

Thomas Caspers, Helmut Reimer

# Hyperscaler – die Globalisierung von Cloud-Diensten

Das Thema Cloud ist nun schon seit mehr als 25 Jahren Gegenstand ausführlicher Auseinandersetzungen in der Informationstechnik. Wie so oft war die Startidee simpel: Die Netzinfrastruktur ließ es zu, Server nicht nur als physikalische Einheiten zu betreiben und auch zu vernetzen, sondern auch virtuelle Dienste anzubieten, die als Infrastrukturkomponenten zunächst Speicherplatz und Datenverarbeitungsleistungen zur Verfügung stellten. Dass damit Datenschutz und Datensicherheit vor neuartige Herausforderungen standen, ist plausibel. Waren doch die in traditionellen IT-Anwendungen geübten Standards auf diesen Gebieten mit hohem technischem Aufwand verbunden und über das Bundesdatenschutzgesetz (BDSG) rechtlich definiert. Es dauerte deshalb einige Jahre, bis auch Cloud-Dienste verfügbar wurden, die ausreichendes Vertrauen in die angebotenen Sicherheitsqualitäten erreichten. Trotzdem ist es so geblieben, dass die Entscheidung, sicherheitsrelevante Anwendungen in die Cloud zu verlagern, einer gründlichen und sensiblen Vorbereitung bedürfen. Vor diesem Hintergrund ist es verständlich, dass auch Cloud-Dienste nach wie vor hinterfragt werden und neue Konzepte für ihre Anwendungen entstehen. Nicht zuletzt ist das auch der nicht aufzuhaltenden Globalisierung der Netzinfrastrukturen geschuldet.

Das Spektrum sowohl von über das Internet erreichbaren als auch von Unternehmen und Behörden selbst betriebenen Fachverfahren in der Cloud ist in seiner Breite bereits heute immens groß. In der täglichen Nutzung von Informationstechnik sind bewusste, oftmals aber auch von Anwendenden unbemerkte Zugriffe auf solche Cloud-Dienste inzwischen der Regelfall. Diese Zugriffszahlen reichen aus der Perspektive des individuellen Cloud-Dienstes von wenigen Einzelfällen bis in mehrstellige Millionenbereiche. Die so angebotenen Dienste in den Cloud-Infrastrukturen müssen daher auf enormen Skalen funktionsfähig und verfügbar sein, sowohl in weitverteilten öffentlichen, „public“ Cloud-Angeboten als auch in den souverän betriebenen, „on premises“ Lösungen. Höchste Skalierungsanforderungen spielen im Cloud-Kontext eine zentrale Rolle. In der Terminologie hat sich dabei der Begriff der „Hyperscaler“ etabliert – wie oftmals im Bereich informationstechnischer Lösungen in seiner Entstehung mit einer unverkennbaren Marketingkomponente.

Die Bezeichnung Hyperscaler wird meist für sehr große Cloud-Anbieter verwendet. In ihrer unmittelbaren Bedeutung bezeichnen die Worte „hyper“ etwas übermäßig Großes oder stark Ausgeprägtes und „skalieren“ die Möglichkeit, einen Zustand um Größenordnungen zu variieren. Im Falle des Cloud Computings sind dies Ressourcen wie die bereitgestellte Rechenleistung, der verfügbare flüchtige und persistente Speicher oder langfristig stabile und äußerst breitbandige Netzzugriffe. Auch wenn diese Ressourcen insgesamt naturgemäß endlich sind, können sie Nutzen

de der Cloud für ihren Bedarf in der subjektiven Wahrnehmung nahezu beliebig erweitern, denn selten stoßen sie an die technischen Grenzen der Cloud. Eine nach eigenen Bedürfnissen erweiterte oder auch reduzierte Nutzung ist zu jedem Zeitpunkt vergleichsweise einfach möglich, indem software-gestützte Prozesse eingesetzt werden, die Cloud-Anbieter in sogenannten „Self Service Portalen“ oder über auch geeignete technische Schnittstellen zur Verfügung stellen.

Flexible Szenarien, wie sie vom Cloud Computing ermöglicht werden, sind in der klassischen Informationstechnik nicht oder nur sehr bedingt möglich. Um solche Szenarien überhaupt realisieren zu können, müssen Cloud-Dienste auf vollständig virtualisierten Server- und Storage-Systeme betrieben werden, die gleichzeitig in eine leistungsfähige Netzinfrastruktur eingebettet werden. Durch Standardisierung und Virtualisierung der Angebote auf Seiten der Betreiber können die Cloud-Dienste dann allerdings auch verhältnismäßig leicht an den Bedarf der Nutzenden angepasst und zwischen ihnen aufgeteilt werden. Dabei erfolgt die Zuweisung der in der intensiven Nutzung entstehenden Lasten auf die verfügbaren Ressourcen des Cloud-Dienstes im Idealfall gleichmäßig. Hyperscaler versuchen, mit überdurchschnittlich gut vernetzten Systemen neben einem großen Datendurchsatz und mehrfachen Redundanzen auch eine besonders hohe Verfügbarkeit und weitgehende Fehlertoleranz selbst für komplexe Dienste zu bieten.

In der fachlich präzisen Beschreibung von Cloud-Anbietern bleibt es jedoch bis heute bei begrifflichen Unschärfen: Es gibt keine fest definierten oder allgemein etablierten Werte, ab denen ein Cloud-Anbieter gemeinhin als Hyperscaler zu bezeichnen ist. Übliche Kriterien sind, dass Hyperscaler global oder zumindest in mehreren Kontinenten agieren und dabei lokale Regionen (z. B. Rechenzentren in der Europäischen Union) betreiben, die in Verfügbarkeitszonen aufgeteilt sind. Hyperscaler bieten in der Regel eine Vielzahl von unterschiedlichen Cloud-Diensten an, die in sehr großen und weltweit zumindest innerhalb der jeweiligen Anbieterinfrastruktur weitgehend standardisierten Rechenzentren erbracht werden. Aufgrund ihrer in einigen Weltregionen dominierenden Marktstellung wird der Begriff Hyperscaler in manchen Zusammenhängen allein für Amazon Web Services (AWS), Microsoft Azure und Google Cloud Platform (GCP) verwendet. Unter fachlich-technischen Gesichtspunkten fallen unter die oben genannten Kriterien jedoch u. a. auch Alibaba Cloud (Aliyun), IBM Cloud, Oracle Cloud und Tencent Cloud. Trotz dieser unter formalen Gesichtspunkten bestehenden Ungenauigkeiten, hat sich der Begriff Hyperscaler in fachlichen Diskussionen durchgesetzt und wird daher auch in den Beiträgen dieser Ausgabe in seiner Bedeutung für den öffentlichen Sektor eingeordnet.