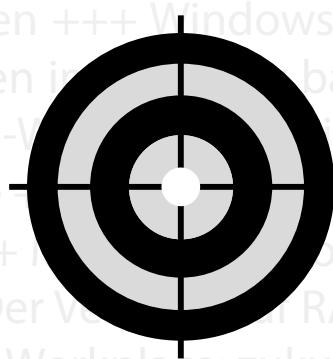


# BEST OF



# DataCenter Insider

## MANAGEMENT & STRATEGIE

Pfalzkom Manet und EMC HostCo schließen Datacenter-Kooperation .....	3
Höchste Zuverlässigkeit bei IBM und Lenovo, schlechteste Werte bei Oracle und HPE.....	9
Wie digital ist der TÜV? .....	12
Appetit auf Cloud? Der Fressfeind Edge-Computing.....	19
Jobkiller und Fachkräftemangel bestimmen die Suche nach Datacenter-Mitarbeitern .....	26
Handeln wie Google - Refurbishment im Datacenter .....	31
OpenStack Summit 2018, Teil1: Kritik? War da was?.....	34
Ist die Deutsche Cloud am Ende? .....	37
Audit-Trails mit IOTA, der Blockchain-Alternative .....	41
Das erste Rechenzentrum in einer Windmühle läuft und die Kundschaft ist happy.....	45
Blockchain-as-a-Service im Unternehmen nutzen .....	49
Erbarmungslos nahe - Der Verzicht auf RAID-Arrays .....	53

## TECHNOLOGIE & ZUKUNFT

Quo vadis Datacenter? .....	60
Open19 Foundation versus Open Compute Project (OCP).....	65
Haushoher Vorteil für Lithium-Ionen-Technik im Rechenzentrum .....	71
Krass: Das Missverhältnis zwischen dem Datacenter als Gebäude und der IT .....	77
Batterietechnik statt Diesel und neue Verfügbarkeitsklassen .....	81
Was ist Composable Infrastructure?.....	86
Datacenter in Schweden und in Deutschland - best Practices versus Ignoranz .....	92
KI: Treibstoff für die intelligente Zukunft der Geschäftswelt .....	98
Strittig: Welche Luftfeuchte sorgt für prima Klima? .....	103
Kubernetes und Cloud Foundry - Konkurrenz oder zwei Asse im Ärmel? .....	107
Was kommt nach den Containern? .....	112
Server gehen immer öfter baden.....	115

### IMPRESSUM:

Vogel IT-Medien GmbH  
 Max-Josef-Metzger-Straße 21  
 86157 Augsburg  
 Tel.: +49(0) 821-2177-0  
 Fax: +49(0) 821-2177-150  
 Email: [zentrale@vogel-it.de](mailto:zentrale@vogel-it.de)  
 Internet: [www.vogel-it.de](http://www.vogel-it.de)

Handelsregister Augsburg  
 HRB 1 19 43  
 Umsatzsteueridentifikationsnummer:  
 DE 127502716  
 Geschäftsführer: Werner Nieberle

### Inhaltlich Verantwortliche gemäß § 55 Absatz 2 RStV:

Nico Litzel, Florian Karlstetter, Ulrike Ostler, Stephan Augsten, Andreas Donner, Peter Schmitz, Rainer Graefen (Anschrift siehe Verlag)

### Vogel IT-Medien

Die **Vogel IT-Medien GmbH**, Augsburg, ist eine 100prozentige Tochtergesellschaft der **Vogel Communications Group**, Würzburg. Seit 1991 gibt der Verlag Fachmedien für Entscheider heraus, die mit der Produktion, der Beschaffung oder dem Einsatz von Informationstechnologie beruflich befasst sind. Dabei bietet er neben Print- und Online-Medien auch ein breites Veranstaltungspfotolio an. Die wichtigsten Angebote des Verlages sind: **IT-BUSINESS, eGovernment Computing, BigData-Insider.de, CloudComputing-Insider.de, DataCenter-Insider.de, Dev-Insider.de, IP-Insider.de, Security-Insider.de, Storage-Insider.de**.

### Vogel IT-Medien

Das Fachmedienhaus **Vogel IT-Medien GmbH** ist einer der führenden deutschen Fachinformationsanbieter mit rund 100 Fachzeitschriften und 60 Webseiten sowie zahlreichen internationalen Aktivitäten. Hauptsitz ist Würzburg. Die Print- und Online-Medien bedienen vor allem die Branchen Industrie, Automobil, Informationstechnologie und Recht/Wirtschaft/Steuern.

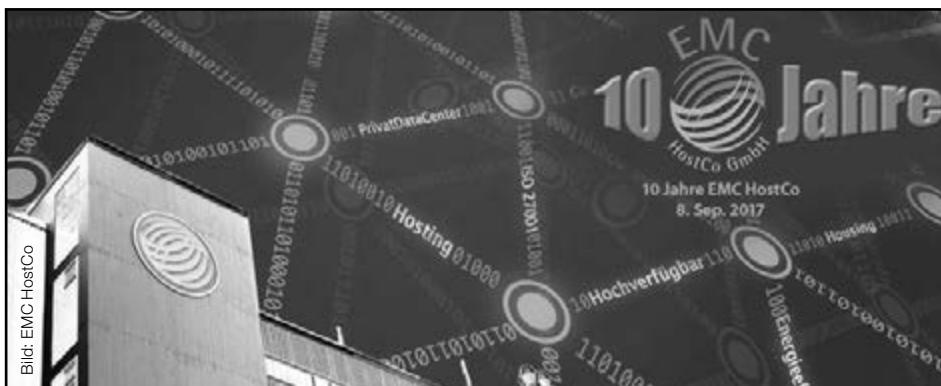
## **Georedundanz bringt Vorteile für Kunden und Anbieter**

# Pfalzkom Manet und EMC HostCo schließen Datacenter-Kooperation

**Pfalzkom Manet und EMC HostCo GmbH bieten hochverfügbare Rechenzentren, das zuerst genannte Unternehmen in der Region Rhein-Neckar und das zweite in München. Durch die räumliche Trennung der Datacenter werden Services für Kunden realisierbar, die zusätzlichen Schutz für besonders unternehmenskritischen Daten und Anwendungen ermöglichen. Das Warum und Wieso erläutert Bernhard Huter, Chef von EMC HostCo, im Interview.**

Im Hinblick auf die aktuell Einzug haltenden Cloud-Thematiken und die immer mehr in den Fokus rückenden persönlichen Haf- tungsrisiken der Geschäftsführung von Unternehmen sind Lösun- gen zu Themen der erweiterten und erhöhten Betriebssicherheit, und dazu zählen im Besonderen Backup- und Redundanzrechen- zentrumslösungen, unumgänglich. Gerade Finanzdienstleister und Banken, Wirtschaftsprüfer und Berater, Rechtsanwaltsfir- men, Produktionsunternehmen, Ingenieure, Chemie-/Pharmaun- ternehmen und Behörden betrifft dies verstärkt.

Das Lösungsspektrum umfasst unter anderem Aktives (Hot) oder Cold-stand-By Rechenzentrum mit hohem Revisionsabstand (rund 400 Kilometer), Business Continuity Rechenzentrum, Dual-Rechenzentrum-Lösung zur Netzwerkperformance-Optimierung und Niederlassungsanbindung, Backup-Rechenzentrum als Rack, Cage oder Hochsicherheitslösung und Online-Backup-Lösungen.



**Der Münchener Betreiber eines Rechenzentrums EMC HostCo konnte im September dieses Jahres sein 10jähriges Bestehen feiern. Zugleich gab der Anbieter von Datacenter-Diensten die Kooperation mit Pfalzkom Manet bekannt.**

## Georedundanz im Cloud-Zeitalter

### **CIOBRIEFING: Was verstehen Sie denn unter Georedundanz?**

**Bernhard Huter:** Echte Georedundanz ergibt sich nicht, wenn man ein Rechenzentrum in lediglich 2 oder 10 Kilometern Abstand spiegelt. Meiner Ansicht nach müssen 300 bis 400 Kilometer zwischen den Standorten liegen und komplett getrennte Netze.

### **CIOBRIEFING: Am 8. September dieses Jahres feierte EMC HostCo GmbH das 10-jährige Firmenjubiläum, in einem Jahr also, in dem Sie die RZ-Fläche verdreifachen. Warum gehen Sie diese Kooperation ein? Sind die Kunden trotzdem unzufrieden?**

**Bernhard Huter:** Als regionaler Anbieter von Rechenzentrumsdiensten analysieren wir die Marktentwicklungen der gesamten IT-Branche. Dazu gehört, dass durch die immer stärker werdenden Clouds und die daraus zu erbringenden Dienstleistungen eine für alle RZ-Betreiber erweiterte und in Teilen neue strategische Herausforderung entstehen.

Wir haben uns innerhalb unseres Netzwerkes dazu etwa mit unserem Partner RFConsult ausgetauscht und gemeinsam unsere Zielrichtung zum Thema Georedundanz entwickelt. RFConsult verfügt über fast 20 Jahre Erfahrung im Bereich international agierender RZ-Betreiber und globaler Kundenlösungen. Bald standen unser Georedundanz-Vorstellung fest, vom Konzept bis zu den verschiedenen Produktausprägungen.

### **CIOBRIEFING: Und wie kamen Sie auf Ihren Geo-Partner, Pfalzkom Manet?**

**Bernhard Huter:** Die hatten dasselbe Beratungsunternehmen, RFConsult. Das brachte den Stein ins Rollen.

### **CIOBRIEFING: Doch wie löst das die Konfrontation mit den nicht immer klar definierten Cloud-Thema?**

**Bernhard Huter:** Aktuell und in den kommenden Jahren werden Unternehmen Cloud-Lösungen sicherlich verstärkt einsetzen. Die großen Software- und Internet-Giganten haben nicht umsonst mit Milliardenaufwendungen weltweit Cloud-Rechenzentren und Lösungen aus dem Boden gestampft.

Unsere Analyse hat ergeben, und da ist man sich auch im Cloud-Markt inzwischen einig, dass Kunden ihre Anwendungen und Prozesse in der Cloud-Welt sicherlich neu strukturieren müssen. Wir glauben, die kommende Lösung wird mit Sicherheit eine hybride Lösung sein. Denn selbst wenn inzwischen viele Unternehmensanwendungen Cloud-fähig sind, wird es noch lange, je nach Branche, mehr oder weniger Anwendungen geben, die nicht in der Cloud lauffähig sind oder schlichtweg nicht in die Cloud gehören.

Unternehmenskritische und lebenswichtige Anwendungen werden unter 100-prozentiger Hoheit des Kunden in seiner dedizierten Private Cloud laufen. Dazu braucht er eigene Rechenzentrumslösungen, idealerweise von einem professionellen Rechenzentrumsbetreiber gemanagt und betrieben.

Weniger kritische Anwendungen werden, je nach Ausprägung, bei einem der großen Public-Cloud-Anbieter genutzt werden und mit der Private Cloud des Kunden verbunden.

Eine so genannte Hybrid Cloud ist kein eigenständiger Cloud-Typ sondern beschreibt vielmehr die Kopplung herkömmlicher IT-Infrastrukturen vor Ort im Rechenzentrum des Kunden und den Ressourcen von Public oder Private Clouds um die Vorteile aller Infrastrukturtypen ideal nutzen zu können.

### **CIOBRIEFING: Und welche Rolle wollen Sie als mittelständischer Anbieter von Rechenzentrumsdiensten einnehmen?**

**Bernhard Huter:** Nun, da Kunden in Zukunft eine stark fragmentierte Applikations- und Dienstleister-Infrastruktur haben werden, haben wir unseren Fokus auf Lösungen für seine Private Cloud gelegt. Wir wollen als „Home of Housing“ für unsere Kunden verstanden werden. Bei uns wird der Kunde auch in Zukunft seine unternehmenskritischen Anwendungen sowie seine nicht-Cloud-fähigen Anwendungen hochausfallsicher in seiner dedizierten Private Cloud betreiben.

Außerdem bezeichnet Cloud Computing ja das dynamisch an den Bedarf angepasste Nutzen von IT-Diensten über ein hochperformantes Netz. Und das ist der zweite Teil unserer Antwort, wir müssen dem Kunden die bestmögliche Voraussetzung schaffen, seine unterschiedlichen Cloud-Bedürfnisse nach seinen individuellen Vorstellungen und Geschäftserfordernissen performant und ausfallsicher zu verbinden.

Dieser Punkt stellt völlig neue Anforderungen an Redundanzstandorte des Rechenzentrumsbetriebes, an eine echte Geo-Redundanz.

## **CIOBRIEFING: Der Begriff wird in Deutschland durchaus anders betrachtet als zum Beispiel in den USA oder Asien.**

**Bernhard Huter:** Richtig, bis heute denken viele Unternehmen und auch Beratungsunternehmen nur in Revisionsabständen von 15, 30 oder 100 Kilometer. Dies ist historisch bedingt. Im Cloud-Zeitalter aber gibt es nicht mehr primär die Probleme mit Laufzeiten bei Rechenzentrumskopplungen, es stellen sich völlig neue Anforderungen an einen hochverfügbaren Rechenzentrumsbetrieb.

## **CIOBRIEFING: Welche Anforderung meinen Sie da?**

**Bernhard Huter:** Die Hochverfügbarkeit der Internet-Verbindung. Die Verbindung mit dem Internet ist eines der wichtigsten Teilstücke einer Cloud, da ohne sie keine Dienste im Internet angeboten werden können. Entsprechend wichtig ist wie beim Aufbau der Netzwerkinfrastruktur auch die Implementierung einer gewissen Fehlertoleranz oder Robustheit.

## **CIOBRIEFING: Was bedeutet das konkret?**

**Bernhard Huter:** Für den Betrieb einer Cloud sollte in Betracht gezogen werden, mehrere voneinander unabhängige Internet-Anbindungen gleichzeitig zu nutzen. Hierdurch ergeben sich nach unserer Meinung folgende Vorteile:

Keine Abhängigkeit von einem einzelnen Internet-Provider. Daher wird auch hier unser Konzept der Netzwerk-Neutralität zukünftig ein wichtiger Baustein bleiben.

Die IP-Verbindung kann immer auf der jeweils schnellsten Internet-Verbindung realisiert werden.

Mögliche Vermeidung von Einschränkungen auch bei Großstörungen im Telekommunikationsnetz.

Aus diesem Grund verstehen wir ähnlich wie die großen Public-Cloud-Dienstleister unter Geo-Redundanz auch 300 bis 400 Kilometer als notwendigen Revisionsabstand. Schauen Sie: Mit zwei Redundanzstandorten in der näheren Umgebung im Großraum München haben Sie selbst bei mittleren Beeinträchtigungen der Netzketten nichts gewonnen. Beide RZ-Standorte sind davon betroffen – Ihre Cloud-Infrastruktur muss ja nicht gleich komplett ausfallen. Es reicht ja, wenn bestimmte Funktionalitäten nicht mehr verfügbar sind. Schon hängt das Zusammenspiel.

Daher haben wir nach einer Lösung gesucht, in der eine Vielzahl der Netze getrennt verläuft und bei einem Standort-Ausfall das andere RZ nicht zwangsläufig davon betroffen ist.

## CIOBRIEFING: Und Pfalzkom Manet ist dafür der richtige Partner?

**Bernhard Huter:** Wir sind ein regionaler, inhabergeführter, mittelständischer RZ-Betreiber, auch wenn wir hier einige 1.000 Quadratmeter Datacenter-Fläche aktiv betreiben. Daher war es für uns wichtig einen RZ-Partner zu finden der zu uns passt, unsere Prozesse und Arbeitsweise versteht und ähnlich flexibel wie wir agieren kann und will.

In Pfalzkom Manet haben wir einen Partner gefunden, der alle Kriterien erfüllt. Der Standort ist weit von München entfernt, hat direkte Anbindung an den Haupt-Internet-Knoten in Deutschland und betreibt zudem einen autarken Internet-Knoten-Standort. Das Rechenzentrum entspricht der Klasse Tier III, das Unternehmen ist ein Mittelständler wie wir.

Für unsere Kunden ist es wichtig, nicht die Kunden-Nr. 917 bei einem global agierenden Unternehmen oder einer großen Aktiengesellschaft zu sein. Flexibilität und Zusammenarbeit auf Augenhöhe ist für sie entscheidend.

Klar müssen wir uns auch ein stückweit öffnen und unsere Prozesse und Dienste offenlegen. Das verlangt Vertrauen in unserer Partner. Doch nur so können wir unsere Prozesse harmonisieren, damit der Kunde das Gefühl hat, mit nur einem Unternehmen zu arbeiten.

Die Notwenigkeit für eine solche Kooperation war mir sofort klar. Im Übrigen sehe ich das auch als große Chance für Mittelständler wie uns. Wenn wir Georedundanz-Services nicht anbieten, übernehmen die großen Wettbewerber das Geschäft.

## CIOBRIEFING: Bedeuten zwei Anbieter doppelte Kosten für Ihre Kunden?

**Bernhard Huter:** Nein. Zunächst benötigen sie nicht an zwei Standorten für alle Bereiche Hochsicherheit. Zudem müssen wir die Prozesse so harmonisieren, dass für die Kunden kein doppelter Aufwand entsteht. Aber: Sie haben Recht; für uns entsteht höherer Aufwand, aber den sollen die Kunden weder im Preis noch im Doing deutlich spüren. Vermutlich schmälert es unsere Marge ein wenig.

## CIOBRIEFING: Welche Services wollen Sie überhaupt durch die Georedundanz bieten?

**Bernhard Huter:** Natürlich alle Anforderungen und Konzepte im Bereich Cloud-Rechenzentrumsbetrieb und hochverfügbarer

unabhängiger Netzanbindungen. Darüber hinaus können wir aber über diese zwei Standorte auch alle K-Fall Szenarien und Bedürfnisse von Unternehmen umsetzen - ganz gleich welche Anforderungen gestellt werden und ob der Kunde nach den Vorgaben des IT-Grundschutzes arbeitet. Ob Backup-Rechenzentrum, Hot-/Cold-Stand-By-Rechenzentrum oder Notfall-Datacenter mit Notfall-Arbeitsplätzen .... Egal, welche Anforderungen der Kunde hat, sie lassen sich damit umsetzen.

**CIOBRIEFING: Das ist ja eine ganze Menge an Themen. Können Sie diese nicht standardisieren?**

**Bernhard Huter:** Nein, leider nicht, oder sollte ich sagen Gott sei Dank?

Lieber und sinnvollerweise setzen wir uns mit dem Kunden und den Fachverantwortlichen zusammen und erarbeiten ein gemeinsames Konzept zur Umsetzung. Standard-Lösungen greifen da einfach zu kurz und stellen nach unserer Meinung somit keine ideale Umsetzung der hybriden Geo-Redundanzstrukturen dar. Und das verstehen wir nicht als unseren zu erbringenden Dienst am Kunden.

**CIOBRIEFING: Ist die Kooperation mit Pfalzkom Manet exklusiv?**

**Bernhard Huter:** Wenn es für eine neue geschäftliche Verbindung Kunden gibt, könnten sich die Georedundanz ausbreiten.

■ *Ulrike Ostler*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



**Hier klicken**

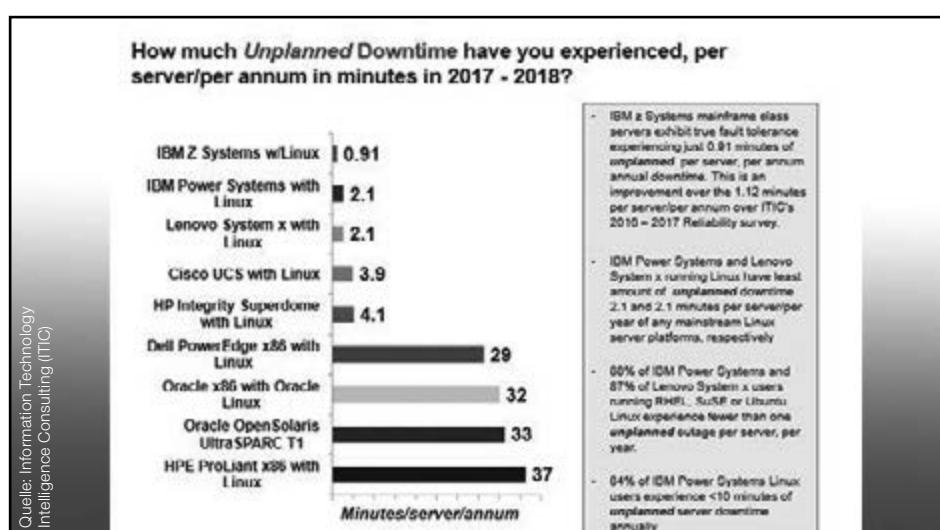
## Global Server Hardware and Server OS Reliability

# Höchste Zuverlässigkeit bei IBM und Lenovo, schlechteste Werte bei Oracle und HPE

Alljährlich befragt Information Technology Intelligence Consulting Corp. (ITIC) 800 Organisationen weltweit zur Zuverlässigkeit ihrer Server und Server-Betriebssysteme. Um möglichst unvoreingenommene Ergebnisse zu präsentieren zu können, akzeptiert das Unternehmen kein Sponsoring von Anbietern. Auch im Report 2017/2018 liegen IBM und Lenovo vorne – zum zehnten Mal hintereinander – Oracle und „HPE Proliant“ nicht.

Nach Ergebnissen des „2017 – 2018 Global Server Hardware and Server OS Reliability“ liefern die „Power“-Systeme 7 und 8 von IBM und die „X6“-x86-Server von Lenovo im Vergleich zu 14 Hardware-Konkurrenzsystemen, 18 Betriebssystemen sowie elf verschiedenen Hardware-Virtualisierungsplattformen eine 17 bis 18 Mal höhere Betriebszeit und Zuverlässigkeit als so manches System, mit denen ITIC sie verglichen hat.

Insgesamt kamen sie im Schnitt auf gerade einmal 2,1 Minuten ungeplante Ausfallzeit pro Jahr. Am schlechtesten kamen die Systeme von Oracle sowie die „HPE Proliant“-Rechner weg. Außerdem zeigt die Untersuchung, dass überhaupt nur 1 Prozent der Power-Systeme und der Lenovo-Server mehr als vier Stunden ungeplanter Downtime pro Jahr aufweisen.



Zuverlässige Systeme können ungeplante Downtime verhindern. Das zeigt die Befragung von Information Technology Intelligence Consulting deutlich. Die Studie ermittelt nicht nur, welches System sich empfiehlt und von welchem die Finger lassen sollten, sondern auch warum.

## Eine Klasse für sich

Gefragt hat ITIC auch nach den Mainframes von IBM: „Eine Klasse für sich“, urteilt der Report. „System Z“ kommt im Schnitt auf weniger als 10 Sekunden ungeplanter Ausfallzeit pro Monat. Auf mehr als vier Stunden ungeplanter Downtime pro Jahr kommt gerade einmal ein halbes Prozent der befragten Mainframe-Nutzer.

Ebenfalls gut im Rennen lagen „Cisco UCS“, „Fujitsu Primequest“, „HPE Integrity“ und „Huawei Kunlun“. Insbesondere Server von Dell EMC, HPE und Oracle, die älter als drei Jahre sind, zeigen zunehmende Schwächen. So fielen 16 Prozent der HPE-Proliant- und der „Poweredge“-Server von Dell EMC jährlich mehr als vier Stunden ungeplant aus.

Weiterhin gab kein einziger der Befragten an, dass bei den Lenovo- und IBM-Systemen die Verfügbarkeit unter 99,9 Prozent gelegen habe, während nur gerade einmal 1 Prozent der HP(E)-Server-Nutzer eine Drei-Neuner-Uptime (99,9 Prozent) bestätigen konnte – das entspricht einer Server-Downtime von 8,76 Stunden pro Server und Jahr. Etwa 6 Prozent der Oracle-Server und 10 Prozent der Dell(EMC)-Systeme erreichen die 99,9 Prozent.

## Die Betriebssysteme

Die jüngste Umfrage unterstreiche auch die Zuverlässigkeit der Linux-Betriebssysteme, so der ITIC-Report. Denn sowohl IBM als auch Lenovo nutzen das Open-Source-Betriebssystem: In 88 Prozent der IBM Power Systems und 87 Prozent der Lenovo-X-System sind die Rechner, mit Red Hat Linux Enterprise (RHLE), Suse Linux Enterprise (SLES) oder Ubuntu laufen, verlässlicher. In Gänze setzen die meisten Power-Kunden „AIX“ ein.

Zu den Bedingungen, die sich ITIC genauer angeschaut hat, sind die Herausforderungen, die die Maschinen und ihre Ausstattungen für die neuen Anwendungen erfüllen müssen: Datenanalysen, Künstliche Intelligenz (KI), Maschinelles Lernen (ML) sowie Security und Management. zum Beispiel haben die Studienmacher nach dem Patch-Management gefragt. Denn die steigende Komplexität der IT, verlangt nach einfacherer Handhabung der Systeme. Immerhin sind auch nach der ITIC-Befragung hauptsächlich menschliche Fehler, Sicherheit und Komplexität die Hauptgründe für die ungeplanten Ausfälle. Die Hersteller der zuverlässigsten Systeme hätten in die genannten Bereiche heftigst investiert.

Erfreulich gute Noten in punkto Verfügbarkeit bekommen auch Fujitsu Primequest und der Newcomer Huawei Kunlun. Eine gute

Entwicklung sieht das Beratungsunternehmen ITIC aber auch für Cisco UCS und HPE Integrity Superdome. Denn diese Systeme schneiden bei jeder Nutzerbefragung besser ab. So zeigten 2 Prozent der Primequest- und drei Prozent der UCS-, Integrity und Kunlun-Server mehr als vier Stunden ungeplante Auszeit pro Jahr.

## Abwärts

Dagegen geht es abwärts mit HP Proliant und Oracle Sparc- und x86 Hardware sowie mit der Zuverlässigkeit von „Solaris“ – und zwar das sechste Jahr in Folge. Jedes der Systeme verursachte 33 beziehungsweise 37 Minuten Downtime im Jahr. Bei Oracle spielt eine Rolle, dass die Systeme relativ alt sind und die Kunden einen Wechsel zu anderen Plattformen verweigern.

Die komplette Studie „Global Server Hardware and Server OS Reliability“ gibt es auf der Website der Information Technology Intelligence Consulting Corp. zur Ansicht. ■ *Ulrike Ostler*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



**Hier klicken**

## Interview mit Dirk Stenkamp, Vorstandsvorsitzender TÜV Nord

# Wie digital ist der TÜV?

**Im Januar dieses Jahres forderte der Chef des Allianz-Sachversicherungsgeschäfts in Deutschland, Joachim Müller, „Treuhänder für Autodaten“. In Frage käme etwa der TÜV. Tatsächlich bekräftigte Dirk Stenkamp, Vorstandsvorsitzender der TÜV Nord Group, auf dem „TSI.Day“ in der vergangenen Woche, dass diese Treuhänderschaft ein angestrebtes Ziel sei. Wie qualifiziert, also: wie digital aber ist der TÜV Nord selbst?**

Jüngst machte der TÜV Süd mit neuen Cloud-Angeboten von sich reden, etwa zur Verarbeitung von Daten aus Hauptuntersuchungen von Kraftfahrzeugen, den Betrieb von Systemen zum Building Information Modeling oder Verarbeitung von Daten für Predictive Maintenance. Die Digitalisierung der eigenen Angebote ist offensichtlich.

### CIOBRIEFING: Wie digital ist der TÜV Nord?

**Dirk Stenkamp:** Der TÜV Nord muss digital werden, weil unsere Kunden diesen Weg gehen. Das betrifft sowohl das Prüfen und das Zertifizieren als auch verstärkt die Beratung in punkto Sicherheit – weltweit.



Bild: Ulrike Ostler/ Vogel IT-Medien GmbH

Anlässlich des „TSI.Day“ der TÜVit GmbH, einem Unternehmen der TÜV Nord Group, in der vergangenen Woche stellte sich der Vorstandsvorsitzende des TÜV Nord, Dr. Dirk Stenkamp, einem Interview über den Umgang der Organisation mit den Herausforderungen durch die Digitalisierung.

Das betrifft zunächst einmal die Digitalisierung der Produkte selbst, die funktionieren müssen, auch wenn der Betreiber und der Prüfer nicht selbst vor Ort sein kann, zum Beispiel bei der digitalen Aufzugsprüfung. Dazu kommen neuartige Prüfungen, wie die Drohnen-gestützte Inspektion oder ein 24x7-Monitoring von Sensoren – etwa in Aufzügen und Autos.

Während früher isoliert einzelne Objekte geprüft wurden, jeder kennt die Hauptuntersuchung bei Fahrzeugen, sind die Prüfobjekte heute selber komplex und zunehmend vernetzt. Neben der physischen Produktprüfung auf Gefahrenpotenzial untersuchen wir heute auch auf Gefährdungen durch Cyber-Angriffe.

Sehen Sie sich das Beispiel „Smart Gas“ von Linde an. Die Gasflasche mit Kohlefaserummantelung, die zu den Composite-Flaschen für Schweißgase gehört, besitzt etwa ein elektronisches Display und Sensoren. Das Display zeigt den Typ und die Art und Weise, wie das Gerät funktioniert, das Gasvolumen im Zylinder sowie die Durchflussmenge und verbleibende Betriebszeit. Schließlich ertönt ein Alarm, wenn der Zylinder fast leer ist.

Für die TÜV-Prüfer ergeben sich somit viele Prüfpunkte bis hin zur Cyber-Security, damit aus einem einfachen Gasbehälter keine Bombe wird.

Diese Herausforderungen implizieren, dass wir als TÜV Nord uns selber transformieren, etwas durch neue Aufgaben und Anforderungen an die Gutachter. Und der TÜV hat einen gesellschaftlichen Auftrag zu erfüllen. Denn Digitalisierung findet nur Akzeptanz, wenn sie sicher ist.

**CIOBRIEFING: Das ist noch nicht die Antwort auf die Frage, wie digital der TÜV Nord selber ist. Als disruptive Techniken führen sie selbst neben Künstlicher Intelligenz etwa Blockchain an. Inwieweit würden Blockchain-Verfahren zentrale Treuhänder für Sicherheit erübrigen?**

**Dirk Stenkamp:** Lassen Sie mich Blockchain zunächst einmal entmystifizieren. Es handelt sich dabei um eine verteilte Datenbank. Jeder Teilnehmer verfügt im Prinzip über das gesamte Datenbankwissen. Das Überweisen eines Geldbetrags wird also vor den Augen aller anderen getätigt, die an einer solchen Transaktionskette teilhaben – allerdings ohne die Identität von Sender und Empfänger zu kennen. Insofern ist eine Blockchain in sich selbst konsistent; denn jeder, der zuschaut, kann die Weitergabe der Informationen auch verifizieren, eine Quittung ausstellen.

Wir als TÜV haben keine Anteile an irgendeiner Blockchain. Allerdings prüfen wir die jeweilige Algorithmik auf Fehleranfälligkeit. Schließlich ist auch die Kryptologie anfällig für Angriffe.

### **CIOBRIEFING: Sehen Sie Potenzial für Blockchain-Anwendungen im eigenen Haus?**

**Dirk Stenkamp:** Tatsächlich könnten fälschungssichere Zertifikate eines der Anwendungsgebiete sein. Das erleben wir häufig, dass Zertifikate, die wir ausgegeben haben, unberechtigter Weise verlängert werden. Indem das Produkt eindeutig mit einem Zertifikat korreliert und das in einer verteilten Datenbank dokumentiert ist, ganz im Sinne von smart contracts, könnte sich etwa das Nachschauen, ob das Zertifikat noch gültig ist, automatisieren lassen.

Das gehört zu den Bereichen TÜV-spezifischer Prozesse, die wir im Innovation Space, Hamburg, aktiv erforschen, bis hin zu einem Roll-out-Programm. Das bedeutet jedoch Jahre an Investitionen, in die vermutlich auch Partner eingebunden werden. Zudem handelt es sich um ein internationales Thema, da ein globales einheitliches System Sinn macht.

