud-Giganten erzielen durch den vollautomatisierten Betrieb von Millionen von Servern enorme Effizienz; allerdings müssen sie vorab enormen Aufwand mit Individualsoftware, speziell angepasster oder selbst assemblierter Hardware sowie hauseigenen DevOps-Prozessen betreiben. Management-Effizienz (und in der Folge hohes Tempo) ist immer definiert durch das Verhältnis des Verwaltungsaufwands zum Umfang der IT-Umgebung: Ist die Umgebung kleiner, muss das Management umso eleganter sein. Eine programmierbare Infrastruktur kann helfen, den IT-Betrieb à la Public Cloud auf die Größe einer Unternehmensumgebung herunter zu skalieren.



> Eine programmierbare Infrastruktur skaliert das Betriebsmodell einer Public Cloud herab auf die Größe einer Enterprise-Umgebung. (Bild: HPE)

Anders als die großen Cloud-Service-Provider müssen IT-Organisationen im Unternehmen, neben schneller Bereitstellung neuer Services, weiter den verlässlichen Betrieb der Bestands-IT bewältigen – jene Parallelbewegung der zwei Geschwindigkeiten, die Gartner „bimodale IT“ nennt.

Mittels der erwähnten einheitlichen High-Level-API muss sich eine programmierbare Infrastruktur deshalb mit beiden Welten verstehen: mit Cloud-Umgebungen wie OpenStack, Docker und Chef, ebenso wie mit den Management-Werkzeugen etwa von Microsoft und VMware. Ist diese Interoperabilität gegeben, lassen sich beide Welten der bimodalen IT zusammenführen: Dann profitiert auch die Bestands-IT von der Ressourcenbereitstellung in Cloud-Geschwindigkeit.