**CIOBRIEFING: Sie fordern internationale Normierung von Datenstandards und Schnittstellen für Industrie 4.0. Denn die Prozessautomatisierung erfolge zu langsam. Insbesondere fehlten verbindliche ITK-Sicherheitsstandards, welche den Stand der Technik abdecken. Mit Ausnahme des IT-Sicherheitsgesetzes für kritische Infrastrukturen (KRITIS) existieren noch keine verbindlichen gesetzlichen Grundlagen. Sie sagen: Deutschland drohe „abgehängt“ zu werden. Doch sind Standards nicht eher ein Hemmschuh? Denn bis es solche gibt, hat sich die Welt schon weitergedreht. Ohnehin schaffen Google & Co. Defacto-Standards. Wer schnell sein will, gehorcht diesen.**

**Dirk Stenkamp:** Tatsächlich gibt es durchaus einen „War of Standard“, einen Kampf der Marktgiganten um die Übermacht. Denn wer die Standards bestimmt, gewinnt Geld, Einfluss, den Markt. Was tatsächlich passiert ist, dass dieser „Krieg“ die Innovation hemmt, die Weiterentwicklung verlangsamt. Müssen die „Zulieferer“ sowohl für „Google Home“, „Apple Home Kit“, „Invoke“ und „Amazon Echo“ unterstützen, steigt der Aufwand exponentiell.

tiell, gegenüber einer genormten Smart-Home-Schnittstelle. Und nach dem Aufpoppen unterschiedlicher Industriestandards folgt zwangsläufig ein Konsolidierungsprozess.

Was ich unter Standard verstehe, und das lässt sich auf alle Industrien übertragen, ist eine Kompatibilität von digitalen Schnittstellen, nicht etwa Uniformität der Produkte. Aber die internationale Normierung von Datenstandards und Schnittstellen für Industrie 4.0 und Prozess-Automatisierung erfolgt zu langsam.

**CIOBRIEFING: Wie wollten Sie die ITK-Sicherheit definieren? Schon immer hecheln die aktuellen Bestimmungen und Gesetze den Malware- und Cyber-Kriminellen hinterher?**

**Dirk Stenkamp:** Das Problem besteht darin, dass anders als bei personenbezogenen Daten (Datenschutzverordnung) weder einen gesetzlichen Exklusivitätsanspruch noch ein Urheberrecht auf Maschinendaten existiert. Das aber resultiert in hohen Unsicherheiten bei Herstellern und Verbrauchern über die Nutzung und die Verkettung von Daten.

Wir erkennen vier Grundelemente der IT-Sicherheit. security by design, privacy by default, intelligent gateways und secure elements. In Bezug auf Security by Design gibt es derzeit lediglich eine Selbstverpflichtung der Hersteller, die IT-Sicherheit in der Produkt-Entwicklung. Es fehlt ein verbindliches Regelwerk. Privacy by Default würde bedeuten, dass die Geräte-Werkseinstellung Privatsphäre und Sicherheit garantiert; auch Auto-Updates brächten etwa kein unbewusstes „Sharing“.

Die Gateway sollten als intelligente Informationsverteiler private und öffentliche Daten trennen und entsprechend abgestuft den Datenzugriff regeln können. Schließlich braucht es spezielle Hardware-Komponenten, die sicherheitskritische Funktionen ausführen, zum Beispiel Verschlüsselung bei Bezahlvorgängen per Smartphone. Die vier Grundelemente müssen mit einer unabhängigen Drittprüfung verbunden und obligatorisch für jedes neue digitale Produkt werden, das in Deutschland entwickelt, hergestellt oder vertrieben wird.

**Der TÜV Nord strebt eine Treuhänderschaft etwa für die Datenfülle an, die Autos erzeugen. Sie werden also zum Cloud-Anbieter mit eigenen Rechenzentren.**

**Dirk Stenkamp:** Wir betreiben seit Mai 2015 ein Hochsicherheits-Datacenter in Hannover, wo alle Systeme übrigens per KI-

Tool kontinuierlich überwacht werden. Denn wir wollen der Herr über unsere Daten sein und sie entsprechend schützen. Immerhin geht es zum Teil um hochsensible Informationen, zum Beispiel aus Kernkraftwerken.

Neu ist das Konzept nicht. Der TÜV war schon immer auch ein Datensammler, früher ging es um analog, sogar handschriftlich erfasste Daten, heute geht es noch immer um Prüfberichte aber auch um Sensordaten.

### **CIOBRIEFING: Wie lange kommen Sie noch mit der Rechenzentrumskapazität aus?**

**Dirk Stenkamp:** Das KI-Tool ermöglicht Analyse und Prognosen der Leistungsdaten und Vorhersagen über das voraussichtliche Wachstum des zukünftigen Datenvolumens. Demnach ist das Rechenzentrum so dimensioniert, dass alle Anforderungen bis mindestens 2030 erfüllt werden können.

Zugleich kommen uns neue Technologien zu Gute. So konnte durch Virtualisierung und konsequente Umstellung der Systeme auf eine Blade-Server-Architektur der Flächenbedarf in den Server-Räumen von jeweils ursprünglich geplanten 600 Quadratmetern auf 300 Quadratmeter reduziert werden.

Die heutige Festplatten- und Archivtechnologie hat sich ebenfalls in den letzten Jahren deutlich gewandelt. Das Speichervolumen je Festplatte hat sich in den letzten 10 Jahren um den Faktor 100 gesteigert. Gleichzeitig haben wir die Speichersysteme konsequent auf SSD-Technologie umgestellt. Auch die Zeit der großen Band- oder CDROM/WORM-Systeme für die Archiv- und Backupsysteme zur Kurz- und Langfristarchivierung sind endgültig vorbei.

Der TÜV NORD hat auf seinen aktiven Laufwerken heute ein Datenvolumen von mehr als 1,2 Petabyte mit einem durchschnittlichen Wachstum von 20-30 Prozent pro Jahr mit steigender Tendenz. Wir archivieren heute nur noch in einem zertifizierten Plattsystem basierend auf SSDs-. Die Backups der Server werden in einem Snapshot-Verfahren gesichert. Server lassen sich so im Minutentakt wiederherstellen.

Somit planen wir derzeit auch keine weiteren Rechenzentren an anderen Standorten. Sollten jedoch weitere Kunden Interesse an einem Betrieb ihrer Systeme durch den TÜV Nord haben, liegen die genehmigungsfähigen Pläne für einen weiteren Rechenzentrumsbau bereits fertig vor und könnten direkt eingereicht werden. Derzeit planen wir mit einer Vorlaufzeit von rund 18 Monaten.

**CIOBRIEFING: Der TÜV, und damit der TÜV Nord, ist eine Institution, die Gewicht hat. Müssten Sie nicht eine aktiver Rolle einnehmen, um die Digitalisierung anzuschieben.**

**Dirk Stenkamp:** Das tun wird ja. Als Verband der TÜV-Organisationen haben wir ein sehr gutes Positionspapier hinterlegt: „Politische Empfehlungen des VdTÜV für den 19. Deutschen Bundestag (2017–2021)“.

Unter anderem sehen wir den 5G Mobilfunk als Voraussetzung für Industrie 4.0 insbesondere den Bergbau 4.0 und um das autonome Fahren zu realisieren. Wir brauchen für die Maschine-zu-Maschine-Kommunikation Latenzen unter 1 Millisekunde; denn das ist mit dem jetzigen Mobilfunk nicht zu machen.

Außerdem muss der Breitbandausbau vorangetrieben werden.

**CIOBRIEFING: Das fordern immer alle. Doch gibt es Stimmen, dass der Wettlauf mit dem Ausland schon längst verloren ist und Deutschland soweit im Rückstand, dass das nicht mehr aufzuholen ist. Und was ist mit der Ansiedlung von Rechenzentren und dem Wettbewerbsnachteil durch die hohen Strompreise hierzulande?**

**Dirk Stenkamp:** Es ist die Aufgabe der neuen Bundesregierung, entsprechende Anreizsysteme zu schaffen, etwa in Bezug auf die Energiekosten. Doch ich will in das Schwarzsehen nicht unbedingt einstimmen. Auch die Hannover Messe 2018 beweist wieder, dass Deutschland das Land für Automatisierung und Prozessoptimierung ist. Die Welt findet sich hier zusammen. Deutschland könnte durch die Digitalisierung wieder eine Vorreiterrolle einnehmen.

**CIOBRIEFING: Sie haben den Datenschutz schon angesprochen. Jetzt kommt die Europäische Datenschutzgrundverordnung (DSGVO). Ist mit der Nivellierung ein Vorteil für die Datenhaltung in Deutschland dahin?**

**Dirk Stenkamp:** Eine Vereinheitlichung ist wichtig, was Europa angeht. Ich bin Verfechter eines starken Europa, um im Wettstreit mit Amerika und Asien bestehen zu können. Ich wünsche mir sogar mehr Initiative, um Mega-Unternehmen zuzulassen,

um Google, Alibaba Co. etwas entgegen setzen zu können. Wir brauchen größere Konglomerate und eine EU-Politik, die zulässt, dass sich Schwergewichte formieren können.

### **CIOBRIEFING: Was bedeutet das für die Ansiedlung von Rechenzentren?**

**Dirk Stenkamp:** Für mich ist ein wettbewerbsfähiger Rechenzentrumsbetrieb eine Frage innovativer Technik, der Incentivierung und der Politik.

■ *Ulrike Ostler*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



# Maverick Research: The Edge Will Eat the Cloud

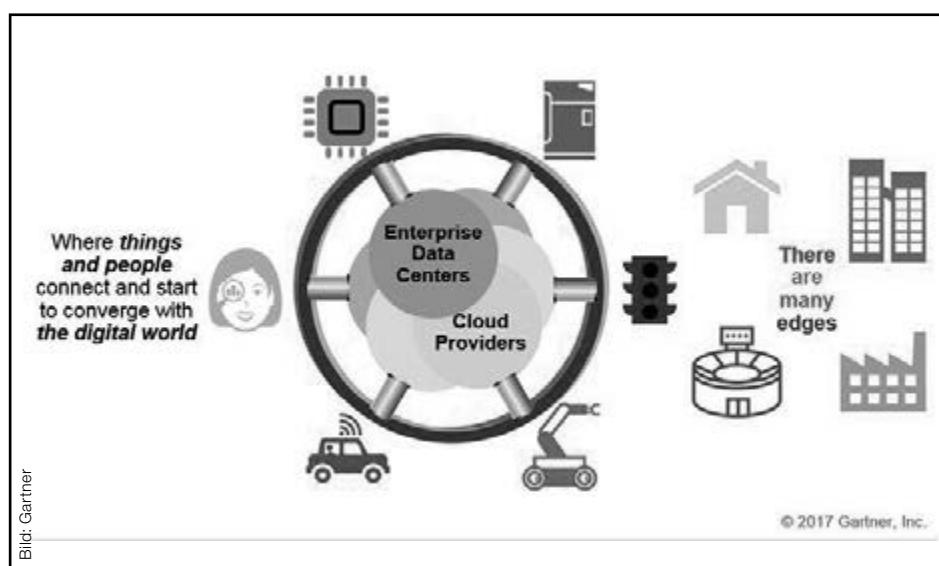
## Appetit auf Cloud?

### Der Fressfeind Edge-Computing

**Schon der Report-Titel von Gartner-Maverick-Analyst Thomas J. Bittman rüttelt auf: „The Edge Will Eat the Cloud“. Edge-Computing wird zu einer großen Herausforderung für Unternehmen sowie für Anbieter von Cloud- und Rechenzentrumsdiensten. Datacenter-Betreiber Equinix hat nun den Report einer breiteren Öffentlichkeit zugänglich gemacht. Was also steht drin?**

Maverick Research ist bekannt für unkonventionelles Denken und ungewöhnliche Ratschläge. Im vergangenen September veröffentlichte Analyst Bittman seine Thesen zum Verhältnis von Edge-Computing und Cloud. Auslöser ist das IoT-Wachstum (siehe: Kasten) und der Trend zu immersiveren, intuitiveren und interaktiven Benutzeroberflächen; denn diese Faktoren werden in den kommenden vier, fünf Jahren den Schwerpunkt der Datenproduktion und -verarbeitung weg von zentralen Rechenzentren hin an den „Rand“ verlagern.

Vor allem hierzulande sind viele IT-Berater noch damit beschäftigt, die Unternehmens-IT in die Cloud (nicht die in eigenen Rechenzentren) zu jagen. Somit widerspricht Maverick Research den vorherrschenden Ansichten über die Zukunft des Cloud Computing, und stellt die Topologie von Computing-Architekturen und mehr noch die Art und Weise in Frage wie Anwendungen auf den Weg zum digitalen Geschäft angepasst werden sollten.



Edge ist da, wo physische auf die digitale Welt trifft.

Somit lautet die Hauptthese: Das Wachstum von Mega-Rechenzentren wird gebremst, Computer und Speicher verlagern sich wegen des Internet der Dinge und neuer Benutzer-/Maschinen-schnittstellen an den (Netzwerk-) Rand. Dazu gesellen sich folgende Hauptpunkte:

- Die schiere Datenmenge, die durch „Dinge“, Maschinen und Services, generiert wird, die Notwendigkeit, lokal zu interagieren und von Echtzeitanalysen werden die Datenverarbeitung näher zu den Dingen bringen und damit an den Rand drängen.
- Immersive Technologien und Echtzeit-Schnittstellen erfordern eine geringe Latenzzeit und werden das Computing-Paradigma umkehren, Computer und Daten rücken insofern näher an die Benutzer.
- Die Rolle der zentralisierten Cloud wird sich in Richtung Koordination, Aggregation, Archivierung, Back-End-Maschinenlernen und traditionelle Back-Office-Verarbeitung verlagern.
- Dabei entsteht eine Zangenbewegung: Cloud-Provider dehnen ihre Techniken und Strategien in Richtung Edge-Computing aus. Zugleich wird die Entwicklung am Edge stärker von den Technologien und Techniken getrieben, die vor Ort bereits physisch etabliert sind.

### **Die hauptsächlichen Empfehlungen an alle Unternehmen:**

- Entwickeln Sie bis 2020 eine Edge-Computing-Strategie und planen Sie bis 2022 den Einsatz von Edge-Computing-Lösungen.
- Identifizieren Sie zunächst die Technologien, Anwendungsarchitekturen und -topologien (einschließlich des Randes), die auf kleinere aber stark im Fluss befindliche Geschäftsprozesse gemünzt sind.
- Planen Sie für die Zukunft massive Investitionen in Benutzeroberflächen, um die Interaktion zwischen Unternehmen, Kunden und Mitarbeitern dynamischer und intuitiver zu gestalten.
- Entwickeln Sie Strategien für zukünftige Anwendungen, die Latenz, Standort, Verteilbarkeit und Autonomie berücksichtigen, und bestimmen Sie die Platzierung am Rand, in der Wolke und dazwischen.

### **Was passiert?**

Hier listet der Report die grundlegenden Annahmen auf, die den Überlegungen zugrunde liegen.

- Vierzig Prozent der Großunternehmen werden Prinzipien des Edge-Computing in ihre 2021er Projekte integrieren; im vergangenen Jahr waren es noch weniger als 1 Prozent.
- IoT und intuitive, interaktive Benutzeroberflächen werden ein Drittel der großen Unternehmen dazu bringen, bis 2021 Edge-Standorte zu bauen oder etwa im Co-Location zu nutzen.
- Bis 2021 werden Latenz- und/oder Bandbreitenanforderungen für 30 Prozent der Workload-Bereitstellungen zu kritischen Faktoren.
- „Die IT-Architektur wird auf den Kopf gestellt“, heißt es im Report und weiter: „Die Digitalisierung Business verwischt die Grenze zwischen der digitalen und der physischen Welt immer mehr - aber dieses 'Verwischen' findet zumeist im Edge statt.

Denn Cloud Computing allein kann den Anforderungen nicht gerecht werden: Selbst im Hyperscale-Umfeld ist ein zentrales Wachstum, das den Bedarf von Milliarden „Dingen“ decken könnte, unmöglich. „Cloud Computing hat Zugang zu massiver zentralisierter Rechen- und Speicherleistung verschafft. Aber es bietet nicht die Agilität und physische und digitale Konvergenz, die man am Rande des Geschehens benötigt. Die 'Wolke' ist zu weit entfernt. [...] Und Compute geht immer zu den Daten.“

Das bedeutet aber auch, dass das „zentralisierte Cloud-Computing“ sich ändern wird. Die künftige Rolle spielt im Backend, als ein wichtiger Bestandteil des digitalen Geschäfts. Insbesondere in der Koordination, Aggregation, Archivierung und dem maschinellen Lernen aber wird die Cloud in den Hintergrund gedrängt. Wachstum und Chancen durch den Einsatz von IT findet im Edge, vor Ort statt.

Denn: Standort und Datenverteilung sind für ein gutes Management entscheidend. Umgekehrt werden Unternehmen und Anbieter, die sich nicht auf neue Anforderungen konzentrieren, nicht mehr wettbewerbsfähig sein. Edge frisst Cloud.

### Was ist „Edge“?

Edge ist dort, wo Dinge und Menschen verbunden werden und mithilfe eines Kommunikationsnetzes mit der digitalen Welt verschmelzen. (siehe: Abbildung 1). Heute ist das ein Smartphone, ein Router, eine Spielkonsole, eine Set-Top-Box, ein Mobilfunkturm, ein „Amazon Echo“, ein Laptop, ein Mikro-Rechenzentrum oder die Intelligenz, die in ein Auto eingebettet ist.

Edge Computing sind Verarbeitung, Speicherung sowie Dienste für Dinge und Menschen, die physisch nah sind. Edge-Services

werden durch Software bereitgestellt, die auf einer Vielzahl von Hardware laufen kann und für viele Anwendungsfälle taugt. Möglich ist eine Hierarchie von Rechendiensten, vom Rand bis zu den zentralen Datacenter geliefert.

## Die Entwicklung

- Anfang der 1980er Jahre: PCs (mit Servern) begannen Mainframes und Terminals zu verdrängen und zogen Computing an den Rand.
- Mitte der 2000er Jahre: Intelligente mobile Geräte (und Anwendungen) begannen, PCs zu verdrängen Apps entstanden und ermöglichen noch mehr Edge-Computing.
- Ende der 2000er Jahre: Cloud begann mit dem Verschlucken von Unternehmensrechenzentren, Es begann zugleich eine Zentralisierung der Daten und Compute -Power in Mega-Rechenzentren.
- Anfang der 2010er Jahre: Es entstand eine völlig neue Klasse von „Nutzern“. Diese „Dinge“ sind Nettodatenproduzenten, nicht nur Konsumenten.
- Anfang der 2020er Jahre: Die Mensch-Maschine-Schnittstelle wird noch multimodaler, interaktiver, umgebungssensitiver, kontextgebundener und intuitiver sein.

## Ja, aber:

Mainframes verschwanden nicht gänzlich, als PCs und Server auftauchten – Doch sie wurden weniger dominant. PCs gingen nicht weg, als mobile Geräte auftauchten, aber die PC-Verkäufe erreichten ihren Höhepunkt etwa im Jahr 2011 und sind seitdem im Großen und Ganzen rückläufig. Die Cloud wird nicht verschwinden, wenn vernetzte Dinge zur Norm werden. Doch es ereilt sie dasselbe Schicksal: Die Cloud wird einfach weniger dominant und nur ein Teil einer breiteren IT-Architektur.

## Und was läuft nun falsch für die Cloud?

Cloud ist ein Computing-Stil - Selbstbedienung, automatisiert. Allerdings geht es beim Cloud Computing und der Storage-Zentralisierung in Mega-Rechenzentren heute vor allem um Größenvorteile, massive Skalierung und Back-End-Verarbeitung. Dies ist ideal, wenn Daten - und deren Berechnung - zentralisiert werden können.

Die heutige Cloud löste ein großes Problem - die Erhöhung der flexiblen Zugänglichkeit von Computing und Storage für Unternehmen, insbesondere für das Backoffice. Aber das Front-Office blieb ein Browser oder eine App, die mit weit entfernten Rechenzentren mit diskreten Anfragen und Antworten sprach.

Content-Delivery-Netzwerke (CDNs) sind entstanden, um den Daten-Chache näher an den Anwender zu bringen - aber das Paradigma besteht nach wie vor darin, Daten in eine Richtung zu streamen, von der Mitte bis zum Rand. Die Explosion IoT wird den Wert noch mehr in Richtung Agilität verschieben und dorthin, wo Anwendungen viel unmittelbarer und interaktiver sein müssen, wo Verzögerungen inakzeptabel sind und wo der Großteil der Daten produziert wird und wo letztlich analysiert und konsumiert werden muss.

Somit gibt es fünf Hauptprobleme, die es zu überwinden gilt: Latenz, Bandbreite, Autonomie, Privatsphäre und Sicherheit.

## Das Latenzproblem

Die Rechenleistung hat sich in den vergangenen Jahrzehnten um das Tausendfache verbessert. Auch die Netzwerkbandbreite hat seit 56 Kbps Einwahlleitungen einen weiten Weg zurückgelegt. Das Problem ist die Netzwerklatenz.

Darunter ist ein begrenzender Faktor zu verstehen, der sich aus Entfernung, Übertragungsmedium, Anzahl der Netzwerk-Hops und -Router, des Protokolls und vor allem der Lichtgeschwindigkeit zu verstehen. Wenn es möglich wäre San Francisco nach New York City und zurück auf einem Glasfaserkabel zu senden, ohne Router, Repeater oder verarbeitende Komponenten, würde es dennoch 42 Millisekunden (ms) dauern.

Und Millisekunden zählen:

- Über den Online-Handel sagte der Amazon-Pionier Greg Linden 2006: „Jede 100 ms Verzögerung kostet 1 Prozent des Umsatzes“, zitiert Analyst Bittman.
- Ein Experiment von Google-VP Marissa Mayer im Jahr 2006 zur Online-Suche ergab, dass eine zusätzliche Verzögerung von 500 ms bei den Suchergebnissen einen Rückgang von 20 Prozent bei Traffic und Einnahmen verursachte.
- Bittman verweist auf die Tabb Group: „... der Wert der Zeit für ein Trading Desk ist entschieden nichtlinear... wenn die

elektronische Handelsplattform eines Brokers 5 ms hinter der Konkurrenz liegt, könnte sie mindestens 1 Prozent ihres Flow verlieren; das sind 4 Millionen Dollar Umsatz pro Millisekunde. Bis zu 10 ms Latenzzeit können zu einem Umsatzrückgang von 10 Prozent führen. Von da an wird es noch schlimmer.“

- Für ein selbstfahrendes Auto, das 120 Kilometer pro Stunde fährt, entsprechen 100 ms etwa 3,3 Meter. Aber wenn es zwei selbstfahrende Autos sind oder zwei Dutzend, die alle zum selben Ort fahren, dann sind 100 ms eine Ewigkeit.

Doch selbst für Menschen sind Millisekunden entscheidend. Bei Verwendung von Virtual-Reality-Brillen ist eine Verzögerung von mehr als fünf bis zehn Millisekunden von der Kopfbewegung bis zur Blickbewegung spürbar und kann zu Reisekrankheit führen.

### Was passiert mit „der Cloud“?

Die Ausführungen zur Bandbreite und Autonomie von Anwendungen, Services und „Dingen“ sowie zu Security und Datenschutz entsprechen dem, was häufig auf Datacenter-Insider Erwähnung findet. Es bleibt die Frage, was mit zentralen Cloud-Angeboten wird, wenn sich das Edge-Computing auswächst.

- Koordination: Cloud Computing verfügt über die derzeit fortschrittlichsten Techniken für automatisiertes Provisioning und Remote Management. Diese Technologien werden in Zukunft auch für das Edge-Computing genutzt beziehungsweise darauf ausgedehnt.
- Aggregation: Zusammenfassen mehrerer Edges, zum Beispiel Daten und Ergebnisse aus allen Außenstellen eines Unternehmens.
- Archivierung: Aufzeichnung von Daten und Protokollen aus den Edge-Anwendungen zur Archivierung oder späteren Verarbeitung.
- Maschinelles Lernen: Während maschinelles Lernen und Trainieren vor Ort stattfindet, könnten längerfristige Lern-, Analyse- und Trainingsmodelle in einer zentralen Cloud ausgearbeitet und ausgeübt werden.
- Traditionelle Back-Office-Verarbeitung: Wo Echtzeitverarbeitung unkritisch ist und Skalierbarkeit mehr zählt als Agilität, übernimmt die zentrale Cloud diese Aufgabe.
- Fallback Edge: Wo es kein intelligenteres, autonomes Edge-Computing gibt, wandern diese Aufgaben in die Cloud. An-

wender nehmen in Kauf, dass die Verarbeitung hier weniger effizient, weniger agil und weniger interaktiv ist.

Allerdings gilt: Wenn die Bedeutung des Edge-Computing wächst, wird es einen deutlich spürbaren Unterschied zwischen den Unternehmen geben, die Edge-Computing ermöglichen können und solchen, die auf remote Verbindungen zu einem zentralen Rechenzentrum angewiesen sind.

### **Wen betrifft das Edge-Computing?**

Welche Unternehmen, welche Branchen bauen wie Edge-Computing auf? Der Maverick-Report listet Folgendes auf.

- Im Heim: Spielkonsolen, Kabel-Set-Top-Boxen, Großgeräte, Wi-Fi-Router, Heim-PCs, intelligente Assistenten
- Bei mobilen Endgeräten: Smartphones, 5G-Mini-Türme
- Bei Unterhaltungsstätten wie Theater und Vergnügungsparks: Mikrorechenzentren, Edge-Server, Campus Wi-Fi
- In der Transportbranche: Hier geht es etwa um Prozessoren in Autos und Flugzeugen
- In den Städten: Hier werden 5G-Minitürme zu sehen sein und intelligente Ampeln
- An den Arbeitsplätzen: Campus Wi-Fi Router, Enterprise-Rechenzentren
- In der Produktion: Mikro-Rechenzentren, Edge-Server
- Beim Militär: Mobile Micro-Datacenter (Cloudlets)

Mit anderen Worten: Es gibt fast keine Branche, die nicht betroffen ist. So gehören zu den verschiedenen Akteuren in erster Linie die Unternehmens-IT, Spiele-Anbieter, Telcos- und Kabelbetreiber, Smartphone-Hersteller und Anbieter von Consumer-Elektronik.

■ *Ulrike Ostler*

## Die Parameter des boomenden Datacenter-Schulungsmarkts

# Jobkiller und Fachkräftemangel bestimmen die Suche nach Datacenter-Mitarbeitern

**Allmählich dämmert es: In den kommenden Jahren wird fast die Hälfte der jetzigen Jobs verschwinden und es wird auch die so genannten Wissensarbeiter treffen. Zugleich herrscht ein Facharbeitermangel, auch in den Rechenzentren, was die Forderung nach Flexibilität etwa der IT-Kräfte verstärkt, die mehr und mehr zu „digitalen Tagelöhnnern“ werden.**

Im vergangenen Jahr sorgte die Studie der University of Oxford für Furore, die zu dem Schluss kam: In den kommenden 25 Jahren werden 47 Prozent der Jobs verschwinden — zumindest in den weit entwickelten Ländern dieser Erde. Die Forscher nahmen 702 Berufsfelder (in den USA) detailliert unter die Lupe.

In einem Forschungsbericht des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit aus dem Jahr 2015 gingen die Wissenschaftler davon aus, dass „bei 13,2 Millionen Beschäftigten Computer 30 bis 70 Prozent ihrer Tätigkeiten erledigen können. Und bei 4,4 Millionen Deutschen beträgt das Substituierbarkeitspotenzial mehr als 70 Prozent. Das heißt: 4,4 Millionen Menschen könnten theoretisch schon jetzt von Computern ersetzt werden“, schrieb die „Wirtschaftswoche“ damals.

Bild: gemeinfrei - MIH83/Pixabay / CC0



**Workshop statt Bildung?  
Training statt Lehrsaal?**

Also: Gibt es nun einen Fachkräftemangel, der sich verschärft, oder eine beginnende Entlassungswelle? Treiben die Forderung nach Agilität und Flexibilität die Unternehmen ins Outsourcing und Bodyshopping, die Arbeitskräfte ins tage- und stundenweise Auftraggeber-Hopping und in die Selbstausbeutung?

### **Mehr Service anstelle von festen Gefügen**

Nach Angaben des Beratungs- und Marktforschungsunternehmens Markets and Markets wächst der Service-Markt für Rechenzentren rasant: Der Grund: Immer mehr Anwender entschieden sich für Services, die die Verfügbarkeit des Rechenzentrums erhöhten und Ausfallzeiten reduzierten.

Nach Evaluierungen des Unternehmens belief sich der Umsatz im weltweiten Servicemarkt für Rechenzentren im Jahr 2017 auf rund 39 Milliarden Dollar; im Jahr 2022 sollen 77,51 Milliarden Dollar verdient werden. Das heißt: Das Wachstum in den kommenden Jahren wird durchschnittlich etwa 14,33 Prozent jährlich betragen und insgesamt um 68 Milliarden Dollar anschwellen.

Die Zahlen entstammen dem Bericht „Service Market for Data Center by Service Type - Global Forecast to 2022“, der auf 123 Seiten Design & Consulting, Installation & Deployment, Training & Development, Maintenance & Support umfasst und nach Tier-Klasse, Unternehmen, Datacenter-Typ, Industrie und Region aufgeschlüsselt ist.

### **Aus- und Weiterbildungsleistungen**

Die wichtigsten Faktoren, die den Servicemarkt für Rechenzentren antreiben, sind, so die Analysten, der wachsende Datenverkehr, die zunehmende Komplexität innerhalb von Rechenzentren, die höhere Nachfrage nach Kosteneffizienz und der Anstieg der Ausgaben für Rechenzentrumstechnologie in Unternehmen. Am meisten zulegen aber sollen die Aus- und Weiterbildungsleistungen. Schulungs- und Entwicklungsdienstleistungen unterstützen Unternehmen bei der Verbesserung der Leistung von Rechenzentren durch Schulungsprogramme. Darüber hinaus helfen diese Services bei der Transformation der Unternehmens-IT bei gleichzeitiger Kostensenkung.

Nachdem der Datacenter-Markt in den USA am stärksten wächst, dürfte es nicht verwundern, dass Nordamerika im Prognosezeitraum voraussichtlich auch den Servicemarkt für Rechenzentren dominieren wird. Allerdings prognostizieren die Analysten, dass

der asiatisch-pazifische Markt voraussichtlich zwischen 2017 und 2022 mit dem höchsten Durchschnittswachstum rechnen kann. Die Haupttriebkräfte für dieses Wachstum sind der zunehmende Datenverkehr in den Rechenzentren aufgrund des Anstiegs der mobilen Geräte in dieser Region.

### **Lehre und Studium: Viel zu lahm**

Unter Umständen sind die außerschulischen Fortbildungen, die Trainings jenseits der Hochschulbildung und Qualifizierungen abseits von Lehre und sonstigen staatlich regulierten Ausbildungsgängen ohnehin die einzige Möglichkeit, sich die notwendigen Fähigkeiten anzueignen. Denn die offiziellen Organe reagieren viel zu behäbig.

So ist davon auszugehen, dass es, um einen neuen Studiengang einzurichten, sechs Jahre braucht. „Mit dieser Einschätzung sind Sie noch gut dabei“, sagte neulich Franz Josef Pschierer, der noch frisch gebackene Bayerische Staatsminister für Wirtschaft, Energie und Technologie, bei der Vorstellung der Initiative „applied AI“ des UnternehmerTUM Projekts (TUM = Technische Universität München). Hier haben sich 26 Technologie- und Industriekonzerne wie Google, Nvidia, Allianz, Linde, Porsche Consulting, SAP und Siemens sowie Partnern aus Politik und Wissenschaft zum Ziel gesetzt, die Anwendung von künstlicher Intelligenz (KI) in Deutschland voranzutreiben. Mit von der Partie: rund 100 Studenten. Allerdings will man Online-Formate entwickeln, mit denen in den kommenden Jahren Einhundertausende Lernwillige erreicht werden sollen.

### **Bildung und Ausbildung bei Datacenter-Jobs**

Die Arbeitsvermittler von Teksystems untersuchen jährlich, wie sich Rekrutierung und die Bindung von IT-Mitarbeitern an die Unternehmen entwickelt. Demnach sagen aktuell rund 81 Prozent der IT-Leiter, dass es schwierig sei, qualitativ hochwertige Kandidaten zu finden, und fast die Hälfte erwartet nicht, eine IT-Stelle innerhalb des erwarteten Zeitrahmens zu besetzen.

Das hat Folgen: Nur rund ein Drittel der Rechenzentrumsmanager und CIOs geben an, dass sie glauben, ihr Unternehmen verfüge über die Fähigkeiten, die Anforderungen eines Rechenzentrums zu erfüllen. Außerdem scheint in manchen Regionen der Mitarbeitermarkt geradezu leergegegt: Teksystems beobachtet, dass der Fachkräftemangel durch Rekrutierung von Militärve-

teranen, an lokalen Universitäten und Qualifizierungsprogrammen immer beliebter werden, um die Talentpipeline im Rechenzentrum zu stärken.

Sie gehen zudem davon aus, dass sich die Situation durch die Veränderungen in der IT noch verschärfen wird. Genannt werden als Faktoren Cloud, Konvergenz, Internet der Dinge (IoT), Virtualisierung und Mobilität. Technologiekonfigurationen und die Abhängigkeit von externen Dienstleistungen änderten die Personal- und Ausbildungsprioritäten. Karsten Scherer, Leiter Global Analyst Relations bei Teksystems, bemerkt: „Die sich entwickelnde digitale Welt und die Cloud erfordern eine Änderung der Rechenzentrumsstrategie mit unterschiedlichen Kompetenzen.“

## Der Wettbewerb

Er sieht zudem eine wachsende Spannung zwischen traditionellen Rechenzentren und den Mega-Rechenzentren von Dienstleistern wie Google, Amazon Web Services und Microsoft Azure. „Diese bauen ihre Rechenzentrumspräsenz bedarfsgerecht aus und konkurrieren mit traditionellen Rechenzentren um den gleichen Pool an Talenten“, so Scherer. „Gepaart mit Fluktuation, einer alternden Belegschaft und jüngeren Menschen, die anderswo arbeiten wollen, haben wir einen zunehmenden Wettbewerb um einen schwindenden Pool an Arbeitskräften.“

Dazu einige Zahlen: Die Gesamtzahl der Unternehmen Rechenzentren weltweit erreichte zuletzt einen Spitzenwert von rund 8,6 Millionen. Doch die absolute Anzahl der Unternehmens-Datacenter schrumpft, stellt der Report „State of the Data Center Industry, An Analysis of Washington's Competitiveness In This Fast-Growing High-Tech Field“ des Washington State Department of Commerce fest. Sie werden konsolidiert, Cloud-Computing nimmt zu genauso wie Co-Location.

Doch in der Gesamtheit betrachtet ist das Wachstum des globalen Rechenzentrumsmarkts mehr als gesund: Rund 33.445.094 Quadratmeter wurden in den vergangenen fünf Jahren neu gebaut.

## Die neuen Anforderungen

So kommt auch die US-Behörde zum Schluss: Das explosionsartige Wachstum von Cloud Computing hat das Design von Rechenzentren dramatisch beeinflusst. Noch vor fünf Jahren wurde in den Staaten ein 100.000 Quadratfuß-5 Megawatt-Projekt

als groß eingestuft. Heute sei es nicht ungewöhnlich, dass eine Ankündigung über 1 Million Quadratmeter und 100 Megawatt auftaucht.

China ist übrigens der Spitzenreiter im Rechenzentrumsmarkt mit über 3,5 Millionen Quadratmetern in sechs bereits existenten Datacenter, ein weiteres Hyperscale-Rechenzentrum befindet sich derzeit im Bau. Nur ein Projekt in den USA, „Super Nap“ in Las Vegas, weist ebenfalls eine solche Größe aus. Allerdings wurden ein halbes Dutzend US-Bauten von rund 92.903 Quadratmetern und mehr in diesem Jahr angekündigt.

Das aber bedeutet wiederum, dass sich während der Pool an qualifizierten Fachkräften für das Rechenzentrum dünn besetzt ist, die Rollen jedoch weiterentwickeln. Die zunehmende Verbreitung von hybriden Umgebungen verlange, so die Studie, dass Rechenzentrumsmanager lernen müssten, mit einer Mischung aus lokalen Ressourcen und externen Diensten umzugehen, die sie von Anbietern über die Cloud beziehen.

## Und nun?

Die Bausteine der Stellenbeschreibungen spiegelten dies wider: Tecsystems-Analyst Scherer stellt eine stärkere Betonung auf Formulierungen wie emotionale Intelligenz, Brückenbau und Silobruch sowie Begriffe wie kreativ, anpassungsfähig, strategisch, innovativ, Allianzbildung und Verhandlungsführung fest, die immer häufiger in den Stellenanforderungen auftauchten.

Er sagt: „Architekturorientierte Rollen und Business Analysten werden als kritisch angesehen, um all diese Teile des Ökosystems zusammenzuführen und ihnen zu helfen, miteinander zu sprechen.“ Auch das Datacenter Infrastructure Management (DCIM) könne mit der Zeit an Bedeutung gewinnen, also mehr Automatismen

Scherer fordert, dass angesichts der zunehmenden Konzentration auf Brokerage-Aufgaben zwischen der Cloud und dem internen Rechenzentrum mehrere Mitarbeiter geschult werden sollten, um als Schnittstelle zwischen Anbietern, IT und Geschäftsführern zu fungieren. Diese Personen müssten gut in der Verwaltung von Verträgen und Service Level Agreements (SLAs), Beschaffung und mehr sein. Dies sind allerdings nicht immer Talente, die für technische Spezialisten selbstverständlich sind, so dass eine gewisse Ausbildung wahrscheinlich notwendig sein wird.

■ *Ulrike Ostler*

# Es muss nicht alles und immer neu sein

## Handeln wie Google - Refurbishment im Datacenter

**Führende auf dem Gebiet der Forschung zum Thema Kreislaufwirtschaft, Kosten- und Energie-Einsparungs-Innovationen, scheint momentan Google zu sein, insbesondere wenn es um Nachhaltigkeit im Tech-Bereich geht. Glauben Sie Ihr Unternehmen ist zu klein in punkto Personal und Umsatz, um diese Ideale zu erreichen? Umdenken! Es gibt verschiedene Möglichkeiten...Und so können Sie diese finden.**

### Was bedeutet Kreislaufwirtschaft überhaupt?

Eine Kreislaufwirtschaft ist die Bewegung zu einem System, das sich hinsichtlich Energie- und Materialverbrauch selber tragen kann. Der Fokus beim Betrieb von Rechenzentren lag hier bisher beim Engergieverbrauch, Ob Senkung des PUE-Werts (Power Use Effectiveness) oder die Verwendung von Solar, Wind- und Bio-Kraftstoff - die meisten Unternehmen konzentrierten sich überwiegend auf den Kohlenstoffausstoß der täglichen Tätigkeit Ihrer Anlagen. Jedoch sind effiziente Kühlung und Gebäudeplanung nicht alles, wenn es um Kohlenstoffausstoß – oder Kosten – in einem Rechenzentrum geht.

Die unwirtschaftliche Nutzung von Computer-Ressourcen hat einen größeren Effekt auf die Effizienz als PUE. Server mit Software-Defined-Storage (SDS) sind daher eine ausgezeichnete Möglichkeit den Kohlenstoffausstoß zu vermindern und Geld zu sparen.

Unternehmen wie Google haben ihre eigene Netzwerkarchitektur entwickelt, um Applikationen zu planen und verwalten. Doch kann jeder, der will, eine große Auswahl an Software auf dem Markt finden, die Ihnen ähnliche Ergebnisse bringen kann.



**Hätten Sie gedacht, dass Google sein Equipment nicht einfach entsorgt, sondern zu einem Großteil aufbereitet und dann erneut verwendet?**

## Wie macht Google das?

Wenn es darum geht, Systeme zu aktualisieren, bieten ältere Server und Komponente eine kostengünstige Lösung. Google lässt seine Upgrades nicht von Berechnungen der Lebenszyklen diktionieren, sondern verwendet so genannte „refurbished“ Komponenten für Reparaturen und Upgrades in seinen Datacenter.

Im Jahr 2015 stammten 75 Prozent aller verwendeten Ersatzteilkomponenten aus einem wiederaufbereiteten Bestand. Im gleichen Jahr kamen die Komponenten für Upgrades zu 52 Prozent aus wiederaufbereitetem Material und 19 Prozent aller Server waren refurbished.

Auf Grund der Unternehmensgröße Größe hat Google die Kapazität für die Lagerung von Ersatzteilen. Kleinere Unternehmen mit weniger Lagerkapazität können jedoch gleichwertige Komponente kaufen. Es gibt eine ganze Reihe von Unternehmen, die sich auf die Wiederaufbereitung von IT Hardware spezialisieren haben. Sie verfügen häufig zudem über ein großes Inventar und bieten oft die Lieferung am Werktag der Bestellung. Global agierende Unternehmen bieten die weltweite Lieferung zumeist am darauffolgenden Werktag, da sie über viele Standorte verfügen. Und: Häufig bieten sie Garantien wie der Hersteller selbst.

## Was bedeutet „refurbished“?

Refurbishment oder Aufbereitung bedeutet Geräte wieder auf Werkseinstellungen zurückzubringen, so dass diese anderweitig wiedereingesetzt werden können. Wenn man in Betracht zieht, wie schnell sich die IT Industrie verändert, sind Ausrüstungs- Upgrades oder -Änderungen meist öfter nötig und noch bevor die Produkt-Lebensdauer abgelaufen ist. Aufbereitete Produkte werden oft im Edge Computing, in der Ausfallsicherung und als Backup eingesetzt.

## Wie funktioniert der Refurbishing-Prozess?

Dies hängt vom Produkt oder der Komponente ab. Festplatten zum Beispiel werden einer kompletten Datenlöschung mit spezieller Software unterzogen. Alle Geräte werden optisch überprüft, gründlich gereinigt und auf Funktionsfähigkeit getestet. Server werden in die einzelnen Komponenten zerlegt und diese dann nach Artikelnummer und Spezifikation protokolliert sortiert. Die Einzelteile können dann entweder einzeln verkauft werden oder

nach Wünschen des Käufers neu zusammengestellt oder als bestehender Rechner neu konfiguriert werden.

### Auf was sollte man achten?

Die Antwort ist einfach: Qualität, Verfügbarkeit und Garantie. Wiederaufbereitete Hardware zu kaufen sollte wirtschaftlich und zugleich zuverlässig sein.

- Suchen Sie sich deshalb ein Unternehmen, welches eine Dreijahres-Garantie bietet.
- Erfragen Sie, wie hoch der Vorrat das Unternehmen ist, damit Sie abschätzen können, wie schnell Sie Ihre Bestellung erhalten können.
- Achten Sie auf Kompetenz - eine ISO Zertifizierung ist ein guter Hinweis für Qualitätskontrolle, aber Extras, zum Beispiel ein Service, der Ihnen individuelle Konfigurationen bietet (CTO), sind weitere Anzeichen für fachliche Kompetenz.

■ *Chris Pooley*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



## Bericht vom Summit 2018 in Vancouver, Teil 1

# OpenStack Summit 2018, Teil1: Kritik? War da was?

**Wer auf deutliche Verbesserungen an kritisierten Punkten gesetzt hatte, könnte enttäuscht sein. Das Open-Source-Projekt OpenStack hat anders reagiert und macht davon wenig Aufheben. Es bleiben allerdings offene Fragen, auf die die Organisation noch keine Antwort gefunden hat.**

Rundum positiv ist der Eindruck nicht, den das OpenStack-Projekt auf seinem Summit 2018 in Vancouver zu vermitteln vermochte. Die Teilnehmerzahlen an den Kongressen sinken. Kein Wort mehr von Zusagen der letzten zwölf Monate oder groß herausgestrichenen Neuerungen. Schwere organisatorische Aufgaben bleiben dem Projekt erhalten.

### Teilnehmerzahlen zeigen Verlagerung der Anwenderinteressen an

Die Frühjahrs-Kongresse galten einmal als wichtiger und waren besser besucht als die Herbst-Events in Europa oder Fernost/Australien. Dem ist nicht mehr. 5600 Besucher kamen zum Herbst-Summit 2016 in Barcelona, 5000 ein halbes Jahr später in Boston. Nach Sydney reisten im Herbst 2017 fast so viele wie jetzt zum Frühjahrs-Event im kanadischen Vancouver kamen: 2600.

Über diese Zahlen zeigte sich SUSE-Manager Alan Clark, Aufsichtsratsvorsitzender der OpenStack Foundation, „nicht enttäuscht“. Er verwies darauf, dass die Organisation das Entwicklertreffen vom Summit abgetrennt habe. Es fand zwei Wochen vor dem Vancouver-Event in Dublin statt. Würde man aber dessen 800 Teilnehmer zu den Besucherzahlen von Vancouver addieren, ergäbe das immer noch keine Erfolgsmeldung. Anscheinend ist OpenStack in Nordamerika nicht mehr so angesagt



Bild: © Jan Engel - stock.adobe.com

**Der Frühjahrs-Summit 2018 fand nicht wie üblich auf dem bisher größten OpenStack-Markt in den USA statt, sondern aus Protest gegen die US-Einreisebeschränkungen im kanadischen Vancouver.**

wie in Europa und Fernost, wo es als die Alternative zu US-amerikanischen Public Clouds gilt.

### **Kritik findet kaum Widerhall**

Auf dem Summit in Sydney hatte es vor einem halben Jahr auf der Bühne Kritik und Forderungen gegeben. Insbesondere Sorabi Saxena, President Business Operations bei AT&T Business Solutions, hatte in seiner Keynote vier Verbesserungen verlangt: Security by Design, Vereinfachung des Betriebs, störungsfreie Upgrades sowie Best Practices und Metriken für den Reifegrad von Projekten. Jonathan Bryce, President der OpenStack Foundation, und Mark Collier, COO der Organisation, hatten damals sofort reagiert und Besserung versprochen.

In ihren Reden zur Eröffnung des Summits in Vancouver war davon nun überhaupt nicht die Rede. Pressevertreter ziehen daraus üblicherweise den Gegenschluss, dass sich nichts getan hat. Prompt war es anderen OpenStack-Vertretern wichtig darauf hinzuweisen, dass man in den Teilprojekten doch auf diese Kritik eingegangen war und Verbesserungen vorgenommen hat.

Es wirft allerdings kein gutes Licht auf eine Open-Source-Organisation, zentrale „Pain Points“ der Anwender so nebenbei zu behandeln. Das macht das Marketing der OpenStack-Distributionen besser (siehe Bilderstrecke), aber damit tun sich Open-Source-Projekte grundsätzlich schwer.

### **Verbesserungen in den Details**

So musste Aufsichtsrat Alan Clark von SUSE im Interview mit DataCenter-Insider (demnächst ausführlich hier) dem Vorstand nacharbeiten. Er verwies darauf, dass bereits bei der aktuellen OpenStack-Version „Queens“ Wert auf bessere Manageability und unproblematische Updates gelegt worden sei. Letzteres auch eine Folge davon, dass das OpenLab, so Teammitglied Melvin Hillsman von Huawei, ein halbes Jahr nach seiner Ankündigung in Sydney mit seinen Testprozeduren arbeitsfähig ist.

In Sydney war es angekündigt, Anfang Februar hatte es Jonathan Bryce in einem Gespräch mit DataCenter-Insider bestätigt: Interessierte Anwender sollen den Reifegrad von OpenStack-Projekten und die für sie besten Teillösungen besser erkennen können. In Vancouver war die Lösung kaum mehr als einen Nebensatz wert. Sie besteht darin, dass die OpenStack-Website komplett überarbeitet wurde. Hier finden sich unter anderem Bei-

spielkonfigurationen für bestimmte Anwendungszwecke samt einer Beurteilung des Reifegrads der empfohlenen Teilprodukte, ein Projekt-Navigator oder beispielsweise die Roadmap.

## Orientierung im OpenStack-Dschungel

Das könnte die Orientierung durchaus erleichtern. Denn OpenStack verfolgt erklärtermaßen die Politik, möglichst viel Auswahl anzubieten. „Das ist der beste Weg, Probleme zu lösen“, erklärt CIO Collier, „denn eine Lösung ist nicht genug“. Eine – inzwischen schon wieder überholte – Grafik weist allein für den OpenStack-Kern 31 Teilprodukte aus. Hinzu kommen 29 Anwendungen für Clients/Programmierung, zugehörige Enabler, Lifecycle-Management und Operations.

Die Grafik weist übrigens noch Produkte mit „Core Functionality“ auf. Diese gehörte früher zum OpenStack-Kern „DefCore“ – im Unterschied zum „Big Tent“, was quasi mehr die äußeren Schalen der Zwiebel kennzeichnen sollte. Inzwischen ist von dieser Differenzierung nur noch sehr ungern die Rede. In Sydney hatte die Leitung erstmals öffentlich von der Notwendigkeit gesprochen, eine andere Organisationsform zu finden. Eine Lösung ist noch nicht gefunden.

## Security bleibt ein Thema

Das Thema Security schlägt sich bei OpenStack unverändert nicht in einer eigenen Lösung nieder, sondern ist Sache jedes einzelnen Projekts – mit OpenLab im Hintergrund. So verfügt das aktuelle Release Queens über rollenbasierte Zugangskontrolle. Johan Christenson von City Network berichtete zwar, das Unternehmen habe sich mit OpenStack als breite Basis für seine Cloud-Angebote einfach an übliche Best Practices gehalten und keine Sicherheitsprobleme erlebt. Joseph Sandoval von Adobe, deren Advertising Cloud zu 90 Prozent mit OpenStack arbeitet (außen vor ist zum Beispiel ERP), meint, Sicherheit sei eher eine Frage der gesamten IT-Umgebung. Doch die Atmail-Vertreter Jason Brown und Jay Sil erklären: „Wir müssen in OpenStack mehr für Sicherheit tun, vor allem bei Containern und Serverless Computing.“

Während sich derart einstige Ankündigungen im Hintergrund verloren, wurde anderes in den Vordergrund geschoben. Das waren zwei Produkte mit irritierenden Implikationen und ein neuer, offenbar rasant wachsender Arbeitsbereich des OpenStack-Projekts. Dazu mehr morgen im zweiten Teil dieses Berichts.

■ *Ludger Schmitz*

## Treuhändermodell angeblich nicht mehr nach gefragt

# Ist die Deutsche Cloud am Ende?

**Vor kurzem ließ Microsoft mit der kolportierten Aussage aufhorchen, dass die Nachfrage nach der Microsoft Cloud Deutschland (MCD) und damit nach einer Public Cloud, die mit deutschen Datenschutzrechtsverständnis konform geht und bei der T-Systems als Treuhänder fungiert, nur wenig nachgefragt würde. Wir haben uns in der Branche umgehört, ob sich das Treuhändermodell für die Deutsche Cloud erledigt hat.**

Selten war der Datenschutz präsenter in den Köpfen als heute, kurz nach Inkrafttreten der DSGVO. Doch Insider treibt das Thema seit Jahren um, insbesondere, wenn es um den Datenschutz vor amerikanischen Behörden aller Art geht. Cloud-Angebote von amerikanischen Unternehmen wurden daher lange Jahre verschmäht. Microsoft reagierte darauf, in dem es die Microsoft Cloud Deutschland (MCD) initiierte. Darin wurden und werden Kundendaten ausschließlich in Deutschland gespeichert. „Die Kontrolle und Entscheidungsgewalt über die Daten obliegt den Kunden selbst. Die Tochtergesellschaft der Deutschen Telekom T-Systems – der Datentreuhänder – agiert unter deutschem Recht und überwacht den Zugriff auf die Kundendaten.“

## Datenschutz hat weiter Priorität

Dieses Modell steht nun am Scheideweg – im März erklärte Microsoft via Handelsblatt, dass man neue Rechenzentren in Deutschland bauen werde – ohne T-Systems als Treuhänder. Die Publikation berichtete, dass die Nachfrage nach der MCD „gering“ sei. Laut „übereinstimmenden“ Aussagen „zahlreicher Branchenkenner“ sei das Angebot „zu teuer, zu rückständig“.

Das wollten wir genauer wissen. Zunächst wendeten wir uns an Frank Strecker - er verantwortet als Senior Vice President das



Bild: © gemeinfrei / CC0

**Branchenstimmen  
zum Status quo  
Deutsche Cloud.**

Cloud-Geschäft von T-Systems. Sieht er einen Hang der Anwender zur Public Cloud OHNE deutschen Datenschutz? Seine kurze Antwort: „Nein.“

Das unterschreibt auch der Analyst Matthias Zacher, Research & Consulting Manager bei IDC: „Unsere Analysen und unsere Gespräche mit Unternehmen zeigen immer wieder, dass die Mehrzahl der Unternehmen auf eine Datenspeicherung in Deutschland oder auf einen Vertrag nach deutschem Recht Wert legt. Viele der großen Cloud-Anbieter bieten aus diesen Grund Cloud Services aus Standorten in Deutschland an. Zudem profitieren die Unternehmen dann ebenfalls von geringen Latenzen. Technische, kulturelle und rechtliche Aspekte gehen bei der Nutzung von Cloud Services Hand in Hand.“

Zacher räumt allerdings ein, dass Unternehmen für viele Anwendungsszenarien aber auch einen europäischen Datenstandort bzw. europäisches Recht akzeptieren. „Für alle Cloud Services gilt, dass die Leistungen zu solchen Kosten angeboten werden müssen, die für eine möglichst große Zahl an Unternehmen interessant sind und das die Angebote State-of-the-Art sind.“ Das sieht auch T-Systems-Cloud-Chef Strecker so: „Es geht darum, welche Daten in der Public Cloud verarbeitet werden sollen. Nehmen wir als Beispiel sensible Patientendaten. Das Start-up Tele-clinic.com, das Sprechstunden aus der Cloud anbietet, setzt auf unsere Public Cloud aus einem deutschen Rechenzentrum. Mit der europäischen Datenschutzgrundverordnung werden viele Unternehmen jetzt zusätzlich sensibilisiert, auf Compliance-Regeln zu achten. Aber natürlich ist es auch so, dass am Ende auch der Preis eine Rolle spielt.“

### Multi-Clouds machen Azure und Co gesellschaftsfähig

Im Zeitalter von Hybrid- und Multi-Cloud unterscheiden also auch deutsche Unternehmen genau, welche Angebote zu welchem Preis für welche Daten sinnvoll sind. Der Datenschutz scheint nicht mehr das Totschlagargument gegen die Nutzung von Azure, AWS oder andere Clouds US-amerikanischer Anbieter zu sein. Grundsätzlich habe sich hier viel bewegt, berichtet Andreas Zipser, Vice President Sales Central Europe bei Sage und damit ausgewiesener Experte für den Mittelstand in Deutschland: „Ich erlebe den deutschen Mittelstand in den vergangenen drei Jahren völlig verändert. Vor drei Jahren gab es noch sehr zynische Kommentare zur Cloud, zur Sicherheit in der Cloud, zur Geschwindigkeit, Verfügbarkeit, und und und. Inzwischen ist die Cloud-Adoption-Rate aber sehr viel höher geworden.“

Doch gerade im Mittelstand gebe es große Unterschiede in der Art der Nutzung: „Vor allem kleine Unternehmen setzen auf unsere Public Cloud-Angebote auf AWS-Basis, um darüber vor allem Finanz- und Lohnbuchhaltung zu machen. Alleine in Deutschland sind das 16.000 bis 18.000 Unternehmen. Bei produzierenden Unternehmen mit streng gehütetem Know-how und entsprechenden Intellectual Properties sieht das natürlich anders aus. Denen raten wir zu unserer Private Cloud-Lösung.“

Bis zu 18.000 mittelständische Sage-Kunden schieben also aktuell ihre sensiblen Daten in einer Public Cloud herum, ganz ohne Beteiligung eines Treuhänders – das lässt allerdings Zweifel an der Zukunftsfähigkeit des deutschen Cloud-Modells aufkommen! Erst im April hat Sage die Sage Business Cloud mit Lösungen für Unternehmen jeder Größenordnung vorgestellt. Die neue Lösung integriert Anwendungen von der Buchhaltung für kleine Unternehmen bis hin zu anspruchsvollem Finanzmanagement und branchenspezifischer Software für größere mittelständische Unternehmen.

## Globalisierung lässt Deutsche Cloud alt aussehen

Wir fragten weiter. Mit Christian Gehring, Pre-Sales Director Germany bei VMware, antwortete uns auch ein Vertreter der US-amerikanischen Anbieter. Er hat eine klare Botschaft: „Unsere Kunden nutzen schon geraume Zeit zahlreiche Cloud-Services. Zunächst wurden vor allem Entwicklungsansätze verfolgt, nun wollen die Kunden mehr und die Public Cloud ausgiebiger nutzen. Dabei erfahren wir, dass das Treuhändermodell eigentlich gar nicht mehr nachgefragt wird.“

Gehring kann das plausibel erläutern: „Der Grund dafür ist, dass die ‚Deutsche Azure Cloud‘ auch rein auf Deutschland fokussiert ist, während unsere Kunden durch die Verknüpfung der internationalen IBM oder AWS-Rechenzentren den VMware Cloud Service global nutzen können. So können Kunden ihre Applikationen überall auf der Welt nutzen. Ein Beispiel wäre etwa ein deutscher Hersteller, der in Asien Workloads oder Anwendungen aus der Cloud benötigt. Er kann die Anwendung in Frankfurt entwickeln und hosten und dann bei Bedarf nach Fernost verschieben.“

In dieselbe Kerbe haut Sören Hühold, Head of IT Transformation und Cloud Services bei Arvato Systems: „Entsprechend den Aussagen unserer Kunden würde ich die Deutsche Cloud mit Treuhändermodell als Auslaufmodell bezeichnen. Per Definition handelt es sich dabei um ein Modell für Deutschland, und in dem steckt man dann fest.“

Selbst kleinere Mittelständler mit nur 100 Mitarbeitern sind oft international tätig, und da funktioniert das Treuhändermodell nicht mehr.“

### Die Deutsche Cloud lebt, wird aber immer unbedeutender

Zusammenfassend kann gesagt werden, dass das Treuhandmodell für Daten in der Cloud lebt – aber offenbar tatsächlich immer weniger nachgefragt wird. Der Druck der Globalisierung auf deutsche Unternehmen, die in der Regel dem Mittelstand zuzurechnen sind, zwingt diese zur Nutzung von Multi-Clouds. Dadurch gewinnen vergleichsweise günstige Public Cloud-Angebote an Boden und werden in immer größeren Umfang gerade von kleineren Mittelständlern eingesetzt, denen der Datenschutz nicht ganz so wichtig ist. Ein Malerbetrieb aus der Eifel kann das auch getrost tun, seine Abrechnungsdaten sind für die NSA eher uninteressant. Global agierende Unternehmen aus Deutschland müssen weiter ganz genau überlegen, welche Daten in welcher Cloud abgelegt werden – für einige dürfte das Treuhandmodell unverzichtbar sein.

■ *Dr. Dietmar Müller*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



Hier klicken

## Der Garant im IoT-Business

# Audit-Trails mit IOTA, der Block-chain-Alternative

**Im Reigen der spannendsten IT-Neuerungen sticht die Distributed-Ledger-Technologie „IOTA“ deutlich heraus. Ihr trauen Experten das größte Veränderungspotential als extrem leistungsfähige und fälschungssichere Dokumentation von Transaktionen für das Internet der Dinge (IoT) zu. Was steckt dahinter?**

IOTA (vom griechischen Buchstaben Iota abgeleitet) gehört wie die Blockchain, die ursprünglich einmal für die Kryptowährung Bitcoin entwickelt wurde, zu den Distributed-Ledger-Anwendungen. Wobei IOTA Schwachpunkte der Blockchain als eine weiterführende Technologie elegant eliminiert.

„IOTA kann das bislang fehlende Puzzleteil sein, mit dem die Machine Economy und das Internet-of-Things (IoT) ihr gewünschtes Potenzial zur vollen Entfaltung bringt“, erläutert Walter Graf, Industrie 4.0-Evangelist und Fujitsu Distinguished Engineer. „Wir sehen IOTA als eine Möglichkeit, ein öffentliches, über Firmengrenzen hinausgehendes Rückgrat für das Internet der Dinge zu realisieren, das eine echte Interoperabilität zwischen allen Geräten ermöglicht - egal wo sich diese befinden.“

## Von der Datenbank zu IOTA

Wohlbekannt, ein revisionssicheres, unveränderbares Tracken von Informationen ist in der IT-Welt nicht ganz so rasant neu. Denn Datenbank-basierte Anwendungen aus dem Content-Management-Bereich leisten schlussendlich das Gleiche. Doch das Problem, das diese Ansätze alle gemein haben, ist, dass sie zwar innerhalb einer Organisation noch ganz gut funktionieren, aber nach dem Überschreiten dieser Grenzen Konflikte zutage treten können.

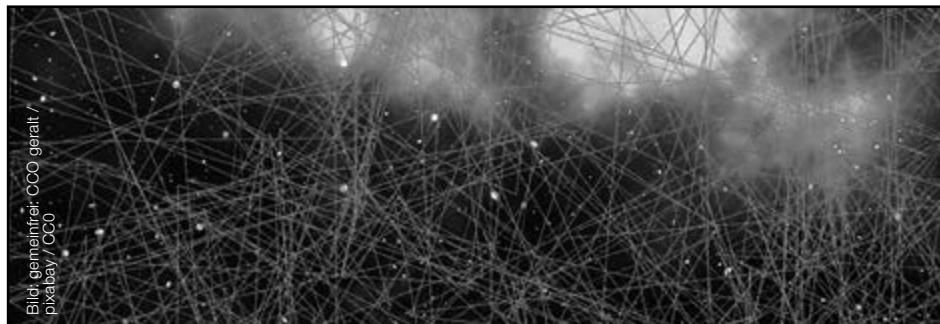


Bild: gemeinfrei: CC0 geralt / pixabay / CC0

**Jeder Teilnehmer im IOTA-Netzwerk, der eine Transaktion durchführt, nimmt aktiv am Konsens teil. Teure Miner sind überflüssig. Damit ist IOTA günstiger als jede Blockchain.**

Sollen zum Beispiel Produktdaten von mehreren Zulieferern in einem Track-Record eines zentralen Systems gespeichert werden, so kann sich das problematisch gestalten. „Statt einer zentralisierten Datenbank bietet sich mit IOTA eine öffentlich verfügbare, neutrale Plattform an, auf der Daten sicher, privat, nachvollziehbar und unveränderbar hinterlegt werden können“, betont Graf. „Darüber hinaus können sich Maschinen gegenseitig über elektronische Brieftaschen bezahlen.“

Im Gegensatz zu einer Datenbank ist die Fälschungssicherheit bereits ein fester Teil von IOTA und das verteilte Wirtschaftsbuch (Distributed Ledger) ist nicht mehr an eine Institution gebunden und bedient insbesondere auch problemlos lange Daten-Lebenszyklen. IOTA kann sich als Netzwerk vom Hersteller, den Zulieferern bis hin zum Kunden über den gesamten Produktlebenszyklus erstrecken. Denn IoT-Produkte sind so designt, dass sie weiterhin Daten sammeln und in einer Plattform ablegen.

## **Einzigartige Features**

IOTA setzt als Verteilte Datenbank nicht wie die Blockchain auf eine verkettete Liste, sondern auf einen gerichteten azyklischen Graphen, der „Tangle“ (Flechtwerk, Gewirr) genannt wird. Dadurch lässt sich eine deutlich höhere Geschwindigkeit (Realtime-Fähigkeit) und fast unendliche Skalierbarkeit bei wesentlich verringerten Transaktionskosten erzielen.

Letzteres geschieht, weil die Miner wegfallen. Denn diese müssen bei der Blockchain die Verifikation und das Abspeichern der Transaktionen durch extrem stromfressende sowie rechen- und zeitaufwendige Prozesse abschließen und dafür auch noch Geld bzw. Kryptowährung erhalten.

Der IOTA-Tangle kennt dagegen kein kostenpflichtiges Miner-Konzept der Validierung. Das bedeutet, dass jeder Teilnehmer des IOTA-Netzwerks, der eine Transaktion durchführen möchte, verpflichtet ist, aktiv am Konsens des Netzwerks teilzunehmen. Dies geschieht, indem er zwei vergangene Transaktionen validiert. Die Auswahl erfolgt dabei über einen Algorithmus.

## **IOTA skaliert mit der Anzahl von Transaktionen**

Das hat weitreichende Konsequenzen: IOTA kann viele Transaktionen pro Sekunde verarbeiten, indem diese parallel verarbeitet werden. Wenn die Anzahl der Teilnehmer zunimmt, werden immer mehr Transaktionen bestätigt. Das heißt, IOTA skaliert pro-

portional mit der Anzahl von Transaktionen. Je größer der Tangle mit zunehmenden Transaktionszahlen, desto schneller und sicherer wird IOTA. „Daher ist gewährleistet, dass niemand in einem IOTA-Netzwerk die Führerschaft übernehmen kann“, unterstreicht Graf. „Diese Tatsche wird ein Erfolgskriterium sein.“

Dank dem Wegfall teurer Miner sind jetzt auch „einfache“ Anwendungen wie zum Beispiel ein fälschungssicherer Tachometerstand möglich. Zu diesem Zweck wird ein digitaler Zwilling vom Kilometerstand in einem IOTA-Tangle abgespeichert. Jeder Fälschungsversuch lässt sich im Distributed-Ledger, dem Wirtschaftsbuch, nun sofort feststellen. Denn dort sind immer die korrekten Daten niedergelegt.

## Showcase für eine IOTA-Plattform

In einem Modell einer simplen Supply-Chain zeigt Fujitsu die Funktion eines verteilten Audit-Trails über zwei Firmen, in dem IOTA seine Stärken gegenüber einer klassischen Datenbank-Implementierung beweist. „Dazu simulieren wir mit zwei Robotersystemen, die Plastikklötzchen auf ein Förderband legen und diese hin und her bewegen lassen, einen fiktiven Produktionsprozess zwischen zwei voneinander getrennten Firmen“, führt Graf aus.

Entscheidend sind in diesem Modell die unterschiedlichen Positionen der Klötzchen, die sie jeweils einnehmen. Deshalb wurden sie vorab mit einem QR-Code in der IOTA identifiziert, um die Erkennung sicherzustellen. Auf dem Förderband werden die Klötzchen gescannt und der Standort erkannt und in der IOTA-Tangle zusammen mit weiteren IoT-Daten als Audit-Trail festgehalten. Auf diese Weise entsteht ein Track-Record aller definierten Stationen und Zustände.

„Wir machten auch die Gegenprobe, indem wir ein Klötzchen für gewisse Passagen vom Förderband nahmen, damit das Klötzchen bestimmte Punkte nicht durchlaufen konnte“, so Graf. „Der nächste Roboter realisierte dann sofort, dass das Klötzchen eine Station verpasst hat, weil er es nicht mehr in der IOTA-Tangle gefunden hat, und sortierte den Klotz aus.“

Prozesse dieser Art können somit weit über die Firmengrenzen hinweg zurückverfolgt werden. Ein weiterer großer Vorteil: Die einzelnen Firmen können ohne eine gemeinsame IT-Infrastruktur über relativ einfache Mechanismen Informationen austauschen.

Gegenwärtig befindet sich IOTA noch in einer Beta-Phase. Das bedeutet, dass es sich bei den aktuellen Java- und Python-Imple-

mentierungen um Referenz-Implementierungen handelt. Daher ist für die nächsten Monate mit weiteren Verbesserungen hinsichtlich Leistung, Funktionalität und Benutzerfreundlichkeit zu rechnen. „Wir von Fujitsu bringen jetzt mit unserem Co-Creator-Ansatz verschiedene Technologie-Stakeholder und insbesondere interessierte Kunden mit ihrem Branchen-Know-how an einen Tisch“, sagt Graf.

„Wir denken, dass es sich speziell bei der Umsetzung eines Audit-Trails, wie wir ihn anhand der beiden Roboter auf der letzten Hannover-Messe vereinfacht gezeigt haben, um einen vergleichsweise naheliegenden und jetzt schon umsetzbaren Use-Case handelt, der sich zusammen mit Kunden auch ohne jegliche Verwendung von Krypto-Währungen realisieren lässt.“ ■ *Otto Geißler*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



## Datacenter-Stromversorgung ohne Umwege

# Das erste Rechenzentrum in einer Windmühle läuft und die Kundschaft ist happy

**Gelegentlich ist es für Rechenzentren in Ballungsgebieten schwierig, an genug Strom zu kommen. Dass es auch anders und vernünftiger geht, zeigt ein Pilotprojekt in Paderborn. Es integriert ein Co-Lokation-Rechenzentrum direkt in den Fuß einer Windenergieanlage von Enercon.**

Das Umland von Paderborn ist windreich. Das hat auch die Energiebranche erkannt und dort eine ganze Reihe von Windparks errichtet. In diesem Monat kann der Kreis Paderborn feiern. Denn übers Jahr gerechnet versorgt er sich vollständig mit Ökostrom – einschließlich aller Industrie- und gewerblichen Unternehmen.

Dennoch kommt es immer wieder vor, dass die Turbinen bei starken Winden und somit zu viel Strom im Angebot abgeregelt werden müssen. In das Netz einspeisen zu wollen und es nicht zu können, ist ärgerlich und bedeutet Energieverschwendungen.

Derzeit lässt sich der Strom wegen des zu langsamen Netzausbau nicht in bedarfsreiche Gegenden abführen. Auch die Bewohner der nahegelegenen Gemeinde Lichtenau können nicht auf Kommando den gesamten Strom verbrauchen.

### Zu viel Strom und keine Abnehmer

Der regionale Provider Westfalen Wind hat seit 2011 etwa 80 Windkraftanlagen gebaut und rund 350 Megawatt Anschlussleistung in der Betriebsführung. Unter anderem gehört dem Unternehmen der Windpark Asseln bei Lichtenau.

Die Bewohner von Lichtenau profitieren von dem Windpark in



**Das eckige im Runden:  
So sieht der Prototyp  
des ersten Rechenzent-  
rums in einer Windkarft-  
anlage aus. Der weitere  
Ausbau sieht mehr  
Stockwerke vor.**

ihrer Nähe unter anderem durch einen vergleichsweise niedrigen-Strompreis von nur 19 Cent pro Kilowattstunde. Patrick Georg, Projektleiter Bau Windparks bei Westfalen Wind, erläutert: „Wir wollen, dass die Menschen hier etwas davon haben, dass wir diese Anlagen bauen.“ Doch die Frage, wie man den regional produzierten Strom erzeugungsnah verbrauchen kann, lässt sich so allein nicht beantworten.

Findigen Ingenieuren und Wissenschaftlern ist diese Verschwendug ebenfalls schon lang ein Dorn im Auge. Ideal wäre, so überlegte sich Gunnar Schomaker, der am Software Innovation Campus Paderborn (SICP) des Universität Paderborn tätig ist sowie Mitbegründer und Ideengeber der Westfalen Wind IT GmbH & Co. KG, die Verbraucher möglichst nah zum Erzeuger zu holen.

Es sei daher eine naheliegende Idee, ein Energie-intensives Rechenzentrum in den riesigen Turm einer Windenergieanlage einzubauen, so Schomaker. Dadurch ergebe sich eine sinnvolle Symbiose aus Windenergie-Anlage und Rechenzentrum. Doch von der Idee bis zur tatsächlichen Umsetzung dauerte es einige Jahre.

## Der Windcore-Prototyp läuft

Im Fuß einer Enercon-115-Windenergie-Anlage entstand der Prototyp eines Rechenzentrums, ein „Windcore“, so die Markenbezeichnung. Derzeit sind Windcores ausschließlich für Windenergie-Anlagen von Enercon konzipiert. Der Stahlbetonturm einer 115er Ankage mit rund 3 Megawatt Leistung hat am unteren Ende einen Innenraum mit einem Durchmesser von rund 13 Metern viel mehr Platz als der Transformator und der Betriebsfahrstuhl der Windturbine benötigen.

Dieser freie Raum wurde als eine gute Grundlage für die Aufstellung von IT-Systemen bewertet. Aufgebaut wurde der Prototyp eines Co-Location-Rechenzentrums, in dem Platz plus Infrastruktur für den Aufbau eigener Rechner und Strom gegen Entgelt zur Verfügung steht.

Der Strom für die Systeme kommt quasi direkt von oben. Bei Bedarf könnte die gesamte Leistung von 3 Megawatt dem Datacenter zur Verfügung gestellt werden oder die Batterien der integrierten USV-Anlage laden. Laut Patrick Georg, Projektleiter Bau Windparks bei Westfalen Wind, „reicht der Strom aber auch bei Schwachwind zu 90 Prozent aus, um die Anlage zu powern.“

Das Rechenzentrum ist derzeit ausgelegt auf einen Strombedarf von rund 60 Kilowatt (vier Racks zu je 15 Kilowatt). Bricht trotz

aller Einbindung in die Stromnetze die Versorgung zusammen, sorgt eine Unterbrechungsfreie Stromversorgung von Socomec, die demnächst gegen solche von Delta ausgetauscht werden, mit ihren Bleigelbatterien für eine Überbrückungsperiode von 20 Minuten, in denen die Systeme heruntergefahren werden können. Allerdings passt sich Westfalen Wind IT als Provider in der Auswahl der USVs den Präferenzen seiner Kunden an.

### **Tier III trotz unkonventionellem Standort**

Die Anlage entspricht sicherheitstechnisch einem Tier-III-Rechenzentrum – alle wesentlichen Komponenten der Infrastruktur sind n+1-fach vorhanden. Die Wände sind aus 40 Zentimeter dickem Stahlbeton, die Ein- und Ausgänge Video-gesichert. Die IT-Racks stecken jeweils in einer Art Panzerschrank, dem „DC IT-Safe“, den ProRZ geliefert hat. Die Systemintegration übernahm Datentechnik Moll.

Telekommunikativ ist das Rechenzentrum über Glasfaser von 1 Gigabit pro Sekunde von der Deutschen Telekom angebunden, Provider ist aber Innofactory. Eine Richtfunkverbindung wird parallel betrieben.

Die Kühltechnik besteht aus in die DC-IT-Safes integrierten und mit einem Wasser-Glykol-Gemisch gefüllten „Coolblades“. Teile der Kühlanlage wurden oberhalb der USVs auf einem Stahlgerüst untergebracht. Nach außen wird die Wärme über einen Freiluft-Tischkühler abgegeben, die einzige RZ-Komponente außerhalb des Turmes.

Bei höheren Außentemperaturen hilft eine im Rechenzentrum stehende Kältemaschine, die gewünschte Eingangstemperatur sicherzustellen. Die derzeit testweise genutzte Eingangstemperatur ist 21 Grad Celsius, die Ausgangstemperatur liegt bei 29 Grad Celsius. Angestrebt wird eine PUE (Power Usage Effectiveness) von 1,2, doch lässt sich laut Westfalen Wind-IT-Geschäftsführer Fiete Dubberke dazu erst nach einem Jahr Betrieb eine seriöse Bewertung abgeben.

Das Rechenzentrum ist unter anderem über zwei separate 110-Kilovolt-Leitungen mit der Stromtransportinfrastruktur von Westnetz und andererseits mit der von Avacon verbunden. Die Versorgungsqualität für Strom ist somit weit höher als die TÜV VK3 Zertifizierung der technischen Anlage fordert. Die Panzerschränke und die USV stehen auf dem Boden der Windenergie-Anlage. Der vorhandene Raum reicht ohne Probleme aus.

## Die ersten Kunden sind eingezogen

Geboren wurde die Idee im Jahr 2011. 2016 erstellte das SICP eine Machbarkeitsstudie, Anfang 2017 wurde das Projekt Windcores von der Westfalen Wind IT ohne Fördermittel finanziert und umgesetzt. 2018 zogen die ersten Kunden in das Co-Location-Rechenzentrum ein: Der Cloud-Provider Teuto.net aus Bielefeld, spezialisiert auf Open-Source-Services, hostet hier Systeme, bald kommt auch Green IT, ein Systemhaus aus Dortmund, hinzu.

Außerdem wird auch die Universität Paderborn ab diesem Monat einige Systeme in der Windenergie-Anlage betreiben und gleichzeitig erforschen, wie die IT-Anlage sich auf die Netzstabilität auswirkt und was sie leistet. Alle Kunden profitieren von einem sehr niedrigen Strompreis von 15 Cent pro Kilowattstunde. Der erste Prototyp hat also seinen Betrieb aufgenommen.

Weder an der Windenergie-Anlage noch an den IT-Systemen mussten für den Einbau besondere konzeptionelle Veränderungen vorgenommen werden. „Wir mussten allerdings den Hersteller der Anlagen davon überzeugen, dass auch nach Einbau der IT-Systeme und Gewerke die Standsicherheit, die Brandlast und weitere rechtliche Anforderungen erfüllt werden“, sagt Georg. Zudem war Enercon der Gedanke, Zusatzaggregate in der Windanlage zu betreiben, anfangs ganz schlicht sehr ungewohnt – das sieht nun anders aus.

Wie geht es nun weiter? „Wir denken bereits über neue Prototypen nach“, sagt Dubberke. Beispielsweise könne man durchaus Konzepte entwerfen, die in Windanlagen anderer Hersteller mit etwas schmälerem Fuß Platz finden.

Der Windenergie- und IT-Spezialist deutet zudem an, dass der Pilot dazu dienen soll, weitere Kunden und Partner zu überzeugen. Zunächst steht aber der Ausbau der Rechenzentrumsanlage auf 40 bis zu 60 konventionellen Racks an - schon kein kleines RZ mehr und ein Strompreis von 0,15€ kW/h wird hier wahrscheinlich neben dem guten Gewissen wegen der Nutzung Kohlendioxid-freier Energie zu einem starken Argument hinsichtlich der Betriebskosten.

■ *Ariane Rüdiger*

## Anwendungsszenarien für Blockchain

# Blockchain-as-a-Service im Unternehmen nutzen

**Blockchain-as-a-Service ist für Unternehmen vor allem dahin gehend interessant, dass sich Bezahltdienste im internen Netzwerk abwickeln lassen. Die dazu notwendige Infrastruktur wird direkt in der Cloud gebucht.**

Die Blockchain-Technologie wurde vor allem durch Kryptowährungen wie Bitcoin oder Ethereum bekannt. Unternehmen können die Technologie für sich nutzen, um Arbeitsabläufe zu automatisieren, abzusichern und zu verbessern.

## Wie Unternehmen von Blockchain profitieren

Der Vorteil von Blockchain ist, dass die notwendigen Daten nicht zentral auf einem Server gespeichert werden, sondern alle beteiligten Computer über Informationen der einzelnen Transaktionen verfügen. In einem solchen Peer-to-Peer-Netzwerk sinkt also die Wahrscheinlichkeit eines Ausfalls, während gleichzeitig die Verfügbarkeit steigt. Außerdem sind Buchungen stets sicher und nachvollziehbar.

Alle Buchungen im Netzwerk sind für alle Beteiligten jederzeit transparent einsehbar und dadurch sicher und überall im Netzwerk nutzbar. Wird eine neue Buchung erstellt, wird die Blockchain, das Register mit allen Buchungen, auf allen beteiligten Rechnern aktualisiert. So sind alle Rechner immer auf dem neusten Stand. Ein Hacken ist sehr schwer, da Fälschungen auf allen anderen Geräten sofort auffallen.

Blockchain ist für Unternehmen daher eine wichtige Technologie, die sich auch ohne eigene Infrastruktur nutzen lässt. Die dazu notwendigen Ressourcen können direkt in der Cloud ge



Bild: © the\_lightwriter - stock.adobe.com

**Anwendungsszenarien und Anbieter von Blockchain-as-a-Service.**

bucht werden. Dazu bieten mittlerweile zahlreiche Unternehmen, darunter Microsoft, Amazon, IBM, SAP, HPE und Oracle entsprechende Clouddienste an, die sich buchen lassen, um eine eigene Blockchain-Infrastruktur aufzubauen, die vollständig unabhängig von anderen Technologien im Internet ist.. Am Beispiel von Microsoft Azure und Amazon Web Services (AWS) können die Dienste auch zusammen mit anderen Ressourcen des Unternehmens genutzt werden.

So kann Blockchain zum Beispiel für Smart Contracts - also intelligente Verträge - genutzt werden. Eine Aktion löst automatisch eine Gegenreaktion aus, die wiederum eine weitere Aktion auslöst, bis alle Aktionen im Vertrag nachvollziehbar vollzogen worden sind. Alle Schritte des Vertrages sind durch alle angebundenen Systeme erfasst und lassen sich lückenlos nachvollziehen. Interessant ist das vor allem für Logistiksysteme, aber auch andere Anwendungsszenarien sind hier denkbar.

### **Blockchain-as-a-Service produktiv nutzen**

Blockchain wird zum Bezahlen vor allem für Dienste im Internet oder für „Micro Payments“ genutzt. Daher ist es sinnvoll auf Technologien und Infrastrukturen zu setzen, die stabil, sicher und skalierbar mit dem Internet verbunden sind. Mit Blockchain-as-a-Service können Unternehmen eigene Währungen erstellen, und für das Bezahlen von Diensten nutzen. Die Bezahlung kann auch für interne Abrechnungen erfolgen, zum Beispiel für verschiedene Abteilungen oder Niederlassungen. Sobald ein Blockchain-Dienst im Einsatz ist, können Unternehmen die Funktion auf eine Vielzahl von Wegen nutzen, wie die bereits erwähnten Smart Contracts.

### **Verschiedene Anbieter und Funktionen für Blockchain-as-a-Service**

Der Aufbau einer Blockchain-as-a-Service-Infrastruktur ist sehr schnell abgeschlossen. Dienstleister wie Amazon ermöglichen das Erstellen über einen Assistenten, der alle notwendigen Ressourcen automatisch erstellt und konfiguriert. Viele Dienstleister setzen dabei auf Technologien von Ethereum, aber auch auf Hyperledger Fabric (ursprünglich von Digital Asset und IBM initiiert und von der Linux Foundation gehostet). Prominente Beispiele dafür sind Amazon, Microsoft und SAP. Dabei handelt es sich um eine Opensource-Technologie, die in die Blockchain-as-a-

Service-Infrastrukturen eingebettet wird. Die Technologie wird von Unternehmen wie Walmart, aber auch Nestlé genutzt. Naheilgenderweise setzen auch sehr viele Banken auf Blockchain-Technologien.

AWS bietet eine eigene Webseite für Informationen rund um die Blockchain-Technologie in AWS. Amazon stellt mit AWS Blockchain-Vorlagen (Templates) vorgefertigte Funktionen bereit, mit denen Unternehmen Blockchain-Netzwerke schnell und einfach selbst in der Amazon-Cloud erstellen können.

Oracle stellt in der „Oracle Cloud“ ebenfalls Blockchain-Technologien zur Verfügung. Auch hier kann über einen Assistenten alles Notwendige gebucht werden. Oracle Blockchain Service ermöglicht das Anbinden verschiedener Technologien und kann intern genauso gut genutzt werden, wie extern, oder auch zusammen mit Lieferanten.

### **Blockchain-Infrastruktur per Assistent erstellen**

Die meisten Anbieter von Blockchain-as-a-Service-Infrastrukturen verfügen über Dienste, in denen sich virtuelle Computer, Container, Speicher sowie die Verwaltung der Identitäten und alles weitere unabhängig buchen lassen. Beim Einsatz von Blockchain-as-a-Service werden diese Dienste miteinander kombiniert und zusammen eingesetzt.

Die Ressourcen werden über Assistenten kombiniert und konfiguriert. Nachdem die Dienste gebucht sind, können sie sofort eingesetzt werden. Über hybride Netzwerke und Dienste können Funktionen aus der Cloud nicht nur im Internet und intern im Clouddienst genutzt werden, sondern auch im internen Unternehmensnetzwerk und zusammen mit externen Partnern und Dienstleistern.

### **Microsoft Azure Blockchain Workbench**

Microsoft Azure hat den Vorteil, viele Funktionen zu bieten, die über Blockchain-as-a-Service hinausgehen. Der Vorteil dabei besteht darin, dass beim Einsatz von Blockchain über Microsoft Azure auch Daten und Funktionen aller Azure-Dienste genutzt werden können. Blockchain in Microsoft Azure kann zur Authentifizierung zum Beispiel auf Azure Active Directory setzen. Aber auch alle anderen Dienste in Microsoft Azure stehen für Blockchain in Azure zur Verfügung. Interessant sind hier vor al-

lem die Dienste rund um Big Data, virtuelle Server, Speicher-dienste, Netzwerke oder Funktionen von Drittherstellern, die im Marketplace gebucht werden können.

## Fazit

Blockchain hat mittlerweile auch in Unternehmen Einzug gehal-ten. Große Cloud-Anbieter wie IBM, Microsoft, Amazon, SAP, Oracle und HPE bieten Funktionen für die Erstellung einer ei-genen Blockchain-Währung in der Cloud an. Der Vorteil dabei besteht darin, dass Unternehmen keine eigene Infrastruktur be-reitstellen müssen, sondern komplett auf Funktionen in der Cloud setzen können.

Die Lösungen in der Cloud sind häufig sicherer und skalierbarer als der Betrieb einer eigenen Lösung. Es lohnt sich also die Mög-lichkeiten in Betracht zu ziehen, die Microsoft, Amazon und Co bieten. Wer bereits Dienste in der Cloud nutzt, zum Beispiel in Microsoft Azure, hat den Vorteil, gleich in die Technologie ein-steigen zu können.

■ *Thomas Joos*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



Hier klicken

## All-Flash ändert alles

# Erbarmungslos nahe - Der Verzicht auf RAID-Arrays

**Die Datenexplosion hält an, der Druck der Echtzeit-Fügbarkeit wächst. Zum rund 30-jährigen Jubiläum ist bei RAID die Luft raus. Und jetzt?**

RAID ist tot. Umständliche Einrichtung, starre Konfiguration, teure Upgrades, überlange Rebuild-Zeiten: Der Verzicht auf RAID war jahrelang in der Mache. Jetzt ist es endlich soweit.

Das jähre Ende von RAID wurde von Branchen-Insidern bis zur Ermüdung gepredigt. Es hieß, die immer komplexeren RAID-Level —RAID-5, RAID-6, bald auch RAID-7.N — würden das Ableben der Technologie besiegen. Denn die HDDs würden immer größer, die Rebuild-Times immer länger und so müssten die betroffenen RAID-Arrays im Ernstfall viel zu lange quasi mit angezogener Handbremse im „Degraded Mode“ unproduktiv vor sich hin werkeln.

Doch Totgesagte leben offenbar länger. Mit dem Einzug von All-Flash-Storage ins Rechenzentrum entfiel auf einmal die einst wichtigste Begründung für die trüben Unkenrufe zum Untergang von RAID: die lahme Schreib- und Lese-Performance von HDDs. Mit Flash-Storage sollte auf einmal alles besser werden.

## Herkömmliches RAID versagt

Doch die Inflexibilität konventioneller RAID-Arrays lässt sich durch den bloßen Austausch von HDDs zugunsten von SSDs nicht aufheben. Ganz im Gegenteil: Bei der Umsetzung von hochperformanten und langlebigen All-Flash-Arrays versagt herkömmliche RAID-Technologie kläglich. Eine der wichtigsten Ursachen



Bild: © Pure Storage

dafür liegt in der beschränkten Lebensdauer von NAND-Speicher, die sich in einer festen maximalen Anzahl von physischen Schreibvorgängen reflektiert.

Es gilt, die Anzahl von Schreibzugriffen zu minimieren und eine ausgewogene Verteilung der Schreiblast über alle Medien hinweg sicherzustellen, um eine gleichmäßige Abnutzung zu gewährleisten. Gängige Verfahren beinhalten Nullblockerkennung, Datendenduplikation und -Kompression, Thin Provisioning und den Abnutzungsausgleich auf NAND-Ebene.

Software definiertes Storage verträgt sich nicht mit der relativ starren Unveränderlichkeit konventioneller RAID-Arrays. Und doch liefern noble Anbieter wie Fujitsu (in „Eternus“), Dell EMC (in Produkten der „VMAX“-, „Unity“- und „SC“-Reihe) oder „Hitachi Vantara“ (mit der „F“-Serie und einer Garantie auf 100 Prozent Datenverfügbarkeit) konventionelle RAID-Arrays für gutes Geld.

Gibt's denn nichts Besseres?

## Softwaregesteuerte Flash-Arrays mit NVMe

In der Vergangenheit basierten die meisten Storage-Systeme der Enterprise-Klasse auf einem Scale-up-Modell mit zwei Controllern (so hat seinerzeit auch Pure Storage angefangen). In den letzten Jahren sind neue Architekturen entstanden, darunter Scale-Out, Software-Defined, konvergente und hyperkonvergente Infrastrukturen usw. Diese neuen Storage-Konzepte bieten Vorteile für hochskalierbare NGA-Workloads (Next-Generation Applications), die bei Dual-Controller-Designs schnell ins Stottern geraten würden.

Im Magic Quadrant for Solid-State Arrays listet das Gartner-Institut als aktuelle Marktführer in erster Linie die Anbieter Pure Storage, Dell EMC, HPE und NetApp, gefolgt von IBM, Hitachi Data Systems und Kaminario. Auch IDC hat Pure Storage im Dezember 2017 die Technologieführerschaft bescheinigt („IDC Market Scape: Worldwide All-Flash Array 2017 Vendor Assessment“).

## Bei Pure Storage

Pure Storage bietet mit „Flash Array“ eine Produktlinie für den Blockspeicher latenzsensitiver Anwendungen wie VM-Volumes. Die eigene Lösung für den Datei- und Objektspeicher taufte der Anbieter auf den Namen „Flash Blade“ und schrieb sich dabei massive Parallelisierbarkeit auf die Fahnen.

Die Flash-Array-Produktfamilie nutzt ein proprietäres RAID namens Raid 3D mit „Direct Flash“ auf NVMe-Basis oder Flash-SAS. Bei NVMe (NVM Express oder Non-Volatile Memory Host Controller Interface Specification, kurz: NVMHCIS) handelt es sich um eine offene logische Geräteschnittstellenspezifikation für den direkten Zugriff auf nichtflüchtige Speichermedien über PCIe (einen PCI Express-Bus).

„Flash Array //X“, die High-End-Edition der Produktfamilie, stellt das weltweit erste Storage-System mit direktem Zugriff auf Flash-Arrays (Direct Flash) dar.

„Purity“ (kurz POE für Purity Operating Environment), die Betriebsumgebung von Pure Storage, hebt sich durch ihre native Unterstützung von NVMe hervor. Direkter Speicherzugriff der CPU auf Flash-Speicher via NVMe fördert durch den Verzicht auf Altlasten-Protokolle mit latenzlastigem Overhead und durch die native Parallelisierung von Zugriffen auf Flash eine präzedenzlose Leistung zu Tage.

## HDD-Emulation ist ineffizient

Konventioneller SSD-Speicher erfordert nach wie vor serielle Altlasten-Technologien wie den NFS-Stack, SAS und SATA; SSDs werden hierbei mit Hilfe einer HDD-Emulation angesteuert. Wer diesen Ballast noch mit schleppt darf sich auch über Latenz nicht wundern.

Die HDD-Emulation erzwingt nämlich nicht nur eine ineffiziente Blockspeichernutzung, sondern auch noch teure Überprovisionierung auf Grund von zusätzlichen Platzanforderungen für Garbage Collection. Flash Array //M unterstützt neben Direct Flash im Übrigen aus Gründen der Kompatibilität unter anderem auch SAS- und SATA-Medien im HDD-Emulationsmodus.

„Flash Array //M“, die Upgrade-fähige Einstiegslösung der Flash-Array-Serie, basiert auf dem üblichen Zwei-Knoten-Controller-Modell mit einem signifikanten Unterschied gegenüber Lösungen der Mitbewerber. Der NVRAM, welcher die Schreibvorgänge mit geringer Latenz bestätigen und Dienste wie die adaptive Datenreduktion ermöglichen soll, befindet sich in den ersten beiden Speicherschachteln und nicht in den Speicher-Kontrollern. Durch die Verlagerung von NVRAM in die persistente Speicherschicht (Flash-SSD) gelang es Pure Storage, eine zustandslose Speicher-Controller-Architektur zu konstruieren.

Im Evergreen Storage-Upgradeprogramm von Pure Storage lässt sich FlashArray //M nicht-disruptiv auf //X aufrüsten: durch den bloßen Austausch der Controller, ein kostenfreies Upgrade von Purity und den Einschub der Direct-Flash-Module. Mit dem Upgrade entfällt die HDD-Emulation und damit erübrigt sich wiederum die latenzlastige Flash-Translation, da Purity jetzt den Speicher von Direct Flash über die neuen Kontroller nativ ansteuert.

## Die Blockzuordnung

Sämtliche Garbage-Collection, den Abnutzungsausgleich (wear leveling), die Blockzuordnung (block allocation), die Flash-Translation (beim Einsatz von SAS- und SATA-Medien im HDD-Emulationsmodus) und dergleichen andere Aufgaben handhabt Purity global in Software statt auf Geräte-Ebene. Die hohe Effizienz dieser Implementierung ermöglicht niedrige Latenzen in stark konsolidierten Umgebungen. Als Resultat eines Upgrades auf Direct Flash mit NVMe halbiert sich die Latenz, verdoppelt sich die Bandbreite und vervierfacht sich die Gesamtleistungsdichte eines FlashArrays (beim Einsatz von 10 Modulen).

Im vergangenen Dezember hat Purity Storage ein aktiv/aktiv Metro-Stretched-Cluster namens „Active Cluster“ auf der Basis von Flash Array vorgestellt. Anwender von FlashArray //M kommen in den Genuss dieser Lösung mit einem kostenlosen Upgrade.

## Der Flash- Stack

Flash Blade nutzt CFM-Speicher (Cooperative Flash Management, die so genannten DirectFlash Modules). Avisierte Anwendungen beinhalten Big Data-Analytics, unter anderem für unstrukturierte Daten aus IoT- und Web-Anwendungen, HPC und Backups. Laut IDC stelle das Produkt „die erste wirklich flash-optimierte Big-Data-Analyseplattform der Branche“ dar.

Pure Storage hat außerdem in Zusammenarbeit mit Cisco eine konvergente Infrastruktur namens „Flash Stack“ entwickelt.

Die Plattform von Pure Storage verwaltet sich mit prädiktiver Intelligenz auf der Basis der hauseigenen KI-Engine Pure1 Meta für „selbstfahrenden Speicher“. Intelligente Software ist offenbar das Geheimrezept für robusten und flotten Flash-Speicher ohne eine HDD-Emulation und ohne RAID-Controller.

## Nachholbedarf bei All-Flash-Arrays und alternative Ansätze

**Dell EMC** stellte die erste Generation der eigenen SC-Storage-

Arrays für All-Flash gerade mal im Herbst des vergangenen Jahres vor. Zuvor war die Reihe (das Resultat der Übernahme von Compellent) für hybride Medien ausgelegt.

In der All-Flash-Ausführung liefern die Arrays bis zu 399,000 Eingabe-/Ausgabeoperationen pro Sekunde (IOPS) und bis zu 10fach mehr pro 10x-Cluster. „SC5020F“ skaliert auf 222-Drives und 2 Petabyte (PB), „SC7020F“ auf 500-Drives und 3 PB. Beide diese Lösungen haben jedoch einen entscheidenden Nachteil: das völlige Fehlen von NVMe-Unterstützung in jeglicher Form. Auch Dells Unity-Storage-Arrays kommen in einer hybriden und in einer All-Flash-Ausführung daher — ebenfalls ohne eine NVMe-Option.

„XtremIO X2 Data Protection“ (XDP) von Dell EMC ist nichts anderes als eine modifizierte RAID-6-Variante mit diagonaler Parität.

Auch **HPE** hat versucht, das eigene Storage-Portfolio im Zuge von mehreren Akquisitionen zu modernisieren. Mit Simplivity erwarb der Anbieter eine Technologieplattform für hyperkonvergente Infrastrukturen in der preislichen Mittelklasse. Mit der Übernahme von Nimble Storage stockte HPE das eigene Portfolio mit einer Alternative zur 3PAR-Reihe auf und bietet nun auch Storage-Lösungen der Mittelklasse mit einem Fokus auf All-Flash-Arrays an.

Die HPE-Simplivity-Plattform nutzt FPGA-Beschleuniger (Field Programmable Gate Arrays) zur Gewährleistung einer vorhersehbaren Leistung bei kontinuierlicher Deduplikation und Datenkompression auf Flash. Bei Nimble Storage handelt es sich um prädiktive All-Flash-Arrays auf der Basis von „Infosight“, einer prädiktiven Analytics-Plattform von HPE. Infosight kann laut dem Anbieter bis zu 86 Prozent der auftretenden Probleme automatisch erkennen und beseitigen.

**Netapp** arbeitet sich durch die Übergangsphase von RAID zu softwaregetriebenen Arrays auch vorsichtig durch. Die All-Flash-Arrays der „AFF-A-Serie“ von Netapp (All-Flash FAS mit SAN über FC oder iSCSI) basieren noch auf RAID DP, einer hauseigenen Implementierung von RAID 6, und auf RAID TEC, einem Feature von Netapps proprietärer Betriebsumgebung „Data Ontap“.

RAID TEC (kurz für Triple Erasure Coding) stellt eine dritte Paritätsdisk bereit. In den Flash-Arrays der EF-Serie verwendet der Anbieter die Raid-Level 0, 1, 5, 6 und 10 mit dynamischer RAID-Level-Migration. In SolidFire-Arrays kommt bereits Netapps eigene, „RAID-lose“ selbstheilende Datenschutztechnolo-

gie „Helix“ für verteilte Datenreplikation zum Zuge. Der Helix-Algorithmus sichert mehrere Kopien der Daten über die Knoten des Cluster hinweg.

### **Softwaregetrieben, aber in Hardware beschleunigt**

Eine Zeitlang wurde Flash für die Verwendung mit unternehmenskritischen, latenzempfindlichen Primärsystemen verwendet und zwar in Workloads, bei denen eine bessere Leistung direkt zu einem besseren Geschäftsergebnis führte. Da jedoch die Kosten für Flash-Storage weiter gesunken sind, ist nun sogar die Verwendung von Flash mit weniger latenzempfindlichen Arbeitslasten aufgrund der sekundären wirtschaftlichen Vorteile von Flash ökonomisch sinnvoll.

Der geringere Energieverbrauch und Platzbedarf erlauben eine höhere Infrastrukturdichte. Die Verwendung von Flash für unstrukturierte Daten erweitert das Spektrum der möglichen Workloads, denn diese waren traditionell (auf Grund technischer Beschränkungen der HDD-Emulation) auf blockbasierte Workloads beschränkt.

Die Verwirrung rund um den Untergang von RAID ist durch eine Inflation von unpräzisen und oft übermäßig vereinfachten Buzzwords verursacht, die von Marketing-Fachleuten in den Markt gepusht und dann von Fachkräften in Datencentern ausgebadet werden müssen. So unterscheiden sich die beiden Begriffe „Storage Virtualization“ und softwaredefinierter Speicher (Software Defined Storage, kurz: SDS) in nur so wenigen Feinheiten, dass sie Missverständnisse verursachen.

Bei SDS ist von einem Software-Layer der Rede, der auf existierende Datacenter-Storage-Ressourcen aufgebracht wird und eine erweiterte Storage-Funktionalität ermöglicht. Der Begriff Storage Virtualization bezieht sich hingegen auf ein Pool von physikalischen und/oder virtuellen Storage-Ressourcen, die sich wie ein größeres Gerät verwenden lassen.

### **Die Zukunft**

Wer die Storage abstrahieren möchte, kann entweder Software Defined Storage oder Storage Virtualization nutzen — mit unterschiedlichen Konsequenzen.

Bei der Storage Virtualization geht es primär darum, größere Datenmengen handhaben zu können. Das Hauptziel der SDS ist die Inte-

gration von Storage-Ressourcen mit gleichzeitiger Performancesteigerung. Daten-Redundanz ist durch SDS nicht explizit abgedeckt.

Laut einem aktuellen Bericht von Gartner soll bis zum Jahre 2021 jedes zweite Datencenter (50 Prozent) für HPC- und Big-Data-Workloads Solid-State-Arrays (SSAs) einsetzen. Letztes Jahr (2017) war es laut Gartner lediglich jedes zehnte Datencenter (10 Prozent). SSD-Anbieter wie Samsung haben bereits angedeutet, dass sowohl die Kapazität von Medien in 3D TLC NAND-Technologie als auch die Kosten für Speicher in aktiver Nutzung (im Gegensatz zum Archiv-Speicher) bald mit HDDs gleichziehen sollen.

■ *Anna Kybylinska*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



**Hier klicken**

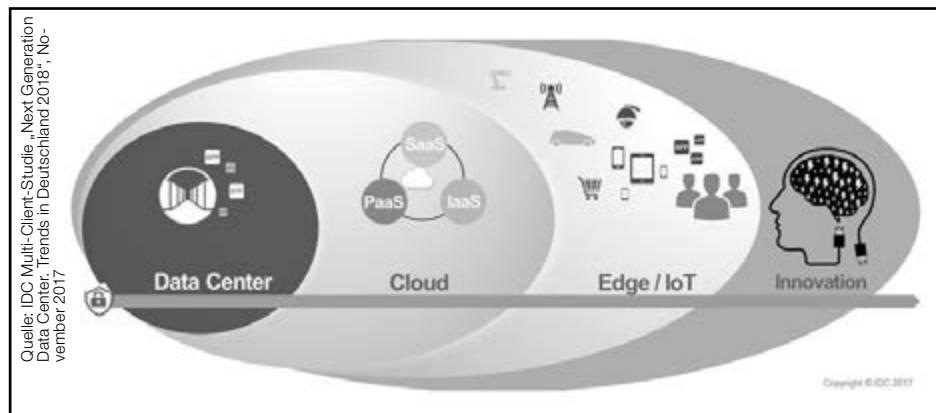
## Die Marschrichtung ist vielen Unternehmen unklar Quo vadis Datacenter?

**Das Marktforschungsinstitut IDC hat interessiert, wie sich die deutschen Unternehmen auf die zukünftigen Herausforderungen von Digitalisierung, Internet of Things und IT-Transformation einstellen. Ergänzt wurde die Präsentation der Umfrageergebnisse am vergangenen Freitag von einer Gesprächsrunde aus Anbietern von Rechenzentrums- und Cloud-Diensten.**

Weitgehend einig waren sich die Manager, dass Cyber-Risiken eine veränderte Infrastruktur im Datacenter und auch mehr Netz- sicherheit erfordern. Ebenso beobachten sie, dass die Heterogenität der IT-Umgebungen dank IoT, Edge- und Cloud- Computing zunimmt und die digitalen Ökosysteme neue Anforderungen an Bandbreiten, Connectivity und Latenz stellen. Dabei rückt auch die engere Verbindung der hauseigenen IT zu Partnern und anderen Marktteilnehmern stärker in den Fokus.

Grundlage der Diskussion war eine Studie, die IDC in diesem Oktober erhoben hat, und zwar durch Befragung in 205 deutschen Firmen mit mindestens 500 Mitarbeitern die IT-Entscheider, die für die Beschaffung von IT-Infrastruktur verantwortlich sind. Inhalt ist der Status Quo ihrer Datacenter und deren Planung.

Kein Wunder also, dass die Befragten in den kommenden zwölf Monaten die Steigerung von Effizienz und Effektivität ihrer Rechenzentren ebenso im Blick haben wie das Dauerthema Kosten- einsparung. Ebenso wichtig ist für sie die Verbesserung von Security und Compliance. Das überrascht nicht; denn 78 Prozent der Befragten verzeichneten in den vergangenen zwölf Monaten bei ihren Datacenter entweder komplett Downtimes oder zu- mindest Einschränkungen bei der Bereitstellung von Services.



**Wer Digitalisierung will, braucht Rechenzentren. Umgekehrt lässt sich eine Datacenter-Transformation nur mit einem übergreifender Ansatz richtig angehen.**

## Fast drei Viertel wollen ihr Datacenter modernisieren

Anbieter von IT-Ressourcen dürften sich freuen, denn 73 Prozent der befragten IT-Manager müssen wegen der „drohenden“ oder bereits eingesetzten digitalen Transformation ihr Data Center erheblich modernisieren. „Eine Erweiterung bestehender Anlagen reicht ihnen nicht aus“, spezifiziert Matthias Zacher, Manager Research & Consulting bei IDC, die Ergebnisse.

Für ihn müssen Digitale Transformation und die Transformation der Datacenter als „ein übergreifender Ansatz“ gesehen werden. Konkret muss das DC um die Cloud (mit SaaS, PaaS und IaaS), Edge- und IoT-Strukturen sowie Innovationen wie etwa „Composable IT“ ergänzt werden.

Erreicht wird das durch sechs Building-Blocks, die die vorhandenen IT-Ressourcen in modernen Data Center ergänzen: Software Defined Infrastructure (vor allem wegen Automatisierung), Multi Cloud (Nutzung von zwei oder mehr Cloud-Diensten für einen Workload), Virtualisierung, Co-Location, für die sichere Verbindungen auch zu Lieferanten und anderen Partnern sowie converged- oder hyperconverged Systeme, die Hardware und Software entkoppeln.

Als innovativer Building Block kommen laut IDC Composable Infrastrukturen zum Tragen, bei denen die physischen Ressourcen im Datacenter (Rechner, Speicher, Netzwerke) als Service betrachtet und bei Bedarf in Pools zusammengefasst werden, also praktisch eine Virtualisierung des gesamten Rechenzentrums.

## Was von dem haben die befragten Unternehmen bereits umgesetzt?

Im Bereich Software Defined Infrastructure (SDI) nutzen 33 Prozent OpenStack als Pilotprojekt oder sogar im produktiven Betrieb. 34 Prozent verfügen über Co-Location-Strukturen und Server Housing, immerhin 26 Prozent lösen ihr SAN komplett durch eine hyperkonverged Infrastruktur ab. Aber mehr als 90 Prozent der Befragten haben ihre Server virtualisiert.

Weniger als ein Prozent nutzen schon die Composable IT - das dynamische Bündeln und erneut Verteilen von IT-Ressourcen hat nicht einmal eine ‚feste‘ Definition, weshalb die Befragten unterschiedliches darunter verstehen - und wer zwei oder mehr Cloud-Dienste parallel (Multi Clouds) im Einsatz hat, konnte IDC nicht ermitteln. Immerhin sollen aber 88 Prozent immerhin über eine Cloud-Strategie verfügen.

Auch über die Verwaltung der (neuen) Datacenter machen sich die Verantwortlichen Gedanken. 33 Prozent wollen in den kommenden zwölf Monaten ein Cloud System Management einführen oder bestehendes verbessern (siehe: Abbildung 10). 29 Prozent haben das Container Management im Fokus, 27 Prozent die Verbesserung von Help Desk und Ticketing. Interessant auch, dass mehr als ein Viertel Analytics für den IT-Betrieb einführen oder verbessern möchte.

## Unternehmen leiden unter Komplexität der IT

IDC organisierte eine Gesprächsrunde mit Vertretern von fünf Anbietern von Datacenter-Produkten und Cloud-Diensten. Ziemlich einig war sich die Runde, dass die IT-Verantwortlichen in den Unternehmen, insbesondere im Mittelstand, unter der wachsenden Komplexität leiden. Manfred Kessler, CEO vom Münchener Cloud Service Provider Global Access Internet Services greift als Beispiel die hyperkonvergenten Systeme heraus, die nicht so einfach zu nutzen sind, wie vielfach propagiert: „Von wegen plug and play“, berichtet er aus der Praxis und weist zudem darauf hin, dass solche Systeme nicht für alles im Datacenter geeignet sind.

„Manche Applikationen benötigen die volle Power, da nutzt ein konvergentes System wenig.“ Zudem sei es schwierig, alte und neue Stacks kombiniert anzubieten, „das klappt noch nirgendwo problemlos“, erläutert er. Alexander Best, Director Technical Business Development bei Datacore, plädiert allerdings für die integrierten Systeme, da diese „den Storage näher an die Applikation heranbringen“.

Ein anderes Thema war die Stabilität der IT. „Das Taxi-Unternehmen Uber darf nicht ausfallen“, sagt Kessler und verweist darauf, dass einige Firmen wie etwa Zalando eigene Layer eingezogen haben, damit der Cloud-Provider schnell gewechselt werden kann. Deshalb setzen 60 Prozent der Unternehmen heute auch auf zwei Standorte für ihre Datacenter, berichtet Donald Badoux, Managing Director von Equinix Germany. Das stabilisiere auch die Kommunikation mit Lieferanten und anderen Partnern, denn „das Eco-System mit der Umwelt wird wichtiger“.

## Der häufigste Grund für ungeplante Downtime: Stromausfall

Datacore-Manger Best weist allerdings darauf hin, dass Stromausfall noch immer der Hauptverantwortliche für Downtime ist. Auch Software-Updates sollen nach Meinung von Dr. Markus

Pleier, Director Germany & Austria bei Nutanix, heutzutage keine Downtime mehr verursachen: „Da muss die Infrastruktur funktionieren“, fordert er. Einig sind sich die Fachleute, dass die derzeitige Bewegung von Top-Cloud-Abiern wie Amazon von einem Tier-3- auf ein Tier-2-Rechenzentrum zu wechseln, deshalb sehr bedenklich sei.

Wie stellen sich die IT-Verantwortlichen in den Unternehmen auf die neue Situation ein? Einerseits sorgt die wachsende Komplexität für Probleme, da sie kaum und wenn, dann nur unter hohen Kosten beherrschbar ist oder – wegen Fachkräftemangel nicht mehr selbst zu bändigen ist. Weichen Firmen dann auf externe Anbieter aus, sieht sich der IT-Leister bedroht und ist es in vielen Fällen auch.

„Die IT-Administratoren haben Angst um ihre Jobs“, berichtet Datacore-Manager Best, obwohl die IT-Mitarbeiter doch für andere Aufgaben eingesetzt werden können. „Zurecht!“ ergänzt CEO Kessler von Global Access Internet Services. „Wer heute nicht agil entwickeln kann oder will oder der Automatisierung m Weg steht, muss gehen.“

### Personalwechsel hinter den Kulissen

Auch Johannes Wagmüller, Director Systems Engineering bei Netapp, erläutert, dass die Automatisierung sich trotzdem nicht aufhalten lasse. Die Frage ist dann, ob sich eine Auslagerung in Cloud-Strukturen rechnet. Nutanix-Chef Pleier vermutet, dass bei zunehmender Automatisierung Vieles wieder zurück ins eigene Rechenzentrum wandern wird.

Die Diskussion darum, ob nicht auch die Anbieter von IT-Ressourcen und -Services ihre Geschäftsmodelle überdenken müssten, beantwortete das Podium damit, dass man den Kunden mehr Ausbildung und Workshops über neue Entwicklungen anbiete oder, wie Netapp es ausdrückt, zu fragen, was am besten für den Kunden passt.

Konkret empfiehlt Nutanix-Manager Pleier, die Unternehmen sollten ihre Workloads hinsichtlich Sicherheit, Lokalität etc. klassifizieren und – ganz wichtig – zu kommunizieren: „Man muss die Menschen bei der Transformation mitnehmen“. Datacore rät zu modularen Schritten um den Übergang in neue Datacenter-Strukturen nahtlos zu gestalten. Und schließlich, so CEO Kessler vom Cloud Service Provider Access, sei ab einer bestimmten Größe ein eigenes Rechenzentrum günstiger als Cloud Computing.

**Fazit:**

Die Wege zum neuen Data Center sind vielfältig, die Marschrichtung unklar, aber Weichen müssen trotzdem jetzt gestellt werden.

■ *Kriemhilde Klippstätter*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



**Hier klicken**

## Offenes Design für Datacenter-Hardware und -Racks

# Open19 Foundation versus Open Compute Project (OCP)

**Der Markt für Datacenter-Hardware ist genauso lukrativ wie hart umkämpft. Doch anders als Hyperscale-Betreiber haben einzelne Rechenzentren branchenüblicher Größe kaum Einfluss auf die Produktentwicklung durch ihre Hardware-Lieferanten. Zwei Initiativen – das Open Compute Project (OCP) und Open19 – wollen Abhilfe schaffen, sind sich aber nicht einig, wie.**

Datencenter-Hardware sei zu teuer, zu umständlich in der Integration und zu wenig interoperabel, bemängeln Datacenter-Betreiber mit der Begründung, bei marktüblichen Lösungen bliebe stets ihre Agilität auf der Strecke. Im Gegensatz dazu können ja die Hyperscale-Datencenter aufgrund massiver Skaleneffekte ihre Hardware zumindest maßgeschneidert fertigen lassen — den anderen sei der Weg bisher aber versperrt. Die Großen ließen sich zudem bei ihrem Handwerk nur ungerne in die Karten schauen.

So die Kritik. Steht denn der Rest der Rechenzentrumsbranche auf verlorenem Posten?

Vor dem Hintergrund der allgemeinen Unzufriedenheit mit dem Status Quo in Sachen Hardware sind OCP und Open19 entstanden. Beide haben es sich zum Ziel gesetzt, das Problem durch quelloffene, interoperable Hardware-Architekturen zu lösen.

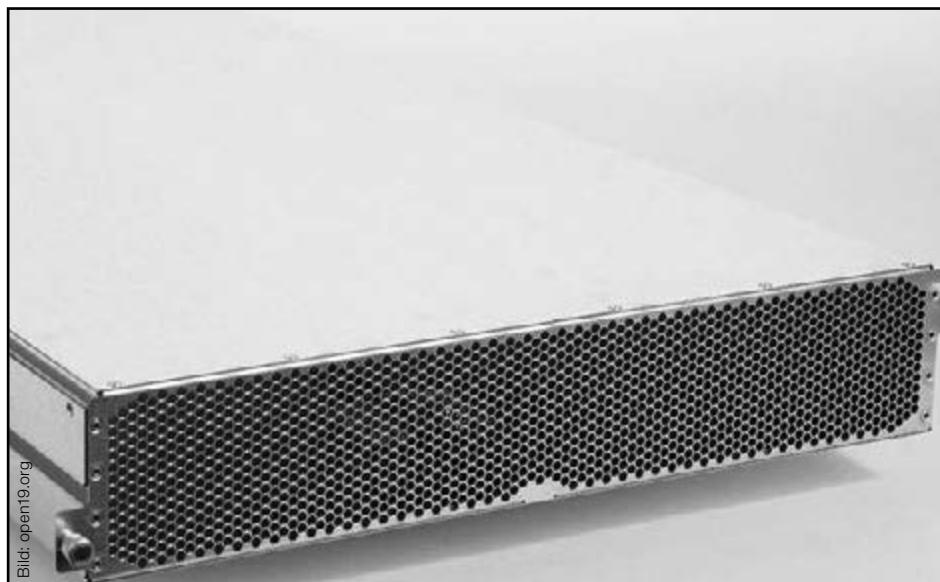


Bild: open19.org

**21 oder 19 Zoll? und andere Fragen: Das Open Compute Project und Open19 konkurrieren um die Spezifikation von Open-Source-Hardware und -Racks.**

## Eine höhere Interoperabilität dank quelloffener Hardware

Datencenter-Hardware existierte bisher in zwei Varianten:

- marktübliche Lösungen von großen Systemanbietern (Dell/EMC, HPE ...), ihren kleineren Herausforderern (Boston Servers, Hyve Solutions, Penguin Computing, Schneider Electric und schließlich noch Baidu, Tencent beispielsweise) und von aufstrebenden Komponentenherstellern wie Wiwynn und Quanta (siehe: „Mit Open nach oben, Das Open Compute Project bei Quanta Cloud Technology“),
- proprietäre Datacenter-Hardware der ultra-effizienten Hyperscale-Rechenzentren auf der Basis top-geheimer Architekturen, zum Beispiel bei Amazon und bis zu seinem Beitritt zu OCP im März 2016 hielt sich auch Google über seine Designs bedeckt.

Unter der Federführung von Facebook entstand dann OCP, das Open Compute Project (siehe: „Datacenter-Design mit quelloffener Hardware für eine Lock-in-freie konvergente Infrastruktur – Wer kennt schon das Open Compute Project?“) in Zusammenarbeit mit Intel, Rackspace Hosting und Goldman Sachs. OCP hat eine Plattform für den Ideenaustausch zwischen Community-Mitgliedern geschaffen und einen Marktplatz für Hardware auf der Basis der eigenen Standards aus dem Boden gestampft.

Doch bei all seinen Erfolgen nahm OCP vorrangig die Bedürfnisse von Hyperscale-Datencenter ins Visier; kleinere Betreiber gingen trotz der vielen Innovationen und Standardisierungsinitiativen im Großen und Ganzen leer aus. LinkedIn sah darin eine Chance, die Zügel an sich zu reißen, um einen breiteren Markt für interoperable Rechenzentrumslösungen mitzugestalten, und gründete in Zusammenarbeit mit Hewlett Packard Enterprise und VaporIO das Open19 Project (Open19 Foundation). Open19 hat sich zum Ziel gesetzt, den Betreibern kleiner und mittelgroßer Datencenter eine Hebelwirkung zu verschaffen, die sich mit der Marktmacht der Hyperscale-Anbieter vergleichen ließe.

## Das Open Compute Project (OCP)

Wie gesagt: Das Design von Hardware für Hyperscale-Datacenter war lange Zeit ein streng gehütetes Geheimnis ihrer Betreiber. Dann kam Facebook mit der OCP-Initiative.

Die Liste der OCP-Mitglieder liest sich heute wie das „Who-is-Who“ der IKT-Branche: Apple, Dell, HPE, IBM, Intel, Nvidia, Microsoft, die Deutsche Telekom (im Übrigen ein Platinum-

Sponsor), die Rittal GmbH & Co. KG (ein Gold-Sponsor). Google hat dem OCP sogar die eigene 48V-DC-Stromverteilungstechnologie beigesteuert, die lange Zeit als eines der am besten gehüteten Geheimnisse des Suchmaschinenriesen galt.

Der OCP-Gemeinde ist es erstaunlich gut gelungen, ein verstärktes Bewusstsein und gesteigertes Interesse an Lösungen für Datencenter-spezifische Produkte zu finden. Doch viele der Innovationen wurden durch den vergleichsweise engen Kreis von Unternehmen getragen, die im Bereich der Hyperscale-Datencenter unterwegs sind, und fanden bei breiteren Scharen von Anwendern kaum Zuspruch. Sind die benötigten Skaleneffekte nicht vorhanden, lösen sich viele Vorteile von OCP-Designs einfach in Nichts auf und so ist auch die Verbreitung in der Praxis entsprechend limitiert.

Das Open Compute Project schlägt sich mit drei Problemen herum, die einem breiteren Erfolg der Initiative bisher noch im Wege stehen:

- die Beiträge von OCP-Mitgliedern haben nicht zu einer Vereinheitlichung der Datencenter-Standards, sondern vielmehr zu einer Fragmentierung geführt,
- OCP-Spezifikationen an sich sind fragmentiert; sie beschreiben jeweils nur eine Teillösung im Rahmen eines größeren Problems,
- die Lieferkette von OCP-Anbietern ist nicht darauf ausgelegt, markttypische Datencenter zu beliefern.

## OCP-Probleme

In der Praxis ist ein Lösungsansatz, der sich für Hyperscaler wie Facebook, Google oder Microsoft eignet, oft nicht auf kleinere Datencenter anwendbar. Zudem haben die OCP-Mitglieder nicht allgemein anwendbare, sondern in vielen Fällen ganz spezielle Lösungen zur OCP-Standardisierung eingereicht, zum Beispiel Racks in 21 Zoll, die sich ohne Anpassungen für 19 Zoll-Hardware nicht nutzen lassen).

Als Resultat daraus kam es zu einer Fragmentierung der Datencenter-Spezifikationen, mangelnder Effizienz in der Lieferkette und zur allgemeinen Verwirrung der IT-Käufer. Zum einen gibt es Hardware mit der Bezeichnung „OCP certified“ - hier hat der Hersteller die zugehörige IP an OCP übertragen -, zum anderen „OCP inspired“ - die einem Design zu Grunde liegende IP gehört nicht der OCP-Stiftung.

Es fehlt OCP eine Art roter Faden, der klar festlegen würde, was sich ein OCP-Design nennen darf und was nicht. Die OCP-Stiftung hat sich bemüht, mit einer Reihe von Standards diese und andere Fragen zu klären. Doch der derzeitige Standardisierungsprozess der OCP steckt noch in den Kinderschuhen; er ist derzeit schlicht nicht in der Lage, zuverlässige Metriken für den Erwerb von OCP-basierter Hardware zu liefern.

Daraus resultiert ein anderes Problem, nämlich die Tatsache, dass eine OCP-Spezifikation erfahrungsgemäß nur eine partielle Lösung einer größeren Aufgabenstellung beschreibt.

### Die Bereitstellungslücke

Bereitstellungsmodelle für Open-Source-Initiativen im Softwarebereich sind offenbar nicht so leicht auf Hardware anzuwenden, da sich hier in der Praxis viele zusätzliche Komplikationen stellen. Die OCP-Stiftung kann zwar mit der einen oder anderen Spezifikation die Grundlagen liefern, doch es liegt immer noch an den jeweiligen Herstellern, diese in greifbare Produkte umzusetzen. Anders als im Falle von Software sind im Falle von Hardware nicht vorrangig die Anwender, sondern die jeweiligen Hersteller für die Komponentenbeschaffung, Fertigung und Logistik, Prüfung und Zertifizierung sowie die Wartung und den Support zuständig.

Für Datencenter-Betreiber, die traditionelle Zulieferer von Hardware und Hardware-Designs gewohnt sind, erfordert OCP trotz aller Standardisierungsbemühungen vor allem die Umstellung auf eine andere Art der Versorgungskette.

Praktisch jeder Anbieter von Server-Hardware, der etwas auf sich hält, nimmt auch für sich in Anspruch, ein OCP-basiertes Produktdesign entwickelt zu haben, nur liefert seine OCP-Hardware nicht in relevanten Stückzahlen aus. Diejenigen, die es geschafft haben, adressieren vorrangig Hyperscale- und HPC-Kunden, welche über eigene ausgewachsene Teams von Hardware-Ingenieuren verfügen, so dass sie ohne externen Support auskommen. Für die Mehrheit der Datencenter sind Modalitäten wie fehlende Garantie- und Serviceleistungen einfach nicht zumutbar.

Die Lieferkette von OCP-Dienstleistern ist schlicht und ergreifend nicht darauf ausgelegt, Mainstream-Datencenter-Betreiber zu beliefern.

### Open19 und seine Vorteile

Open19, vorgestellt im Juli 2016, entstand aus der Initiative von LinkedIn; durch seine Übernahme des sozialen Netzwerks ist der

OCP-Mitglied Microsoft de facto jetzt auch an Open19 beteiligt. Open19 verfolgt die Vision offener Server- und Rack-Technologie für das Datencenter der Zukunft.

Das Projekt hatte ursprünglich zum Ziel, für die eigenen Datencenter von Linkedin durch optimierte Designs zusätzliche Effizienzen herauszuarbeiten. Die Offenlegung der Spezifikationen hat eine Gemeinde gleichgesinnter Unternehmen hervorgerufen, welche nun mit der Hebelwirkung eines Hyperscale-Betreibers den Markt mitgestalten soll.

Open19 beabsichtigt, einen quelloffenen Standard — frei von Patentansprüchen und -gebühren Dritter — für Datencenter- und Edge-Computing-Lösungen in dem Formfaktor der 19-Zoll-EIA-Racks (Electronic Industry Alliance) zu etablieren — daher auch der Name: Open19. So soll die Anbieter-Basis zur Produktion von Open19-kompatiblen Servern breit aufgestellt werden. Open19 strebt niedrigere Kosten pro Rack, niedrigere Kosten pro Server, eine vereinfachte Integration und eine optimierte Leistungsauslastung an.

## Auslöser: Nervige Installationsroutinen

Open19 definiert einen gemeinsamen Formfaktor sowie die Eckdaten der Stromversorgung und der Netzwerkkonnektivität für Server, Storage und andere Netzwerkausrüstung. Anders als OCP legt die Open19-Spezifikation weder das Design von Motherboards, noch die zulässigen Prozessortypen oder Netzwerkarten fest.

Linkedin betreibt weltweit fünf eigene Datencenter mit einer Stellfläche von bis zu 150.000 Servern. Die Schwierigkeiten in der Integration neuer Systeme machten sich bei Linkedin schmerhaft bemerkbar. Ein qualifizierter Techniker soll für die Bereitstellung eines neuen Racks mit 96 Servern bis zu acht Stunden Zeit gebraucht haben; mit modular aufgebauter Hardware nach den Spezifikationen von Open19 ist dieselbe Aufgabe in einer Stunde 20 Minuten erledigt.

Das wichtigste Geheimnis: Es ist nicht nötig, jeden Server-Knoten an das Netzwerk und die Stromversorgung anzuschließen; der Techniker schraubt den so genannten Brick-Käfig im Rack fest; die einzelnen Brick-Server gleiten dann einfach hinein und fertigt. Diese Art der Hardwarebereitstellung ist sowohl im Kernrechenzentrum als auch in Edge-Rechenzentren an der Netzwerk-kante anwendbar.

## Open-19-Eckdaten

Die Open19-Plattform basiert auf Standardbausteinen mit den folgenden Eckdaten:

- 19-Zoll-Rack mit 4 Pfosten nach dem EIA-Standard, der sich für Server, Storage und Netzwerke gleichermaßen eignet,
- Brick-Käfig in zwei Baugrößen mit 12 und 8 Rack-Einheiten (der Brick-Käfig (siehe: Abbildung 7) ist ein passives Blech-Element, welches den Formfaktor festlegt und automatisch Snap-in-Öffnungen zum Einrasten von Kabeln bereitstellt); jede Rack-Einheit lässt sich für einen Brick (B), einen doppelten Brick (DB), und einen doppelten 2HE-Brick (Double High Brick, kurz: DHB) umbauen,
- Versorgungsregal für die 12-Volt-Stromverteilung, OTS-Leistungsmodule,
- optionale Batterie-Backup-Einheit (Battery Backup Unit, kurz: BBU),
- optionaler Top-of-Rack-Switch (ToR),
- Snap-on-Stromkabel/PCB-200-250 Watt pro Ziegel,
- Snap-on-Datenkabel bis zu 100 G pro Baustein.

Solange sich die Hersteller von Datacenter-Hardware an diese Spezifikationen halten, brauchen sie ihre Designs nicht offen zu legen.

■ *Filipe Pereira Martins und Anna Kobylinska*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



## Ein Vergleich verschiedener Batterietechniken für USV-Anlagen

# Haushoher Vorteil für Lithium-Ionen-Technik im Rechenzentrum

**Alternative, zumeist mobile Stromquellen sorgen dafür, dass die Unterbrechungsfreie Stromversorgung (USV) laufen, wenn der Energieversorger ausfällt. Dafür stehen unterschiedliche Techniken mit spezifischen Vor- und Nachteilen zur Verfügung. Zur Absicherung von IT-Systemen haben sich USV-Anlagen mit Blei-Säure-Batterien durchgesetzt. Jetzt bekommen diese Konkurrenz.**

Die Batteriegeschichte reicht bis vor die Mitte des 18. Jahrhunderts zurück, als 1745 mit der Leidener Flasche ein Vorläufer des Kondensators entwickelt wurde. Um 1800 konstruierte der italienische Physiker Alessandro Volta mit der Voltaschen Säule die erste funktionierende Batterie.

Ein ähnliches System, jedoch nach der Entladung wieder aufladbar, entwickelte der deutsche Physiker und Philosoph Johann Wilhelm Ritter. Es war die Vorform der heute bekannten Akkumulatoren. 1854 baute dann der deutsche Mediziner und Physiker Wilhelm Josef Sinsteden den ersten Blei-Akku aus zwei Bleiplatten in einem Gefäß mit verdünnter Schwefelsäure. Gaston Planté verbesserte dieses Konzept durch eine spiralförmige Anordnung der Bleiplatten – eine Konstruktion, die auch heute noch in Bleiakkumulatoren eingesetzt wird.

Unabhängig von der Technologie werden bei Energiespeichern bestimmte Eigenschaften erwartet, wie lange Lebensdauer und



Bild: Socomec

**Lithium-Ionen-Batterie für USV-Anlagen von Socomec. Sie punkten mit geringerem Gewicht und Platzbedarf, einer Lebenszeit von zwölf 12 bis 15 Jahren, hoher Zyklusfestigkeit, kurzer Ladezeit, geringem Kühlungs- und Wartungsbedarf, sowie ökologischer Verträglichkeit.**

hohe Zyklusfestigkeit, ein hoher Lade-Entlade-Wirkungsgrad, eine geringe Selbstentladung und geringe Standby-Verluste. Zudem sollen sie wartungsarm, preisgünstig, ökologisch unbedenklich und anspruchslos in der Handhabung sein.

Allerdings kann keine der verfügbaren Speichertechnologien bei allen Eigenschaften gleichermaßen punkten. Welche Technologie im konkreten Fall verwendet wird, hängt von den Anforderungen der jeweiligen Anwendung an Kapazität, Verfügbarkeit, Rentabilität, Ladeverfahren, Wartungs- und Betriebskosten ab und ist in der Regel das Ergebnis einer Abwägung.

### **Die Standard-Lösung: Blei-Säure-Batterien**

Eine für unsere digitalisierte Welt wesentliche Anwendung von Energiespeichern sind Rechenzentren. Dort dienen USV-Anlagen dazu, auftretende Netzstörungen, wie sie in der europäischen Norm EN62040-3 klassifiziert sind, zu eliminieren. Dazu gehören Einbrüche, Spitzen, Schwankungen der Spannung sowie Transienten.

Bei einem kurzzeitigen vollständigen Ausfall des Netzstromes stellen sie den kontinuierlichen Betrieb der IT-Last und kritischer Infrastruktur-Komponenten sicher. Dafür nutzen sie gespeicherte Energie. In der Regel wird eine Überbrückungszeit von zehn bis 15 Minuten gewählt. So soll im Fall einer länger andauernden Unterbrechung des Netzstroms ausreichend Überbrückungszeit verfügbar sein, damit die Versorgung der Last an Netzersatzanlagen wie Dieselgeneratoren übergeben werden kann.

Die seit Jahren meistverwendeten Energiespeicher für diese Anwendung sind Blei-Säure-Batterien. Sie werden als umfangreiche Batterie-Anlagen eingesetzt, und zwar mit erheblichem Platz- sowie Klimatisierungsbedarf. Weitere Nachteile bestehen darin, dass sich mit Gleichrichter, Wechselrichter und Netzgerät drei potenzielle Fehlerquellen in der Versorgungskette befinden. Zudem verursachen die Wandlungsprozesse erhebliche Energieverluste.

Blei-Säure-Batterien müssen regelmäßig gewartet und je nach Gebrauchsdauererwartung zum Beispiel alle sechs bis acht Jahre ausgetauscht werden, da ihre Leistung mit jedem Lade-, Entladezyklus auf 80 Prozent der Kapazität (Eurobat) am Ende ihrer Lebensdauer abnimmt. Diese Faktoren treiben, verstärkt durch steigende Anforderungen an Rechenzentren, vor allem hinsichtlich Energie-Effizienz und Platzbedarf, die Entwicklung alternativer Speicherverfahren.

## Flywheels für Kurzzeit-Ausfälle

Eine Ersatzlösung für Batterien, die seit Jahren serienreif und in der Praxis erprobt ist, ist die Schwungrad- oder Flywheel-Technologie, bei der Energie als kinetische Energie in rotierenden Massen gespeichert wird. Hauptmerkmale sind ein der Energiespeicherung in Batterien deutlich überlegener Wirkungsgrad von 99 Prozent sowie eine geringe Verlustleistung von 300 Watt bei 300 Kilowatt Wirkleistung. Die verschleißfreie Lagerung des Rotors hält den Wartungsaufwand niedrig, zudem ist eine Flywheel-Anlage weitgehend unempfindlich gegenüber rauen Umgebungsbedingungen.

Obwohl Faktoren wie eine lange Lebenszeit, Zyklenfestigkeit, Zuverlässigkeit und geringer Platzbedarf für die Technologie sprechen, werden USV-Anlagen mit Schwungradtechnik nur selten im Rechenzentrum eingesetzt. Geschuldet ist dies ihrer kurzen Überbrückungszeit von unter einer Minute und der höheren Erstinvestition im Vergleich zu Batterielösungen. Die Anschaffungskosten amortisieren sich kalkulatorisch meist erst nach mehreren Jahren. Ihre Stärken können Schwungrad-USV-Anlagen in Anwendungen im industriellen Bereich ausspielen, wenn beispielsweise Anlagen und Prozesse gegen Mikrounterbrechungen abgesichert werden müssen, oder in Umgebungen, in denen Substanzen wie Blei nicht zugelassen sind.

## Ultra-Caps für die Kurzstrecke

Für die Absicherung von Kurzzeitausfällen oder als Bridging-Lösung kommen neben dem Schwungrad auch Ultra-Caps in Frage. Sie eignen sich allerdings nur für kleine Leistungsbereiche. Einen serienreifen Speicher für Autonomiezeiten im Sekundenbereich stellen zudem Supercaps dar. Ihr Wirkungsgrad liegt jedoch nur bei 90 Prozent, und die Anschaffungskosten sind deutlich höher als für Standard-Speichermedien.

Alle diese Nachteile der Kurzzeit-Puffer umgeht die von Socomec angebotene Innovation, der Lithium-Ionen-Kondensator. Es handelt sich dabei um einen Hybriden zwischen einer Lithium-Ionen-Batterie und einem Superkondensator, bei dem die Kathode wie bei einer Lithium-Ionen-Batterie Lithium enthält und die Anode wie bei einem Superkondensator Aktivkohle.

Auch von seinen Produkteigenschaften her verbindet der Lithium-Ionen-Kondensator das Beste aus zwei Welten: Er verfügt unter anderem über eine hohe Energie- und Leistungsdichte und eine niedrige Selbstentladungsrate, unterliegt keiner zyklusbe-

dingten Verschlechterung und ist nach einer Entladung schnell wieder verfügbar. Zudem kann er im Temperaturbereich von 20 Grad bis +70 Grad eingesetzt werden.

## Die Brennstoffzelle als Alternative

Wie Akkus, Batterien und Kondensatoren gehört die Brennstoffzelle ebenfalls zu den langbekannten Technologien. Erfunden wurde sie im Jahr 1839 von Sir William Grove. Das ihr zugrundeliegende Arbeitsprinzip unterscheidet sich insofern von Batterien und Akkus, als der Energieträger von außen, aus einem Tank, zugeführt werden muss. Ab den 1960er Jahren spielte die Brennstoffzelle eine wichtige Rolle in der Raumfahrt und wurde infolge der Ölkrise 1973 über die Verwendung in der Luft- und Raumfahrt hinaus bekannt.

Allerdings verläuft ihre Einführung in Branchen wie der Automobilindustrie bis heute stockend; in der Anwendung mit USV-Anlagen ist sie bisher nicht über das Stadium von Pilot-Installationen hinausgekommen. Von ihren Leistungsmerkmalen her könnte sie eine Batterielösung komplett ersetzen.

Zu beachten ist jedoch, dass sie eine Anlaufzeit von 20 bis 30 Sekunden benötigt. Für die Anwendung im Rechenzentrum ist daher eine Überbrückung durch eine klassische USV-Anlage oder eine Batterielösung notwendig, was die Brennstoffzelle weniger zu einer Alternative für den Batteriesatz als für den Diesel-Generator macht.

## Disruptiv im Datacenter: Lithium-Ionen-Batterien

Eine inzwischen ausgereifte Speichertechnologie sind Lithium-Ionen-Batterien. Seit Sony sie 1991 für den Einsatz in Video-Kameras auf den Markt gebracht hat, wurden sie in der chemischen Zusammensetzung verbessert, im Batteriedesign sicherer gemacht und an unterschiedliche Anwendungen angepasst. In den vergangenen Jahren hat vor allem die Automobilindustrie mit ihren hohen Anforderungen an Sicherheit und Temperaturbeständigkeit erhebliche Fortschritte bei der Energie- und Leistungsdichte und bei der Zuverlässigkeit erzielt.

Diese Verbesserungen kommen jetzt dem Einsatz der Lithium-Ionen-Batterien in USV-Anwendungen zugute. Für die USV-Systeme der Baureihe „Delphys Green Power“ verwendet Socomec aufgrund der Leistungs- und Sicherheitseigenschaften Lithium-Ionen-Batterien in der aus der Fahrzeugindustrie bekannten Ausführung als Lithium-Mangan-Akkus. Ein Batterieschrank ent-

hält 16 bis 18 Blöcke mit je acht Zellen in Serie. Ein integriertes, interaktives Batterie-Management-System steuert alle Parameter der Zellen und korrigiert kritische Bedingungen, die die Batterieleistung beeinträchtigen können.

## **Energie-effizient und umweltfreundlich**

Im Einzelnen sind für den Datacenter-Einsatz folgende Eigenschaften der Lithium-Ionen-Batterien hochinteressant:

- Aufgrund der hohen Energiedichte benötigen sie nur etwa halb so viel Platz wie Blei-Säure-Batterien. Flächen, die als Batterieräume verwendet werden oder vorgesehen sind, lassen sich für andere Zwecke nutzen, beispielsweise als Stellflächen für zusätzliche Server-Racks oder für Leistungserweiterungen.
- Sie sind leichter als konventionelle Batterien und wiegen nur etwa ein Drittel so viel. So bringt das für kleine bis mittlere Datacenter geeignete 200 Kilovoltampere-System der Baureihe Delphys Green Power für acht Minuten Backup-Zeit mit Lithium-Ionen-Batterien 1.010 Kilogramm auf die Waage gegenüber 3.100 Kilogramm mit herkömmlichen Batterien und braucht zudem nur die Hälfte der Stellfläche. Die 500 Kilovoltampere-Anlage für große Rechenzentren wiegt 3.150 Kilogramm gegenüber 7.020 Kilogramm der Socomec-Blei-Säure-Alternative und spart knapp 40 Prozent der Stellfläche.
- Bei der Lebensdauer sind Lithium-Ionen-Batterien den Bleibatterien mit zwölf bis 15 Jahren deutlich überlegen, selbst wenn sie in Anwendungen mit häufigen Lade-/ Entladezyklen eingesetzt werden.
- Zudem sind sie weniger empfindlich gegenüber hohen Temperaturen und können ohne weiteres in Umgebungen bis 40 Grad betrieben werden. Dadurch sinkt der Kühlungsbedarf und mit ihm Energieverbrauch und Energiekosten.
- Bei weiteren Umweltaspekten liegen Lithium-Ionen-Batterien ebenfalls vorne: Sie entsprechen mangels Blei als Inhaltsstoff den RoHs-Richtlinien und können vollständig recycelt werden.
- Da Lithium-Ionen Batterien keine Gase abgeben und als vergleichsweise sicher gelten, gibt es bis heute keine besonderen Vorschriften und Beschränkungen für den Batterieraum. Sie können mit den USV-Systemen in einem Raum oder bei kleineren Rechenzentren auch auf der aktiven Fläche aufgestellt werden.
- Da die Raumtemperatur für aktives IT Equipment, USV und

Batterie erhöht werden kann, lassen sich Einsparungen bei der Kühlung realisieren.

Obwohl eine Lithium-Ionen-USV um 20 bis 50 Prozent teurer ist als eine USV-Anlage mit Blei-Säure-Batterien, lohnt sich demzufolge die Anschaffung. Die Einsparungen bei Raumbedarf und Kühlung, der Wegfall von Kosten für Wartung und Batterieaustausch führen dazu, dass die Gewinnschwelle nach zwei Jahren erreicht werden kann. Signifikante Einsparungen bei den Betriebskosten ergeben sich nach sechs bis acht Jahren, nach denen bei einer herkömmlichen Lösung der Batteriesatz ausgetauscht werden muss.

■ *Steffen Breiter*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



## Facility Management und IT: Miteinander reden wäre hilfreich

# Krass: Das Missverhältnis zwischen dem Datacenter als Gebäude und der IT

**Was geht mich das an? IT-ler und die Elektronik- und Kühlungsspezialisten im Rechenzentrum wissen nicht viel voneinander. Sollten Sie aber – unbedingt. Denn wenn die IT sich verändert, kommt das Facility-Management oft nicht hinterher – das Resultat ist ein Missverhältnis zwischen den RZ-Infrastruktur und der IT. Manchmal fehlt schon die notwendige Power.**

Eine ganze Weile galt ein niedriger PUE-Wert als wichtigster Ausweis eines gelungenen RZ-Designs. Doch die vermeintlich so aussagekräftige Zahl täuscht. Damit ein Rechenzentrum energetisch und auch technisch optimal läuft, ist eine umfassende Gesamtbetrachtung nötig.

Das zeigt ein einfaches Beispiel: Viele, gerade leistungsstarke Server oder die sich allmählich verbreitenden rechenstarken grafischen Verarbeitungseinheiten (GPU) nutzen eine On-Board-Kühlung. Die gilt, da in der IT befindlich, bei der PUE-Berechnung nicht als Bestandteil des Kühlsystems, sondern gehört zur IT.

Denn PUE (Power Usage Effectiveness) definiert das Verhältnis zwischen Stromverbrauch für die externe Kühlung und Stromverbrauch für die IT. Letztere steht im Nenner der Gleichung. Wenn also die Summe von Stromverbrauch für Prozessor und On-board-Kühlung besonders hoch ist, sinkt der PUE-Wert – ein mathematisch richtiges, nichtsdestotrotz aber widersinniges Ergebnis.



Bild: © foto\_desch - stock.adobe.com

**Der Inhalt eines IT-Rack aus der Sicht eines Rechenzentrumsplaners.**

## Strom und Kühlung sind das Problem

Zudem befinden sich Kühlmöglichkeiten und IT heute ohnehin oftmals im Missverhältnis: Durch die umfassende Virtualisierung vormals vorhandener Einzel-Server leeren sich die Racks. Doch die neuen, leistungsstärkeren Rack-Server oder Blades, die die alten ersetzen und in virtualisierten Infrastrukturen stecken, verbrauchen mehr Strom und erzeugen damit mehr Abwärme.

So entstehen hohe Punktlasten dort, wo sich das betreffende Rack befindet, dem häufig keine adäquate Kühlmöglichkeit gegenübersteht: Der ganze Raum ließe sich zwar gut temperieren, nicht aber die räumlich konzentrierten Hotspots, die diese Systeme erzeugen – selbst dann nicht, wenn die Kühlleistung der vorhandenen Anlagen nominell durchaus im Stande wäre, die erzeugte Wärme abzuführen.

Deshalb ist bei der Einführung neuer Server das Dilemma häufig groß, weiß Peter Dümig, Senior Server Product Manager bei Dell EMC. „Die Rechenzentren sind oft zehn bis fünfzehn Jahre alt, auf alte Systeme ausgelegt und nicht an sich so stark ändernden Anforderungen ausgerichtet. Hat man Glück, lässt sich eine neue Klima-Anlage einbauen, doch auch das verursacht Downtime und zusätzliche Kosten“, weiß er.

## Punktuelle Kühlung ist schwierig

Oft seien die punktuellen Kühlmöglichkeiten so defizitär, dass in einem Rack, in dem eigentlich 42 HE untergebracht werden könnten, am Ende vielleicht nur ein Chassis mit 16 Einschüben eingebaut ist, weil diese Blades bereits die weggängbaren beispielsweise 9 Kilowatt beanspruchen. „Der Trend zu Multicore-Systemen verstärkt dieses Problem“, meint Dümig. Denn jeder weitere Core erhöht auch den Stromverbrauch und damit den Kühlbedarf.

Ähnliche Erfahrungen macht auch Dr. David Hoeflmayr, Geschäftsführer und Vorstand bei der Thomas Krenn AG, bei Kundengesprächen. „Werden Infrastrukturen virtualisiert, ist die alte Klima-Anlage gerade bei kleineren Rechenzentren oft derart überdimensioniert, dass die Kühlleistung selbst dann zu groß ist, wenn man sie weit weg vom optimalen Betriebspunkt dauerhaft auf niedrigster Stufe betreibt.“

Da der Umzug in einen neuen, besser geeigneten Raum häufig genauso zu teuer sei wie die Anschaffung eines neuen Klimageräts, komme es sogar vor, dass ein Heizlüfter im virtualisierten

Server-Raum dafür sorge, dass die alte Klima-Anlage genügend zu kühlen hat und somit die eigentlich fällige Investition unterbleiben kann.

### Unterschiedliche Sprachen

Gängig sei bei Facility- und IT-Verantwortlichen noch immer ein Lagerdenken. „Beide sprechen unterschiedliche Sprachen, haben unterschiedliche Zeithorizonte und sehen sich auch unterschiedlichen Anforderungen gegenüber“, betont Hoeflmayr. Dazu trage auch der von Herstellern und Entwicklern forcierte schnelle Technologiewandel maßgeblich bei – und natürlich der Drang mancher Administratoren, möglichst sofort von der jeweils aktuellen Technologie zu profitieren.

Der wird von den Geschäftsleitungen angesichts stetig steigender Flexibilitätsansprüche im Kerngeschäft und entsprechenden Versprechungen der Hersteller manchmal durchaus geteilt. „Je nach dem Stil des jeweiligen IT-Managements werden Systeme zwischen jährlich und alle fünf Jahre ausgetauscht“, sagt Hoeflmayr. Fragen an die IT-Verantwortlichen, welche Entwicklung innerhalb der nächsten fünf Jahre on premise geplant werde, seien angesichts der bestehenden Veränderungsgeschwindigkeit tatsächlich kaum zu beantworten. „Das RZ-Gebäude plant man aber für ein bis zwei Jahrzehnte“, sagt er.

Hoeflmayrs Empfehlung: IT-Management und Facility Management sollten mehr miteinander reden und auch planen. Ansätze dafür gebe es aber bisher noch wenig. „In einigen wenigen Unternehmen bekommen IT- und Facility-Management inzwischen einen gemeinsamen Vorgesetzten, etwa den Finanzverantwortlichen oder den CEO. Dann sind sie gezwungen, sich besser miteinander abzustimmen“, sagt Hoeflmayr.

### Wie wäre es mit Regulierung?

Helfen könne auch Regulierung: Die Stadt Frankfurt, ein besonders prominenter deutscher Datacenter-Standort, gebe seit diesem Jahr zukünftigen Rechenzentrumsneubauten einen maximalen Wärme-Output vor. Freilich nutzen solche Regulierungen erst dann wirklich, wenn sie flächendeckend verordnet werden.

Innovative, integrative technische Ansätze wie das im vergangenen Jahr anlässlich der Rechenzentrumsmesse „Future Thinking“ vorgestellte Hot-Fluid-Computing hätten es noch schwer, die Mauern in den Köpfen zu durchdringen, berichtet Hoeflmayr.

Bei dieser von der Thomas Krenn AG entwickelten Technologie wird die vom Serverblade erzeugte Wärme direkt durch eine Wärmeleit-Flüssigkeit abgeführt und zum Beispiel zu Heizzwecken verwendet.

„Es gibt großes Interesse - wir verhandeln mit mehreren Kunden, konnten aber noch nirgends implementieren“, berichtet Hoeflmayr. Zudem finden viele Interessenten zwar das Energiesparpotential der Flüssigkühlung attraktiv, haben aber (noch) keine sinnvolle Nutzungsmöglichkeit für die Abwärme gefunden.

Um hier möglichst flexible Alternativen anzubieten, arbeitet Thomas-Krenn AG jetzt mit einem mittelständischen Anbieter innovativer Adsorptionskälteanlagen zusammen. „Dessen Aggregate könnten mit der Abwärme von Servern und Switches gefüttert werden und sie in Kälte umwandeln“, erläutert Hoeflmayr.

## Aktualisieren und Modularisieren

Doch es gibt auch im Rahmen konventioneller IT-Technologien einige Möglichkeiten, krasse Missverhältnisse zwischen IT und Kühlung zu vermeiden. So empfiehlt Dümg, die Infrastruktur regelmäßig anzupassen und inzwischen bewährte Maßnahmen wie die Einhausung von Warm- und/oder Kaltgängen durchzuführen oder Technologien wie gekapselte Racks einzusetzen.

Modernere Technik sei häufig besser als der gute, alte Doppelboden. „Oft reicht dessen Kühlleistung nicht, um auch die im Rack oben stehenden Geräte ausreichend zu kühlen.“ Sinnvoll sei es, auf Racks mit Seitenkühler auszuweichen, deren Leistung auf volle Belegung mit modernen Multicore-Systemen ausgelegt sei.

Zudem sollten Systeme nicht, wie gern noch üblich, überdimensioniert gekauft werden. „Die CPU-Last von Servern liegt in der Regel bei 30 bis 35 Prozent, daher kann das Netzteil meist kleiner dimensioniert werden“, sagt Dümg. Nebenwirkung: weniger Kosten, Stromverbrauch und Kühlbedarf. Im Einzelfall überflüssige Komponenten wie DVD-Laufwerke solle man gleich weglassen. Und auch bei der Redundanz solle man das Nötige, des Guten aber nicht zu viel tun.

Ein wichtiges Konzept ist die modulare Bauweise von Rechenzentren. „Alles, was in Richtung schrittweiser Ausbau und Modularisierung des RZ-Gebäudes läuft, hilft, Missverhältnisse, unnötigen Raum- und Stromverbrauch zu vermeiden“, sagt Hoeflmayr.

■ *Ariane Rüdiger*

## Open Standard for Datacenter Availability – OSDA

# Batterietechnik statt Diesel und neue Verfügbarkeitsklassen

**Wer ein zertifiziertes Rechenzentrum betreiben will, konnte bisher den Einsatz von Dieselgeneratoren als Notstromerzeuger kaum vermeiden. Deshalb entwickelt The Green Grid nun einen neuen Standard, der auch Batterien und Erneuerbare als Erzeuger als Notstromreserve zulässt.**

Der Stromverbrauch der Rechenzentrumsindustrie wächst unvermindert weiter. Lex Coors, Chief Datacentre Technology and Engineering Officer beim Datacenter-Provider Interxion und daneben unter anderem Gründer und Mitglied des Executive Advisory Board des Uptime Institute, Mitglied des gemeinsamen Forschungsausschusses der Europäischen Kommission zur Nachhaltigkeit sowie der European Data Centre Code of Conduct Metrics Group, sagt dazu: „2020 werden es etwa 60 Gigawattstunden(GWh) sein.“

Das entspricht dem gesamten Verbrauch eines Landes wie Polen oder Thailand. 2005 lag die elektrische Leistung der Rechenzentren weltweit noch bei 17 GWh, was 102 Milliarden Tonnen CO2-Äquivalenten entspricht. Die Versorgung war komplett nicht kohlenstoffneutral.

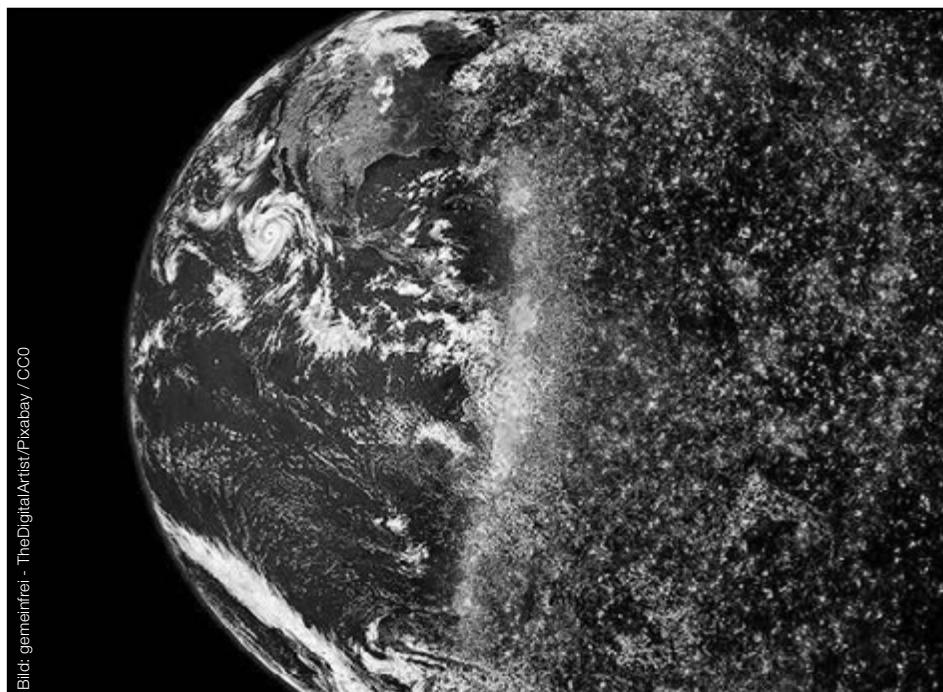


Bild: gemeinfrei - TheDigitalArtist/Pikaboo /CC0

**Der Klimawandel fordert die Nutzung erneuerbarer, nicht immer zuverlässiger Energien, doch die Rechenzentren müssen höchstverfügbar sein. Ein Umdenken ist erforderlich plus neue Batterien und neue Metriken. Der Open Standard for Datacenter Availability (OSDA) von The Green Grid macht sich darn.**

Im Hintergrund wartete überall der Diesel auf seinen Einsatz, und der Primärstrom kam üblicherweise von den damals ebenfalls erst in Erneuerbare einsteigenden Stromversorgern. 2020 werden es, steigt der Verbrauch, wie Coors und andere es erwarten, dreimal so viele CO2-Äquivalente sein. Gleichzeitig habe sich aber die EU das Ziel gesetzt, ihren Kohlendioxidausstoß bis 2020 um 20 Prozent und bis 2050 um 80 bis 95 Prozent zu senken. „Also müssen wir was tun“, sagt Coors (Die neue deutsche Bundesregierung hat ihre Klimaziele schon vor Regierungsantritt gekippt).

### **Nachhaltigkeit soll sein, aber ....**

Damit befindet sich die Rechenzentrumsindustrie in derselben Bredouille wie etwa der Mobilitätssektor: Einem scheinbar unaufhörlich steigenden Bedarf stehen Nachhaltigkeitsbedürfnisse oder -zwänge gegenüber, denn die Zeit tickt. Guter Rat ist also teuer, wenn die EU ihre Nachhaltigkeitsziele tatsächlich einhalten möchte.

Damit, was nun unter anderem im Datacenter-Sektor zu tun ist, beschäftigte sich am 5. Februar des laufenden Jahres die EU-Kommission. Auch Coors war zu der Sitzung eingeladen, bei der es um Initiativen ging, mit denen sich die Kohlendioxiderzeugung im IT-Bereich und bei der Mobilität wirksam reduzieren lassen, ohne Mobilität und IT-Nutzung einzuschränken.

Während im Mobilitätsbereich auch dank des Dieselbetrugs langsam einschneidende und beschränkende Maßnahmen näher zu rücken scheinen, gibt es in der IT begründete Hoffnung auf umsetzbare technische Lösungen. Sie sollen dem Kohlendioxidproduzenten Rechenzentrum zu Leibe rücken, ohne seine Leistungsfähigkeit einzuschränken.

### **Neue Energiespeicher müssen her**

Gefeilt wird an zwei Bestandteilen: verbesserter Energiespeicher-technik und dem von The Green Grid vorgeschlagenen „Open Standard for Datacenter Availability“, der es ermöglichen soll, auch für nichttraditionelle Rechenzentrumsdesigns verlässliche Verfügbarkeitslevel zu ermitteln. Mit den derzeitigen Klassifikationen und Tools geht das nicht.

Zwar gibt es schon viel erneuerbaren Strom, doch ihn als alleinige Ressource für Rechenzentren einzusetzen, ist nach wie vor schwierig. Das wichtigste Hindernis dabei ist die diskontinuierliche Stromerzeugung durch die Erneuerbaren – ruht der Wind und

verdüstern Wolken die Sonne, dann geht im Rechenzentrum das Licht aus, oder unter aktuellen Bedingungen der Dieselgenerator, eine echte Kohlendioxid-Schleuder, an.

„Bisher war Diesel die einzige Alternative, ein sicheres Datacenter-Elektrizitäts-Backup bereitzustellen“, sagt Coors. „Wenn ein Rechenzentrum mit 15 Megawatt DC für 10 Stunden versorgt werden soll, braucht man entsprechend viel Speicherkapazität, und das ist mit den Batterietypen, die es bisher gibt, schwer zu realisieren.“

### Die Nickel-Salz-Batterie: Alter Wein in neuen Schläuchen

Doch neue Batterietechnik macht jetzt Hoffnung: Kombiniert mit den Erneuerbaren, gebe es durch sie, so Coors, nun erstmals die Chance, Gesamtlösungen ohne Diesel zu entwickeln.

Am Massachusetts Institute for Technology (MIT) entwickelt aktuell, unterstützt auch von der Gates-Foundation, Donald R. Sadoway, der dort die John-F.-Elliott-Professur für Materialchemie innehat, mit seinem Team einen im Jahr 1968 entstandenen Batterietyp so weiter, dass er für diese Aufgabe taugen könnte.

Die wiederaufladbare Batterie basiert auf Elektroden aus Salz und Nickel-Chlorid. Bislang war der praktische Einsatz dieses Batteriedesigns an der Membran gescheitert, durch die die Elektronen bei ihrem Weg durch die Batterie diffundieren müssen: Die bislang verwendeten Keramik-Membrane waren zu spröde und zerbrachen deshalb oft unter den nicht immer sanften Einsatzbedingungen. Nur in einigen Spezialfällen bewährten sich Nickel-Salz-Batterien daher praktisch.

### Batterien werden zur Hauptenergiequelle

Das Team um Sadoway ersetzte die Keramikmembran nun durch ein beschichtetes Metallmaschennetz, das sehr viel stärker und widerstandsfähiger ist. Da die übrigen Grundmaterialien günstig, reichlich verfügbar und sicher sind, eröffnet das neue Horizonte. Die Batterien selbst sehr stabil, was viele Lade-/Entladezyklen bedeutet. Sie sind zudem stapelbar und erzeugen kaum Hitze, so dass auch keine aufwändigen gekühlten Umgebungen am Batteriestandort notwendig sind. Es könnte sich also tatsächlich um einen Durchbruch handeln.

Denn mit diesem Batterietyp könnte man Flauten und Wolken am Himmel, die die Solar- und Winderzeugung ausschalten, begreifen: Erneuerbare Energie würde, wann immer möglich, di-

rekt in den Datacenter-Batterieblock eingespeist. Dieser würde das Rechenzentrum ständig mit Strom versorgen, vergleichbar etwa mit einer Online-USV.

Bei Flaute und schlechtem Wetter tritt die Batterie als primäre Ressource in Aktion. Sobald eine der vorhandenen erneuerbaren Ressourcen wieder verfügbar ist, füllen sie die Batterien erneut auf. Die gewünschte Überbrückungsdauer wird durch einen entsprechend dimensionierten Batterieblock sichergestellt.

### Für wen taugt das?

Sinnvoll ist diese Art der Versorgung insbesondere für kleinere und mittlere Rechenzentren, für sehr große würden auch die Batterieblöcke entsprechend groß. Der Charme dieser Idee: Es ist egal, woher der Strom kommt und ob er kontinuierlich von außen in die Batterien fließt, so lange nur die Dimensionierung des Batterieblocks der gewünschten Überbrückungsdauer und Leistung entspricht.

Coors bestätigt: „Sinnvoll ist diese Art der Versorgung insbesondere für kleinere und mittlere Rechenzentren, größere brauchen andere Überbrückungsressourcen.“ Soweit die Theorie. Ob das Ganze praktisch funktioniert, wird sich eine geplante Versuchsanwendung zeigen.

Doch um diese Technologie praktisch im Rechenzentrumsmarkt durchzusetzen, ist noch mehr erforderlich: Nämlich eine Norm, die die zuverlässige Berechnung von Verfügbarkeiten auch für derartige Installationen ohne Notstromaggregat ermöglicht.

### Anders planen!

Genau das tut OSDA (Open Standard for Datacenter Availability): Mit dem vorgeschlagenen Verfahren lässt sich die Verfügbarkeit eines Rechenzentrums sicher prognostizieren, egal, welche Form der Energieversorgung es verwendet und egal wie hoch die angepeilte Verfügbarkeit sein soll.

OSDA baut auf den bestehenden vier Tier-Klassifikationen auf (Nicht redundant, grundlegend redundant, unterbrechungsfrei wartbar und fehlertolerant) und soll in den frühen Planungsstadien des Datacenter-Designs verwendet werden, wenn Bauherren und Architekten grundlegende Entscheidungen über Verfügbarkeit, Dimension und Skalierbarkeit sowie Energieversorgung treffen.

OSDA, so ein Arbeitspapier von The Green Grid, das für Nichtmitglieder des Verbands gegen Entgelt von dessen Website herun-

tergeladen werden kann, basiert auf einer generischen Kalkulationsmethode. Mit ihr lässt sich jedes Design nach den Kriterien Nachhaltigkeit, Energie-Effizienz und Verfügbarkeit bewerten. Sie setzt die Bewertung auch zu den bestehenden vier Verfügbarkeitslevel in Beziehung. Die spätere Nutzungsform des Rechenzentrums ist dabei unerheblich. Am Ende wird für das Gesamtdesign des Datacenter ein Punktwert vergeben, der es ermöglicht, unterschiedliche Rechenzentren miteinander zu vergleichen.

Im Hintergrund stehen Wahrscheinlichkeitsberechnungen von in Serie oder parallel geschalteten Komponentenkonfigurationen, um die vier Level zu beschreiben. Dabei wird für jede Komponente eine eher pessimistische Verfügbarkeitsschätzung verwendet, um überoptimistische Einschätzungen zu vermeiden. Faktoren wie die Skalierbarkeit werden ebenfalls in das Modell einbezogen. Auch die statistische Verfügbarkeit von Designs, die heute keiner Level-Definition entsprächen, sowie die jedweden zukünftigen Designs ließen sich mit OSDA in Zukunft berechnen.

### Für den Einsatz von OSDA beim Datacenter-Design schlägt The Green Grid vier Schritte vor:

1. Ziele definieren und Systemauswahl gemäß dem DCMM (Datacenter Maturity Model).
2. Auswahl der Energieversorgung und Auslegung des Energiesystems
3. OSDA-Bewertung errechnen
4. Nummer 1 bis 3 so lange wiederholen, bis das angestrebte Gleichgewicht zwischen Verfügbarkeit und Nachhaltigkeit erreicht ist.

Geplant ist ein Online-Tool, das es ermöglicht, diese Berechnungen für das eigene Rechenzentrum durchzuführen. Derzeit läuft ein offener Konsultationsprozess, in den sich alle Interessenten mit Kritik und Verbesserungs- oder Änderungsvorschlägen einbringen können.

■ *Ulrike Ostler*

## Disaggregation und Hyperkonvergenz im Auftrag softwaredefinierter zusammensetzbare Infrastruktur

# Was ist Composable Infrastructure?

**Eine dynamische Workload-Provisionierung ist das heilige Gral des Rechenzentrumsdesigns. Composable Infrastructure soll der Datacenter-Infrastruktur zur dringend benötigten Flexibilisierung verhelfen – sei es durch „echte“ Hyperkonvergenz oder durch Disaggregation.**

Der Weg zur höheren Flexibilisierung des Datencenters führt durch die Implementierung von Infrastruktur in Code (Infrastructure as code). Dies setzt voraus, dass sich die Ressourcenklassen Compute, Storage und Netzwerk separat voneinander erweitern und über ein vereinheitlichtes API zusammensetzen lassen.

Composability bildet de facto eine Voraussetzung für extreme Service-Agilität. Frost & Sullivan spricht im übertragenen Sinne von einer Sammlung von logischen Legosteinen der Infrastruktur, die sich bedarfsgerecht replizieren ließen und über vorprogrammierte Intelligenz verfügten, um sich wunschgemäß und zweckgebunden in der benötigten Ausstattung dynamisch zusammenzusetzen.

### Zustandslose Infrastruktur auf der Basis von Network-Fabric

Composable Infrastructure (zusammensetzbare Infrastruktur) abstrahiert die Ressourcen Compute, Storage und Netzwerkgewebe (Fabric) und macht sie als separate Bausteine via Code steuerbar. Dies ermöglicht die Umsetzung fortgeschrittener Kontrollfunktionen wie der automatischen Hardware-Erkennung, Selbstdiagnose und -Heilung von Betriebsstörungen. Gartner

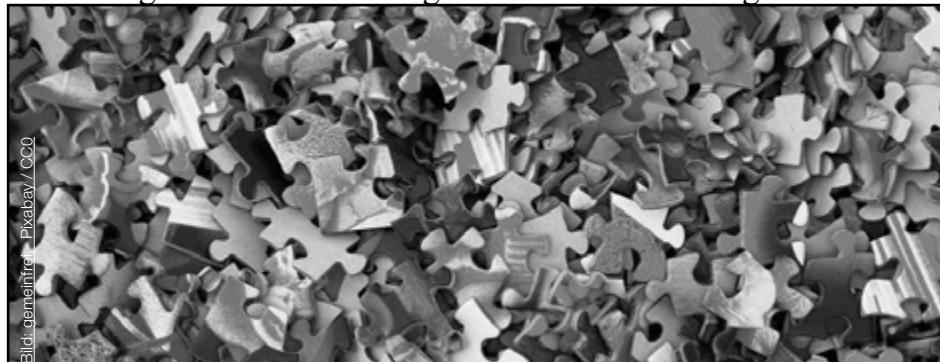


Bild: gemeinfrei / Pixabay / CC0

**Composable Infrastructure abstrahiert die Ressourcen Compute, Storage und Netzwerk in Software und macht sie via Code steuerbar**

Analysten zufolge stellt Composable Infrastructure den nächsten logischen Schritt auf dem Weg zur vollständig zustandslosen dynamischen Infrastruktur-als-ein-Service.

Das Beratungshaus IDC konnte in dem aufkeimenden Markt bereits zwei Lösungsansätze ausmachen. Die HPE-Plattform „Synergy“ ist ein Beispiel für die so genannte softwaredefiniert zusammensetzbare Infrastruktur (Software Defined Composability). Eine softwaredefiniert zusammensetzbare Infrastruktur entsteht auf der Basis eines Frameworks beziehungsweise eines Stacks von Werkzeugen. So entsprang beispielsweise HPEs Synergy-Produktlinie den Blade-Systemen des Anbieters durch die Zugabe von Features aus den Appliances „Synergy Composer“ und „Image Streamer“ (siehe auch den Bericht: „Keine HPE-Commodity-Hardware für die Hyperscaler – Was macht die Konkurrenz?“).

Der umgekehrte Lösungsansatz trägt die Bezeichnung Disaggregation; er strebt die völlige Entkopplung der Ressourcenklassen Compute, Storage und Netzwerk/Fabric in einzelne Knoten von einander, um sie wieder auf Hardware-Ebene zusammenzusetzen. In diese Kategorie fällt unter anderem „Rack Scale Design“ (kurz: RDS) von Intel für Hyperscaler.

In einem System wie HPE Synergy lassen sich die benötigten Workloads mit Hilfe softwaredefinierter Vorlagen in einem einzigen Schritt provisionieren, ohne die Verfügbarkeit laufender Dienste zu unterbrechen. Auch Änderungsvorgänge wie das Aktualisieren von Firmware führen nicht mehr zwangsweise zu Dienststörungen.

## Die Ansprache einzelner Ressourcen

Die softwaredefinierte Schicht der zusammensetzbaren Infrastruktur verwaltet autark die darunterliegende Hardware. So können die IT-Verantwortlichen ihre Aufmerksamkeit verstärkt der logischen Dimension der Infrastruktur widmen.

Ressourcen wie der Datenspeicher und das Netzwerk nutzen derart „ausgewachsene“ Protokoll-Stacks (TCP/IP, iSCSI, NFS und andere), dass sich noch eine weitere Softwareebene auf die Gesamtperformance der Infrastruktur kaum auswirkt (siehe dazu auch den Bericht: „Der Verzicht auf RAID-Arrays“). Im Gegensatz dazu sprechen Ressourcen mit Zugriffszeiten im Nanosekunden-Bereich nicht Protokoll-Stacks, sondern CPU-native Befehlssätze.

In diese Kategorie fallen Hardwarekomponenten der Speicher-

domäne wie DRAM, SCM (storage class memory oder Speicherklassenspeicher) und GPUs/FPGAs. Bei der Vervollständigung zusammensetbarer Infrastrukturen kommt gerade diesen schnellen Ressourcen eine entscheidende Bedeutung zu. Die Industrie sucht händeringend nach neuen Ansätzen.

## Das Gen-Z-Interconnect

Infrastrukturauteile der Speicherdomäne aktueller Systemarchitekturen, darunter Hardwarebeschleuniger, I/O (die Ein- und Ausgabe) und RAM, entziehen sich der Disaggregation und somit auch der Zusammensetbarkeit auf Hardware-Ebene. Jeder Versuch, die verfügbaren Ressourcen bis in die Ebene der Speichersemantik zu disaggregieren, würde sich sofort auf die Bandbreite und Latenz der betreffenden Komponenten auswirken und die Vorteile zusammensetbarer Infrastruktur schnell wieder nivellieren. (Bei Speichersemantik, memory semantics, handelt es sich um die Prozesslogik, mit deren Hilfe ein System den Zugriff auf gemeinsam genutzte Adressen des physischen Speichers oder gemeinsam genutzte Variablen durch mehrere Prozessoren oder mehrere Threads steuert.)

Denn während sich die Speicherkapazität pro CPU-Kern seit Jahren kaum geändert hat, nimmt die Bandbreite des Arbeitsspeichers pro CPU-Kern von Generation zu Generation sogar noch ab. Es wird eine völlig neue, Speicher-zentrische Architektur benötigt, und ein neues Interconnect. Eine solche Technologie entsteht unter dem Namen Gen-Z im Rahmen des Gen-Z Consortium.

Bei Gen-Z handelt es sich um einen offenen Standard für ein Interconnect auf Systemebene für den Zugriff auf Daten und Geräte unter Verwendung von Speichersemantik über eine Direktanbindung, Switches oder Netzwerkgewebe (Fabric). Gemeint ist ein neuer Bus, und zwar eine direkte, in Hardware implementierte Hochgeschwindigkeitsanbindung zwischen CPU-Kernen, SoCs, FPGAs, GPUs, Speicherpools und vernetztem Datenspeicher.

## Bandbreite und Latenz

Gen-Z hat zum Ziel, eine Bandbreite jenseits von 100 Gigabit pro Sekunde (GBps) und eine Latenz von weniger als 100 Nanosekunden zu erreichen. Es geht hierbei um die Bandbreite zwischen den Ressourcen Compute, zum Beispiel den CPU-Kernen, auf

der einen Seite und Block-basiertem Speicher auf der anderen Seite. Eben diese Anbindung stellt derzeit einen Flaschenhals im Hinblick auf die Latenz dar.

Gen-Z-Hardwarekomponenten lassen sich in die bestehende Unternehmens-IT vollständig transparent einfügen; es sind hierzu weder Änderungen am Betriebssystem noch an der Middleware erforderlich. Der Standard nutzt allerdings neuartige modulare Stecker; zum Teil sollen sie mit bestehenden Anschlüssen und bestehender Verkabelung nutzbar sein. Das Interconnect soll vorerst zwischen verschiedenen Serverknoten vermitteln und auf einen Rack skalieren können.

Unternehmen versprechen sich von der Architektur unter anderem auch neue Möglichkeiten der Leistungsoptimierung bestehender Infrastruktur, eine verbesserte Energie-Effizienz und eine höhere Branchenagilität dank der neu gewonnenen Fähigkeiten zur Datenanalyse in Echtzeit. Die Technologie ist offen und frei von Lizenzgebühren — eine Kriegserklärung der IT-Branche an Intel und seine proprietären Interconnects. Zu den Mitgliedern des Gen-Z Consortiums zählen unter anderem AMD, ARM, Broadcom, Cray, Dell EMC, HPE, Huawei, IBM, Lenovo, VMware und Samsung.

## Disaggregierte Integration?

Hyperkonvergente Infrastrukturen basieren auf einem gegenteiligen Ansatz, nämlich auf der Ressourcen-Aggregation. Integrierte Systeme würden nicht automatisch die beste Lösung für beliebige Workloads darstellen, bestätigt das Analystenhaus Gartner in einem Bericht. Hyperkonvergenten Systemen stünde ein großes Redesign bevor.

Insbesondere bei Deployments jenseits von 16 bis 20 Knoten stößt Hyperkonvergenz in der aktuellen Implementierung schnell an ihre Grenzen; bei latenzkritischen Arbeitslasten versagt sie kläglich. Zu den Problemfeldern der Architektur zählen außerdem die niedrige de-facto nutzbare Speicherkapazität (viel „Verschnitt“ als ein Nebeneffekt der Replikation, anbieterspezifisch), die beschränkte Belastbarkeit und die eingebauten Performance-Bremsen, allen voran das Netzwerk.

Die Schwächen aktueller Implementierungen fallen insbesondere bei KI-lastigen Workloads, Big Data-Anwendungen auf der Basis von HANA, Hadoop oder Cloudera und anderen nicht-linearen, speicherlatenzsensiblen, transaktionalen Arbeitslasten ins Ge-

wicht. Die Kosten zusätzlicher Leistung steigern bei HCI üblicherweise doppelt so schnell wie die Performance.

Anbieter wie Stratoscale versuchen bereits, diese Probleme durch eine Hardware-agnostische, „echte“ Hyperkonvergenz und Disaggregation auszumerzen. (Die Stratoscale-Plattform setzt wie Nutanix mit „AHV“, auf dem KVM-Hypervisor auf und unterstützt eine Migration von vSphere oder Hyper-V wie auch von anderen KVM-Implementierungen.

Gartner identifiziert drei Entwicklungsphasen integrierter Systeme.

- Die bereits abgeschlossene erste Phase fand zwischen 2005 und 2015 statt; sie umfasst den Höhepunkt von Blade-Systemen.
- Die zweite Phase begann etwa 2010 und dürfte voraussichtlich bis 2020 anhalten; sie markiert die Ankunft konvergenter Infrastrukturen und das Aufkommen von HCIS für spezielle Anwendungsfälle.
- Die dritte Phase, die ja auch bereits 2016 anfing und sich bis etwa 2025 ausstrecken könnte, repräsentiert aus Sicht des Analystenhauses die kontinuierliche Anwendung und Bereitstellung von Microservices auf HCIS-Plattformen wie der von Stratoscale. Die dritte Phase der integrierten Systeme werde demnach „dynamische, zusammensetzbare und Fabric-basierte Infrastrukturen bereitstellen“, indem sie auch modulare und disaggregierte Hardwarebausteine böte, die eine kontinuierliche Anwendungsbereitstellung und kontinuierliche wirtschaftliche Optimierung ermöglichen sollen, schrieb Gartner in einem Bericht.

Ein nicht zu unterschätzender Vorteil zusammensetbarer Infrastrukturen gegenüber herkömmlichen Architekturen besteht in der Fähigkeit, sowohl Bare-Metal-Deployments als auch virtualisierte und containerisierte Umgebungen auf einer einzigen Plattform zu implementieren. Die nächste Generation der Hyperkonvergenz werde letztendlich die zugrundeliegende Infrastruktur verschwinden lassen; die Infrastruktur würde zu einem anpassungsfähigen Werkzeug morphen, welches unter der Kontrolle intelligenter Software automatisiert würde, um „ITaaS“-Implementierungen (IT-as-a-Service) zu ermöglichen. „HCIS ist kein Ziel, sondern eine evolutionäre Reise“, urteilt Andrew Butler, Vice President und distinguiertes Analyst bei Gartner.

Gartner prognostiziert, dass 20 Prozent geschäftskritischer Anwendungen, die heute (2018) auf Three-tier-Infrastruktur bereitgestellt werden, bis zum Jahre 2020 auf hyperkonvergente Sys-

teme migrieren würden. Anfang 2019 rechnen die Analysten mit einem Marktvolumen von rund 5 Milliarden Dollar.

Richard Fichera, VP und Principal Analyst Für DevOps von Forrester Research, urteilt: „Die Aufnahme zusammensetzbarer Infrastrukturtechnologie in den aktuellen Mix aus Virtualisierungsabstraktionen und Automatisierungswerkzeugen (...) bringt das vollständige Potenzial eines SDDCs (...) näher an die Realität“. Geoff Hollingworth, Head of Product Marketing Cloud Systems bei Ericsson, schlägt in dieselbe Kerbe. „Alles muss programmierbar sein (...) damit es sich automatisieren lässt“, so Hollingworth.

■ *Filipe Pereira*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



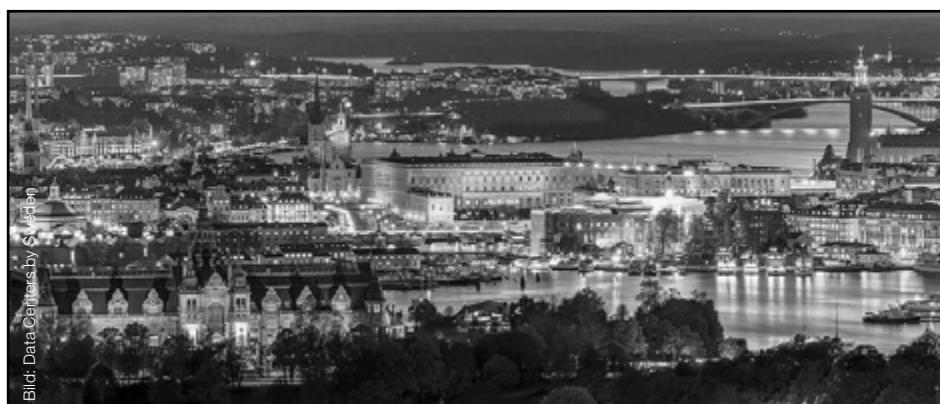
**Es könnte besser laufen - bei uns!**

# Datacenter in Schweden und in Deutschland - best Practices versus Ignoranz

**Zugegeben: Es ist ein wenig unfair, eine kleine Veranstaltung in Frankfurt mit einer großen in Stockholm zu vergleichen, obwohl es in beiden um die Art und Weise ging, Rechenzentren zu betreiben – möglichst nachhaltig. Allerdings: In Schweden setzen die Verantwortlichen, von der Politik über Behörden bis zu Energielieferanten und den Rechenzentrumsbetreibern, die Zukunft bereits um. Das spricht sich herum und zieht Publikum an, während in Deutschland das Nachdenken erst einsetzt.**

Im März dieses Jahres hatte Data Center Dynamics (DCD) zur Rechenzentrumskonferenz „Energy Smart“, die vom schwedischen Staatssekretär im Ministerium für Unternehmen und Innovation, Niklas Johansson eröffnet wurde, geladen und der Andrang von mehr als 350 Besuchern überwältigte uns geradezu, sagt Thomas Sokolnicki, Senior Investment Advisor Business Sweden, Data Centers by Sweden. Doch nicht nur die Politik unterstrich die Bedeutung auch Vertreter von IBM, Stockholm Exergi (früher: Fortum Värme), NTT, Google, Facebook, Microsoft, OX Wind, der Stadt Stockholm, Digiplex, Hydro66, e-Shelter, The Green Grid, Interxion, der Europäischen Kommission, Baidu, Statkraft, GreenIT Amsterdam, Uptime Institute, Aligned Energy, Base-layer verliehen der Veranstaltung Bedeutung.

Bereits im Februar dieses Jahres hatte das Netzwerk Energieeffiziente Rechenzentren in Frankfurt am Main zum Experten-Workshop „Rechenzentrum von Morgen“ eingeladen.



**Beispiele aus Stockholm zeigen, wie Datacenter-Ansiedlung attraktiv für Investoren und für die Bevölkerung wird.**

Thema in Frankfurt sowie in Stockholm waren unter anderem die Potenziale zur Nutzung von Abwärme aus Rechenzentren und Cooling as a Service. Dabei war etwa das Energierreferat der Stadt Frankfurt, in persona Paul Fay, der lokale Energie-Lieferant Mainova, vertreten durch Ingo Franke, und Jan-Oliver Schmitt von Hessen Trade & Invest, Initiative Digitales Hessen.

## Peinlich

Das wohl Positivste, das sich berichten lässt: Die Strategie tauber Ohren scheint durchlässig zu werden, wenn es um die Nutzung der Abwärme beziehungsweise, wenn es um das Kühlen von Rechenzentren geht. Peinlich ist allerdings der Wissensstand, ist doch Frankfurt einer der größten Rechenzentrumsstandorte weltweit.

Sieht nicht kompliziert aus: Solche Verbindungsstücke benötigten Rechenzentren, um Wärme abzugeben und Kälte zu beziehen. Nicht einbezogen in diese Darstellung ist die Tatsache, dass die Abwärme bei 30 Grad oder leicht darüber liegt. Für die Einspeisung braucht es aber mindestens 50 Grad.

Sieht nicht kompliziert aus: Solche Verbindungsstücke benötigten Rechenzentren, um Wärme abzugeben und Kälte zu beziehen. Nicht einbezogen in diese Darstellung ist die Tatsache, dass die Abwärme bei 30 Grad oder leicht darüber liegt. Für die Einspeisung braucht es aber mindestens 50 Grad. Unklarheit herrscht auch, wenn es um die Temperaturen für die Einspeisung in ein Fernwärmennetz geht. In der Regel geben die Rechenzentren 30 Grad für die Abwärme an; Fernwärmennetze benötigen aber 60,70 Grad. Damit müsste die Temperatur angehoben werden. Dafür gibt es entweder Hot-Fluid-Systeme, die die Wärme direkt aus dem Server, in dem die Recheneinheiten 70, 80 Grad heiß werden, ziehen – Thomas-Krenn AG und Cloud & Heat nutzen das – oder Wärmepumpen.

Die wiederum brauchen Strom. Stehen sie in den Datacenter, müssen die Betreiber für das Aufheizen bezahlen, mitsamt EEG-Umlage und allen sonstigen Gebühren. Zudem gibt es etwa bei Mainova keine Idee, in welcher Höhe und ob man überhaupt für diese Wärme zahlen sollte.

Bei dem lokalen Energieversorger, der die meisten Rechenzentren in Frankfurt am Main versorgt, träumt man auch lieber von einem zusätzlichen Geschäft: Die Kühlung aus dem Netz, für das die Datacenter-Betreiber natürlich zahlen sollen. Doch damit ist

mit konkreten Überlegungen auch schon Schluss. Bei dem Workshop konnte der Mainova-Vertreter Franke weder beantworten, welche Temperaturen er zu welchen Konditionen liefern könne, ab wann das möglich wäre und wie Temperaturschwankungen auszugleichen wären: Kopfschütteln bei der anwesenden potenziellen Kundschaft bei dem demonstrierten Desinteresse.

Alles nicht so schlimm? Die Initiative Digitales Hessen bekundete jedenfalls, sämtliche Bestrebungen zu unterstützen. Das klingt gut. Allerdings denkt man hier in Zeiträumen von 20, ja, 30 Jahren. Und wer alles beteiligt sein soll ... Da gibt es hier einen Professor, dort ein Institut, fernerhin diese Organisation .... Koordination und Zielstrebigkeit ist etwas anderes.

Es bleibt also auch im digitalen Hessen, und das wähnt sich ganz vorne in punkto Nachhaltigkeit. Rechenzentren und digitaler Infrastruktur wie Dauermahner und Energie-Experte Staffan Revemann sagt: „Deutschland verschwendet Energie.“

### **Was machen die Schweden besser?**

Die Frage also, warum eine Datacenter-Konferenz in Stockholm so viele Investoren, Rechenzentrumsbetreiber und deren potenzielle Kunden anzieht, erklärt sich somit fast von selbst. Doch was läuft in Schweden besser als hierzulande? Alles. Alles? Alles.

Das beginnt bei einem gesellschaftlichen Konsens. Nachhaltigkeit und Energiesparen scheinen quasi in der DNA der Schweden verankert - ein Widerspruch zwischen Ökonomie und Umweltbewusstsein existiert nicht. Auch die Chancen in der Digitalisierung zu sehen ist gesetzt. Da es Digitalisierung ohne Rechenzentren und Datennetze nicht gibt, wird für beides gesorgt.

Das heißt nicht, dass nicht nachgerechnet wird. „Ich bin froh, schon ganz früh angestoßen zu haben, welche ökonomischen Folgen die Facebook-Ansiedlung in Nord-Schweden haben würde,“ sagt Sokolnicki von Data Centers by Sweden. (siehe: Der wirtschaftliche Nutzen der Facebook- (und AWS-) Ansiedlung in Schweden, Facebook-Datacenter in Nord-Schweden verdoppelt sich“. Noch nicht eingerechnet ist die Start-up-Kultur, die Rechenzentren nach sich ziehen.

### **Alle ziehen an einem Strang**

Heute scheinen Politik und Gesetzgebung, Behörden und Industrie an einem Strang zu ziehen; im Herbst 2016 beschloss die Regierung, die Stromsteuer für Rechenzentren drastisch zu senken.

Sokolnicki: „Mein Konzept war es, eine mit Business Sweden und darin verankert Data Center by Sweden eine nationale Plattform zu schaffen, auf der die verschiedenen schwedischen Regionen und ihre Interessensvertretungen zusammenarbeiten können, etwa Nord Pole, Invest in Norrbotten, Invest in Gävleborg, Skelleftea, The Power Region und Big Green Data. Haben sich die Firmen, die sie anlocken wollten, für Schweden entschieden, entsteht eine Konkurrenzsituation. Das ist wie bei einem Nationalteam im Fußball; die Spieler agieren international als eine Mannschaft, aber außerhalb spielt jeder bei seinem eigenen Verein.“

## Es begann 2008

Persönlich rutschte er ins Datacenter-Geschäft, als sich 2008 IBM mit einem eigenen Rechenzentrum in Stockholm niederließ. Er bestätigt zudem die positive Zustimmung in der Bevölkerung, was die weitere Ansiedlung der Energie-intensiven Branche angeht, obwohl die Senkung der Steuern für diese Branche für Diskussionen gesorgt haben dürfte.

„Nein, es gab keine großen Auseinandersetzungen“, widerspricht er. Die Leute seien einverstanden mit der Maßnahme, die die Attraktivität für stromintensive Industrien wie die Datacenter-Branche erhöhen, auch wenn es sich um Co-Locator, Hyperscaler wie AWS und Facebook handle. Die Ausnahme bilde lediglich das Krypto-Mining. „Dazu gibt es offene Fragen“, sagt Sokolnicki. Die Standards seien niedrig, der Energieverbrauch immens, die Geschäftsmodelle unklar und der Nutzen für die Bevölkerung dahingestellt.

2008 war auch das Gründungsjahr für Data Center by Sweden. 2009 war Sokolnicki das erste Mal auf Werbetour in den USA. Schon damals entpuppte sich die Verwendung der Rechenzentrumshitze für die Fernwärmenetze als Türöffner, erinnert sich der Manager. Heute lassen Firmen aus 30 Ländern in Schweden rechnen. Die aktuelle Wachstumsrate beträgt 14 Prozent.

## Die Data Parks

Zu den Zukunftsbranchen zählt Sokolnicki neben der Datacenter-Branche auch IT im Allgemeinen, Life Science, Batterieherstellung und Automotive. „Wir stellen sicher, dass die Firmen, die sich hier ansiedeln wollen oder Partner suchen, alles an Informationen und Kontakten bekommen, was sie benötigen.“

Die Data Parks in Stockholm allerdings bieten weit mehr als Information und Networking. Hier ist die Infrastruktur bereits vorbereitet: Es steht genügend Strom aus verschiedenen Richtungen mit sehr hoher Verlässlichkeit zur Verfügung. Dazu kommt ein bemerkenswertes Datennetz: Stockholm besitzt das weltweit größte Dark Fiber Netzwerk, das 30 Millisekunden Roundtrip ermöglicht und ist praktisch der Hub für die Nordics.

Der Strompreis ist der niedrigste in der EU. Die Kilowattstunde aus erneuerbarer Energie kostet 0,04 Euro. (Nur 10 Prozent des Stroms stammt in Schweden aus anderen Stromquellen.)

Das Fernwärmennetz umfasst in Stockholm rund 2.800 Kilometer. 10.000 Haushalte sind bereits angeschlossen, zu 95 Prozent im Kern von Stockholm. Das entspricht rund 12 Terawattstunden pro Jahr, 1 Prozent der Heizungen in Stockholm; im Jahr 2035 sollen es 10 Prozent sein. Das heißt aber auch: Ein Rechenzentrum mit einer Last von 10 Megawatt kann rund 20.000 Wohnungen heizen.

Es gibt bereits 30 Rechenzentren, die ihre Abwärme einspeisen. Zu den Referenzen zählen Ericsson, H&M, Interxion, Bahnhof und Digiplex. 100 Millionen Euro gespart

Für die Einspeisung gibt es verschiedene Modelle, erläutert Erik Rylander, Head of Stockholm Data Parks. Zum einen können die Rechenzentren mit eigenen Wärmepumpen selbst das erforderliche Temperaturniveau von mindestens 68 Grad erreichen oder der Fernnetzbetreiber Stockholm Exergi unterstützt das.

Das Einspeisen selbst rechnet sich (siehe: Abbildung). In dem Beispiel, das Ryklander vorlegt, rechnet sich die Anfangsinvestition in die Wärmepumpe für ein 10 Megawatt-Rechenzentrum in fünf Jahren. Bei einem Festpreis, den der lokale Stromversorger für zehn Jahre bietet, komme ein solches Rechenzentrum in diesem Zeitraum auf mehr als 100 Millionen Euro an Einsparungen.

Doch auch für den Abnehmer der Abwärme rechnet sich das Geschäftsmodell. „Der Preis, den wir für die Einspeisung bezahlen, liegt unter den Kosten, die wir für das Generieren der Fernwärme hätten“ lässt Stockholm Exergi wissen.

## Cooling as a Service

Noch mehr sparen lässt sich, wenn Rechenzentren das Angebot einer Kühlung als Service annehmen. Das dafür zur Verfügung stehende Netz umfasst derzeit rund 300 Kilometer. Interxion und

Ericsson nutzen das geschlossene System. Diese Rechenzentren haben eigene Kühlanlagen nur noch als Backup.

„Wir wollen dieses Angebot noch erheblich ausbauen. Die Idee ist, dass Rechenzentren, die einen geringeren Bedarf als 10 Megawatt haben, die Kühlung für einen Preis beziehen, der unterhalb der Eigenversorgung liegt. Rechenzentren, die größer sind und mehr Wärme abgeben, bekommen 10 Megawatt für lau“, so Ryklander.

Die Nutzung der Abwärme aus den Rechenzentren und die Kälte aus dem Netz führen zu einer positiven Klimabilanz. Denn die CO2-Emmission reduziert sich deutlich (siehe: Abbildung), da die einmal erzeugte Wärme weiter verwendet wird und dieser Anteil nicht eigens generiert werden muss.

Um das Konzept auszubauen und eventuell auf weitere Industrien anzuwenden, muss die Stadtentwicklung in die Betrachtungen einbezogen werden. Stockholm ist eine Boom-Town. Stockholm rangiert auf Platz 10 unter den Städten mit den meisten Firmensitzen globaler Firmen - Nr. 1 ist Tokyo, es folgen London, Hong Kong, New York, Peking, Seoul, Paris, Taipeh und Houston. Außerdem ist Stockholm die am schnellsten wachsende Stadt in Europa:

Die Bevölkerung wächst jährlich um 20 Prozent, der CO2-Ausstoß aber sinkt zugleich um 25 Prozent. Stockholm hat eine der höchsten Gründungsraten mit 16,1 neuen Unternehmen je 1.000 Einwohner und viele Gründungen sind erfolgreich. Was milliardenschwere Startups angeht, rangiert Stockholm weltweit auf Platz fünf, pro Kopf auf Rang zwei hinter dem Silicon Valley, zählen doch Firmen wie Klarna, Skype und Spotify dazu.

■ *Ulrike Ostler*

## Künstliche Intelligenz in Geschäftsanwendungen

# KI: Treibstoff für die intelligente Zukunft der Geschäftswelt

**Beim Gedanken an die Nutzung künstlicher Intelligenz (KI) im Geschäftsleben werden einige mit Schrecken an den Supercomputer HAL aus dem Film „2001: Odyssee im Weltraum“ denken. Der Computer der Serie 9000 übernimmt hier die Kontrolle über das auf dem Weg zum Jupiter befindliche Raumschiff und tötet die gesamte Besatzung. In der Realität werden die Auswirkungen allerdings nicht so dramatisch sein – zumindest hoffen wir das. Vielmehr besitzt KI die Fähigkeit, jenen Unternehmen, die mutig genug sind, einen echten Vorteil zu verschaffen.**

Wie viele andere neue Technologien auch, wird KI nicht allein den größten Konzernen zur Verfügung stehen. Tatsächlich ist es eher so, dass die meisten aufstrebenden Unternehmen in vielen Fällen einen Vorteil gegenüber den bereits fester etablierten Organisationen haben, wenn es um die nächste Welle intelligenter Automatisierung geht. Sie sind weniger durch entstandene Infrastrukturkosten eingeschränkt, reaktionsschneller und besitzen mit größerer Wahrscheinlichkeit die Fähigkeit, einige der Möglichkeiten, mit denen KI die moderne Geschäftswelt umgestalten wird, für sich zu nutzen.

Wie Rupert Murdoch schon sagte, geht es nicht mehr darum, dass die Großen die Kleinen schlagen, sondern die Schnellen besiegen die Langsamen. Dank neuer intelligenter Technologien können moderne Unternehmen in die Zukunft blicken und florieren – gestützt auf Entscheidungen, die mithilfe zweckdienlicher, zeitnahe, präziser, vorbeugender und belastbarer Echtzeit-Informationen gefällt werden.

Das Tempo in der Geschäftswelt, die Geschwindigkeit der Verän-

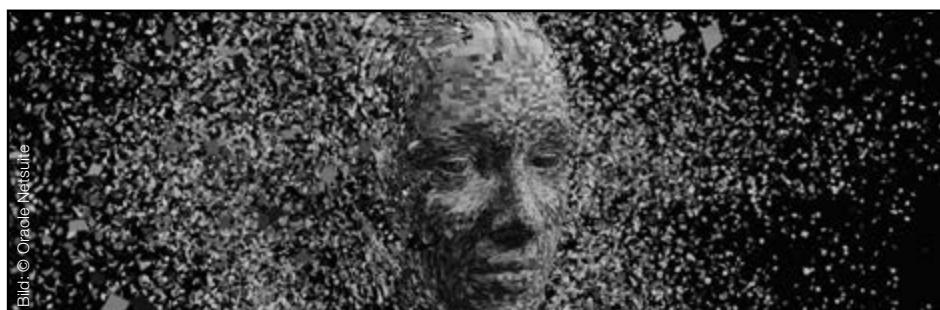


Bild © Oracle NetSuite

**KI ist eine mächtige Schlüsseltechnologie – allerdings nur, wenn sie auf dem richtigen Fundament gründet.**

derungen und das Wachstum der Datenmengen vollziehen sich heute zunehmend exponentiell. Informationen liefern die nötigen Erkenntnisse für ein schnelles Voranschreiten. Angesichts der enorm wachsenden Datenmengen aber stellt sich die Frage, wie Unternehmen all die Informationen verarbeiten können, um Erkenntnisse für eine schnellere Evolution und Innovation zu gewinnen, ohne dass sie durch die Suche nach Erkenntnissen ausgebremst werden.

Denken Sie nur an einige der wichtigsten Schlagwörter von heute: KI, Machine Learning, Bots und weitere intelligente Arten der Automation. Schon das Spektrum der Dienste, die damit geboten werden sollen, ist verwirrend – von einer vereinfachten Datenverarbeitung über das Präsentieren von Entscheidungen für die Benutzer bis hin zum „Training“ einer Maschine nach diesen Entscheidungen zu handeln. Wie kann ein Unternehmen, das sich seit Jahren auf die Hilfe seiner Geschäfts- und Finanz-Software gestützt hat, um Prozesse zu automatisieren und retrospektive geschäftliche Intelligenz bereitzustellen, den Schritt in die intelligente Welt vollziehen?

KI ist eine mächtige Schlüsseltechnologie – allerdings nur, wenn sie auf dem richtigen Fundament gründet. Unternehmen, die über eine einheitliche Plattform für ihre Geschäftsprozesse verfügen, sind bestens aufgestellt für die Nutzung von KI. Ein Unternehmen mit durchgängigen Applikationen in der Cloud wird nicht durch Datensätze behindert, die auf mehrere Datentöpfe verteilt, in Tabellenblättern verborgen sind und sich, wenn überhaupt, dann nur schwierig integrieren lassen.

Unternehmen mit echten Cloud-Lösungen sind außerdem Bestandteil einer großen Gemeinschaft von Unternehmen. Sie können geschäftliche Muster und Praktiken nutzen, die über ihren eigenen Erfahrungsschatz hinausgehen. Abgesehen davon haben sie den zusätzlichen Vorteil, dass ihnen stets die aktuellste Version vorliegt, sodass dem Zugriff auf die neueste Technologie keinerlei Hindernisse im Weg stehen. Auf kosteneffiziente Weise helfen vertikale und branchenspezifische Applikationen den Unternehmen dabei, einen geschäftlichen Nutzen aus der KI zu ziehen. Insbesondere sind es drei Arten, auf die diese Unternehmen künftig von künstlicher Intelligenz profitieren können:

## 1. Intelligente Analysen

Anstelle der rückblickenden Perspektive, die mit Business Intelligence möglich ist, analysiert KI große Datenmengen, um Ent-

scheidungen zu empfehlen oder – wie im Fall des Machine Learning – auf der Basis von Daten tatsächlich zu handeln.

Mithilfe von KI können sowohl Ausreißer als auch Trends, die sich auf das Geschäft auswirken, in Echtzeit identifiziert werden. Zum Beispiel wird eine voraussichtliche Verzögerung bei der Bestellung eines billigen Rohstoffs die rechtzeitige Auslieferung hochwertiger Aufträge einen Monat später beeinflussen und damit entsprechende Folgewirkungen nach sich ziehen. Die daraus resultierenden Probleme mit der Kundenzufriedenheit werden wiederum zu verzögerten Zahlungen, Außenständen, höheren Rabatten und weniger Folgeaufträgen führen und damit Auswirkungen auf die Einnahmen in sechs Monaten haben. Wenn sie sofort über das Problem informiert werden, haben die Einkäufer die Möglichkeit, einen passenden Ersatz zu finden, um auf diese Weise die Auswirkungen auf die künftigen Einnahmen zu umgehen. Mithilfe dieser Analyse können Anwender Konsequenzen vorhersagen und früher die richtige Entscheidung fällen, um teure Fehler zu vermeiden.

## 2. Intelligente Interaktion

Anwendungen können Hunderte rollenabhängiger Dashboards besitzen, die fest vorgegeben sind. Sie basieren auf jahrelanger Entwicklung und Implementierung in mehreren Branchen und bieten Funktionen, damit die Mitarbeiter ihre Aufgaben effektiv erledigen können. In der Zukunft aber wird intelligente Interaktion für die dynamische Konstruktion von Dashboards sorgen. Dabei wird die Einschätzung dessen zugrunde gelegt, was der jeweilige Anwender je nach seinen aktuellen Aktivitäten und dem Verhalten Tausender ähnlicher Rollen vermutlich sehen muss. Wir erleben derzeit außerdem das Aufkommen intelligenter sprachaktivierter Lösungen. Die Intelligenz nimmt hier unstrukturierte Anweisungen entgegen, um zu ermitteln, was der Benutzer meint, und um anschließend zu gewährleisten, dass die richtigen Informationen ausgegeben werden.

## 3. Intelligente Automation

Unternehmens-Lösungen können mit sehr leistungsstarken Workflow Engines ausgestattet sein, die Unternehmen das Verändern und Automatisieren von geschäftlichen Abläufen ermöglichen, ohne dass sie dafür Code schreiben müssen. Die Engines arbeiten regelbasiert: Wenn A = B, dann C oder sonst D. Gestützt

auf dem Erlernen geschäftlicher Muster und Verhaltensweisen, werden neue intelligente Systeme Prozesse erlernen, vorschlagen und automatisieren. Zum Beispiel kann die Maschine je nach Bezahl- und Bestellverhalten Kunden unterschiedlich behandeln.

Intelligente Analyse, Interaktion und Automation hilft Finanzabteilungen und CFOs dabei, effektiver zu werden und den allgemeinen Erfolg eines Unternehmens voranzubringen. Die folgenden vier potenziellen Anwendungsfälle zeigen, wie CFOs vom wahren Nutzen der Intelligenz profitieren können, um künftige Trends im Verhalten von Kunden oder Mitarbeitern zu ermitteln.

1. vorhersagende Analytik: Für CFOs dreht sich der wahre Nutzen der vorhersagenden Analytik um das Ermitteln bevorstehender Trends im Kundenverhalten. Die erste große Chance besteht in der Identifikation von Kunden, bei denen das Risiko besteht, dass der Auftrag verlorengeht – mit entsprechenden Auswirkungen auf den Umsatz etc. Das Zahlungsverhalten, die bei Vertriebsinteraktionen erhobenen Daten etc. lassen sich bündeln, um das künftige Verhalten zu verstehen. Als zweites resultiert hieraus die Chance zur Identifikation von Produkt- oder Angebotslücken, die mit traditioneller Denkweise oder Analytik übersehen würden.
2. vorhersagende Buchhaltung: Bei der vorhersagenden Buchhaltung geht es darum, das traditionelle, historische Rechnungslegungs-Konzept zugunsten eines verbesserten Management-Buchhaltungs- und Analysemodells zu verlassen. Man möchte den Entscheidern hiermit Echtzeit-Erkenntnisse zur Verfügung zu stellen, die sie für Entscheidungen nutzen können. Dies würde dem System die Überwachung von Transaktionen ermöglichen, denn sie sind für die Suche nach bestimmten Mustern konzipiert, die wichtig für die Lösung geschäftlicher Probleme sein können. Außerdem können sie Journaleinträge vorschlagen.
3. vorbeugendes Auditing: Das vorbeugende Auditing geht über das traditionelle, kontinuierliche Auditing-Konzept hinaus. Bei der fortlaufenden Überprüfung von Daten versteht das prädiktive Auditing mit der Zeit, welchen Weg die Daten nehmen, um so festzustellen, wenn etwas außerhalb der Norm liegt. Theoretisch könnte sich hiermit ein potenzieller Betrug

schon sehr früh im Anfangsstadium aufdecken lassen, und nicht erst dann, wenn das Ausmaß schon so groß ist, dass es auch mit traditionellen Mitteln erkannt wird. Derartige Fähigkeiten könnten die Kosten und die Dauer von Audits deutlich reduzieren.

4. vorbeugendes Risikomanagement: Die Funktionsweise des vorbeugenden Risikomanagements besteht darin, in umfangreichen Datensätzen Muster zu erkennen, die auf Betrug oder andere Probleme hindeuten können. Viele Finanz-Institutionen streben zunehmend den Einsatz von Machine-Learning-Lösungen an, um die regelmäßig eintreffenden Berichtsdaten sowie unstrukturierte Informationen zu verwalten und zu untersuchen. Dies kann ganz erheblich die analytischen Fähigkeiten im Risikomanagement- und Compliance-Bereich verbessern (z. B. Geldwäsche, Modellierung von Kreditrisiken und Regelwerke). Außerdem werden strukturierte und unstrukturierte Daten kombiniert, die aus sozialen Medien und der Überwachung des Internets sowie von Emails, Textverarbeitungs-Dokumenten, Videos, Fotos und Audiodateien stammen, um böswillige Aktivitäten und Trends zu identifizieren und Risiken beispielsweise für Betrug oder Cyber-Attacken abzuwehren.

Im Gegensatz zu dem Chaos, das HAL in „2001: Odyssee im Weltraum“ stiftete, werden sich Unternehmen, die auf intelligente Technologien setzen, beste Voraussetzungen dafür schaffen, ihre Mitbewerber hinter sich zu lassen, ihr Prozessineffizienzen zu senken und ihren Umsatz und Gewinn zu steigern.

■ *Hartmut Hamann*

## Kühler Dampf - aus dem Bestseller „IT-Räume und Rechenzentren planen und betreiben“

# Strittig: Welche Luftfeuchte sorgt für prima Klima?

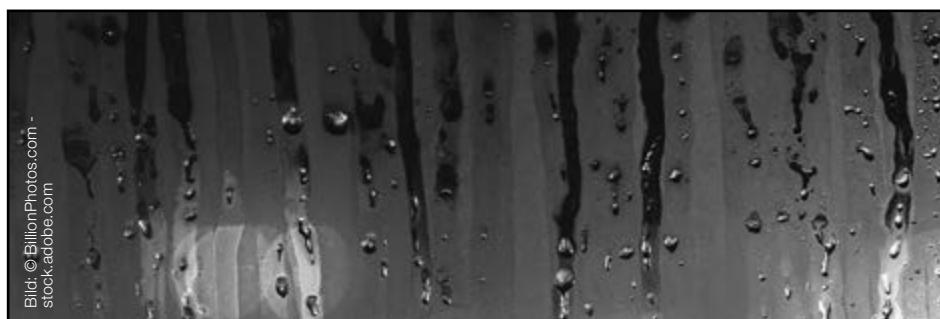
**Viel haben wir schon berichtet über die Klimatisierung von Rechenzentren, wobei das wesentliche zwangsläufig übersehen wurde: Die Luft. Sie beziehungsweise ihre Feuchtigkeit ist laut Bernd Dürr einer der umstrittensten Punkte bei der Planung und dem Betrieb von Rechenzentren.**

Dürr ist Autor des Bestsellers „IT-Räume und Rechenzentren planen und betreiben“, aus dem wir alle paar Wochen ausgewählte Kapitel in gekürzter Form präsentieren. Das diesmal behandelte Thema „Luftfeuchtigkeit“ ist Teil des Bereiches „Klimatisierung“, an dem Peter Döbert, , Rodgau, kräftig mitgewirkt hat.

Die Zeiten ändern sich. Vor 15 Jahren wurde laut Bernd Dürr ein so genanntes Präzisionsklima für die Rechnerräume propagiert. Aus dieser Zeit stamme dann auch die Bezeichnungen einiger Hersteller für Ihre Produkte als „Präzisionsklimagerät“. Heute dagegen liege der Fokus auf der Energie-Effizienz von Rechenzentren, aktuelle Produkte würden eher als „Energysaver“ oder ähnliches bezeichnet. „Auch für die Regulierung der Luftfeuchte gilt es, Energie-optimierte Systeme zu verwenden“, so der Autor.

### Parameter für die Luftfeuchtigkeit

Als Folge dieser Entwicklung seien auch die oberen und unteren Grenzwerte der DIN EN 50600 für die zulässig relative Raumfeuchte deutlich auseinandergewandert. Die DIN EN 50600-2-3 fordere nun für alle Räume mit elektrostatisch empfindlichen Geräten, dass die relative Luftfeuchte genau gemäß den Anweisungen der Hersteller der Geräte eingehalten werden muss. Nur wo



**Bei der Planung eines Rechenzentrums muss zunächst eine Feuchtebegrenzung festgelegt werden.**

keine Informationen vorhanden sind, muss ein minimaler Taupunkt von 5,5 Grad eingehalten werden.

Bei der Planung eines Rechenzentrums müsse zunächst eine Rechenzentrumsklassifizierung oder Feuchtebegrenzung festgelegt werden. Die amerikanischen Klassifizierungen lassen hier laut Dürr einen wesentlich größeren und somit effizienteren Spielraum zu als die deutschen Regelungen, die eine Mindestfeuchte von 39 Volumenprozent relativ Feuchte (Vol.-% r.F.) verlangen.

Zudem und grundsätzlich seien die vorgeschriebenen Umgebungsbedingungen der Hardwarehersteller maßgeblich. „Oftmals geben diese Grenzwerte vor, an die man sich tunlichst halten sollte, sofern man die Gewährleistungen nicht verlieren möchte“, heißt es von Dürr.

Das Temperaturniveau im Rechenzentrum ist ausschlaggebend, um den oberen Grenzwert festzulegen. Bei einer Zulufttemperatur vor den Servern von 24 Grad sollte sich für die meisten Anwendungen die Luftfeuchte zwischen 30 und 70 Vol.-% r.F. eingrenzen. Zu feuchte Luft könnte durch Taubildung Schäden nach sich ziehen.

## Regulierung der Luftfeuchtigkeit

Die benötigte Befeuchtungsleistung muss - genau wie die Kühlleistung - für jedes Projekt individuell berechnet werden. Dabei gibt es unterschiedliche Möglichkeiten, die Luftfeuchte zu regeln:

- Wird mit Umluftkühlgeräten gekühlt, wird ein Teil der Geräte mit einem Dampfbefeuchter ausgestattet. Dies ist in der Regel ausreichend, um die Luftfeuchte entsprechend der Grenzen zu halten. Jedoch gibt es auch eigenständige Luftbefeuchter die nicht in anderen Geräten verbaut sind, sondern als separate Einheit mit eigener Steuerung im Rechenzentrum oder IT-Raum montiert werden.
- Dampfbefeuchter sind die einfachste Art, um Luftfeuchte zu erzeugen. Er ist in etwa aufgebaut wie ein Heißwasserkocher aus dem Haushalt. Wasser wird durch eine Elektrode oder einen Widerstand verkocht, bis es den gasförmigen Zustand erreicht. Der erzeugte Wasserdampf wird dann über Ventilatoren dem Raum zugeführt, wo er sich wieder abkühlt. Diese Technik sei zwar in der Investition sowohl die Günstigste als auch die Einfachste, von den Betriebskosten her gesehen komme sie aber am teuersten.

- Wesentlich effizienter als isotherme Dampfsysteme seien so genannte adiabate Systeme, bei denen das Wasser als Nebel in die Luft gebracht wird. Hier führt Dürr aus: „Diese Vernebelung wird durch feine Zerstäubung, zum Beispiel mittels Ultraschallwellen, erreicht.“

Mittels eines piezoelektrischen Wandlers wird die elektronisch hergestellte Ultraschallfrequenz in mechanische Energie umgewandelt. Die Oberfläche des Schwingungswandlers schwingt in derart großer Geschwindigkeit, dass das Wasser der Bewegung aufgrund der Massenträgheit nicht mehr folgen kann und somit ein momentanes Vakuum und eine Kompression entstehen. Die Blasen, die hierbei entstehen, treffen mit großer Kraft aufeinander. Dieser Vorgang wird Kavitation genannt.“

### Kaltnebel-Befeuchtung

Durch die Fokussierung der Schall-Leistung dicht unter der Wasseroberfläche würden zudem Kreuzwellen erzeugt, in deren Kreuzungspunkte sich kleinste Wassertropfen lösen und einen Nebel erzeugen. Dieser werde vom Luftstrom aufgenommen und in die Fläche verteilt.

Dürr: „Die Wasserpartikel sind im Durchschnitt nur 0,001 Millimeter groß. Die Herstellung des Nebels durch adiabate Befeuchtung wird umgangssprachlich auch Kaltnebel-Befeuchtung genannt. Die Befeuchtungsanlagen mittels Ultraschallwellen sind in der Anschaffung zwar teurer als die Elektrodenbefeuchter, allerdings spart man im Betrieb bis zu 90 Prozent der elektrischen Energie.“ Zudem kühle man die Luft durch den Befeuchtungsprozess, wodurch dann auch weniger mechanische Kälteleistung erzeugt werden müsse.

Als weitere Vorteile nennt der Experte, dass die Luftbefeuchtung sofort zur Verfügung steht und das Wasser nicht erst erhitzt werden muss. Die Befeuchtungsleistung könne ohne Abschlämmevorgänge aufgrund der Aufkonzentration von Mineralien genau zwischen null und 100 Prozent geregelt werden.

„Die Lebensdauer der Schwinger ist sehr lang und beträgt mindestens 10.000 Stunden. Die Ultraschallbefeuchter werden mit Frischwasser versorgt, weshalb die Befeuchtung sehr hygienisch ist und in der Hygieneverordnung auch dementsprechend behandelt wird. Der Betrieb mit entmineralisiertem Wasser ist ebenfalls möglich. Die dazu notwendigen Umkehrosmose-Anlagen sind in den Investitionskosten der letzten zehn Jahre sehr stark gesunken“, erörtert Dürr.

## Enthärtungsanlagen

Um kleinste Kalk- und Salzablagerungen im Rechenzentrum zu vermeiden, sollten insbesondere bei großen Wassermengen eine Enthärtungsanlage zur Vorenhärtung beziehungsweise zur Effizienzsteigerung der Umkehrosmose-Anlage installiert werden. Auch sei eine Nachbehandlung des Umkehrosmose-Wassers mittels Mischbettionentauschers zwingend angesagt. Die Leitfähigkeit des Befeuchter-Einspeisewassers müsse unter 5 Mikrosiemens ( $\mu\text{S}$ ) reduziert werden.

Eine solche Anlage sei sehr wartungsarm und müsse nur noch zyklisch überprüft werden. Bei Dampfbefeuchtern seien in Abhängigkeit von der Qualität des Leitungswassers die Kunststoff-Dampfzylinder möglicherweise öfters fällig. „Eine moderne Membran in der Umkehrosmoseanlage hält im Vergleich dazu zwei bis drei Jahre länger“, so Dürr.

## Frischluft im Rechenzentrum

Hier sei noch ein kurzer Abstecher erlaubt: Frischluft sei auch im Rechenzentrum durchaus empfehlenswert, „schon aus arbeitsrechtlichen Aspekten“, wie der Autor anführt. Dabei sei ein halber bis einfacher Raumluftwechsel völlig ausreichend. „Dies bedeutet, dass die komplette Luft einhalb- bis einmal pro Stunde komplett erneuert wird und dabei auch Gerüche abtransportiert. Dies erfolgt in der Regel über separate kleine Anlagen oder man schließt das Rechenzentrum an eine zentrale Lüftungsanlage an.“

## Was erwartet den Handbuch-Leser?

In der Neuauflage des erstmals 2013 erschienenen Buches sind zahlreiche Themen dazugekommen, etwa die Rechenzentrumsnorm DIN EN 50600 oder die „Begrünung“ von Rechenzentren durch noch höhere Energieeffizienz. Bereits auf Basis der Erstausgabe des Standardwerkes hat DataCenter-Insider verschiedene Artikel generiert, nun stellen wir weitere Themen daraus vor.

Das 685-Seiten starke Handbuch „IT-Räume und Rechenzentren planen und betreiben“ von Bernd Dürr ist im Verlag Bau+Technik erschienen und unter der Nummer ISBN 978-3-7640-0626-6 zu bestellen. Wir halten es für ein unverzichtbares Werk für alle Praktiker im Rechenzentrumsumfeld. ■ *Dr. Dietmar Müller*

## Der Anwender gewinnt

# Kubernetes und Cloud Foundry - Konkurrenz oder zwei Asse im Ärmel?

**Sie sind die aktuell stärksten Container-Projekte am Markt – doch sind sie deshalb gute Partner oder nicht doch Gegenspieler? Michael Jores von Suse zeigt: Mit der Containerisierung von Cloud Foundry kann Kubernetes die Verwaltung übernehmen und aus zwei Kräften entsteht ein belastbares Team.**

Cloud-Spezialisten orakelten schon lange, was im Jahr 2018 eintreten könnte. Der wachsende Einsatz von Anwendungscontainern war dabei oben auf der Liste der Voraussagen. Und tatsächlich: Container-IT überzeugt immer mehr Unternehmen – nach einer Erhebung des Analysehauses 451 Research wird der Markt von 762 Millionen Dollar im Jahr 2016 auf 2,7 Milliarden Dollar im Jahr 2020 wachsen.

In diesem expandierenden Markt tummeln sich unter anderen die zwei großen Namen Kubernetes und Cloud Foundry, von manchen IT-Experten als konkurrierende Projekte eingeschätzt. In der Tat gibt es einige funktionale Überschneidungen zwischen den Technologien: Allerdings dienen beide komplementären Zwecken, die auf das gleiche Ziel hinarbeiten. Und beide verfolgen Ansätze, die Unternehmen die Verwaltung ihrer gesamten Cloud-Umgebung erleichtern.

Dennoch sind Kubernetes und Cloud Foundry im gemeinsamen Einsatz ein fast unschlagbares Team. Drei Gründe sprechen für dafür.



Bild: © gemeinfrei / CC0

**Mit Cloud Foundry und Kubernetes auf der Hand kann fast nicht mehr schief gehen, schreibt Michael Jores, Regional Director Central Europe für Suse in Deutschland, Österreich und der Schweiz.**

## Nummer 1: Containerisierung ist die neue Virtualisierung

Container bieten neue Möglichkeiten zur Virtualisierung von Rechenleistungen. Sie sind schneller und effizienter als herkömmliche virtuelle Maschinen. Sie sind dank Tools wie Docker einfacher zu bauen und produktiv zu setzen.

Mit dem rasanten technischen Fortschritt und der Marktzepanz von Kubernetes steuert die IT nun auf einen neuen Standard mit containerzentrierten Infrastrukturplattformen zu. Kubernetes hat sich dabei durchgesetzt. Schon bald werden IT-Organisationen mit Kubernetes als allgegenwärtige Schicht in jeder Systeminfrastruktur arbeiten.

Für Anwender von Cloud Foundry ist Kubernetes eine spannende Entwicklung – denn die zunehmende Verfügbarkeit hochwertiger, containerbasierter Infrastruktur erleichtert ihnen die Arbeit. Cloud Foundry beschäftigt sich seit langem mit der Anwendungsbereitstellung über Container, auch wenn es selbst in einer traditionellen VM-Umgebung verwaltet wird.

Ein beliebter Anwendungsfall von Containern, vor allem in der Form von Container-as-a-Service-Angeboten, ist die Automatisierung von Entwicklungsprozessen im DevOps-Modell. Hier setzen immer mehr Unternehmen auf Container, um ihre eigene Entwicklung mit einer Build Pipeline zu automatisieren – das heißt, dass ihre Anwendungen automatisch von einer Plattform bereitgestellt werden, und zwar bereits containerisiert.

Meist handelt es sich dabei um Web-Anwendungen. Dabei können neben klassischen Server-Applikationen auch andere Bereiche wie Big Data bis hin zu Machine Learning von der beschleunigten Bereitstellung durch Container branchenübergreifend profitieren.

Daraus folgt, dass eine wachsende Zahl an Unternehmen, Behörden, Institutionen und unabhängige Software-Anbieter die Vorteile dieser neuen Infrastruktur nutzen und ihre Anwendungen in Container verpacken. So ist es nur konsequent, dass Cloud Foundry diesem Beispiel folgt.

Jüngste Initiativen zur Containerisierung wie das Open-Source-Projekt „Fissile“ und die „Suse Cloud Application Platform“ setzen diese Idee heute um. So kann Cloud Foundry einfach und überall eingesetzt werden und benötigt nur einen Bruchteil der Ressourcen VM-basierter Implementierungen.

In seiner containerisierten Form wird Cloud Foundry durch Kubernetes bereitgestellt, was die Wiederherstellung und Skalie-

rung der Cloud Foundry Plattform erst ermöglicht. Und da immer mehr Menschen Kubernetes-Know-how erwerben, lassen sich ihre Fähigkeiten nutzen, um die Kosten des Cloud Foundry-Managements zu senken. Das sind enorme Vorteile, die keinesfalls unterschätzt werden sollten.

## **Nummer 2: Anwendungen sind keine Container**

So mächtig die Containerisierung und so großartig Kubernetes auch sind, bestehen weiterhin Probleme auf Containerebene: Wo und wann sollte ich einen neuen Container in Betrieb nehmen? Wie funktionieren meine Container?

Das Arbeiten mit Anwendungen ist völlig anders gelagert. Das liegt daran, dass Anwendungen ein übergeordnetes Konstrukt sind, das aus mehreren und oft zahlreichen über ein Netzwerk verteilten Containern bestehen. Um eine Anwendung bereitzustellen, müssen Entwickler all diese Container in einer koordinierten Weise erstellen und verwalten. Und wenn sie Anwendungen in großem Maßstab bereitzustellen möchten, benötigen sie Werkzeuge, um diesen Aufwand zu automatisieren.

Genau das unterstützt die Cloud Foundry. Sie bietet Abstraktionen auf höherer Ebene zur Vereinfachung und Automatisierung der Anwendungsbereitstellung. Durch die Arbeit mit diesen Abstraktionen und das unsichtbare Erstellen und Verwalten der zugrundeliegenden Container erzielen Anwender und Betreiber enorme Produktivitätsvorteile.

## **Nummer 3: Kubernetes kann Cloud Foundry verwalten**

Für Kubernetes-Anwender sollte Cloud Foundry das Mittel der Wahl sein, wenn es um das Hinzufügen von Anwendungs-Management-Funktionen geht. Bis vor kurzem war die Kombination dieser Technologien jedoch etwas umständlich, da sich die beiden Plattformen völlig unabhängig voneinander entwickelt hatten.

Die gute Nachricht ist, dass sich das nun ändert. Da die Cloud Foundry containerisiert wurde und von Kubernetes eingesetzt und verwaltet werden kann, ist es relativ einfach, die Leistungsfähigkeit der Cloud Foundry einer bestehenden Kubernetes-Umgebung hinzuzufügen. Dies ist eine sehr angenehme Neuigkeit für Anwender, für die es bisher keine anspruchsvollen und ausgereiften Tools für das Anwendungs-Management gab.

Was heißt das nun konkret? Kubernetes als Orchestrierer bietet

die Grundlage der Arbeit mit Containern. Cloud Foundry kann als All-in-One-Plattform genutzt werden, etwa in Form der Suse Cloud Application Platform, die als Cloud Foundry Distribution Entwicklern ein Gesamtpaket bietet.

Die Plattform nimmt vom Entwickler den Quellcode entgegen und entdeckt anhand von Build Packs, welche Sprache und welche Arten von Container verwendet werden sollen. Die Anwendung wird automatisch kompiliert, gebaut und containerisiert bereitgestellt. Die Plattform sorgt sogar für die gewünschte Skalierung.

Kubernetes bietet ebenfalls alle diese Möglichkeiten – doch in Produkten wie der Suse-Plattform sind die Automatisierungsschritte bereits eingebaut. So können Entwickler deutlich Zeit und Aufwand sparen.

## Zwei sind mächtiger als einer

Es gibt noch einen weiteren sehr pragmatischen Grund, Kubernetes und Cloud Foundry gemeinsam einzusetzen. Denn wenn Entwickler container- und anwendungszentrierte Code-Management-Modelle kombinieren, nutzen sie im Rahmen einer einzigen Anwendung das Beste aus beiden Welten. Cloud Foundry unterstützt einen anwendungsorientierten Ansatz für das Code-Management und basiert auf einer definierten Reihe von Anwendungsarchitekturen und Bereitstellungsmustern.

Wenn Entwickler nun vor allem cloud-basierte Anwendungen mit diesen Mustern bereitstellen wollen, liefert Cloud Foundry eine deutliche Produktivitätssteigerung. Allerdings passt nicht jeder Code zu diesen Mustern. Deshalb brauchen Entwickler auch ein flexibleres, containerzentriertes Managementmodell, wie es von Kubernetes unterstützt wird, um die Codeverwaltungslücke zu schließen.

In diesen Fällen bietet es sich an, einen Teil der Produktivität der Cloud Foundry zu opfern, um von der höheren Flexibilität von Kubernetes zu profitieren. So können Entwickler eine neue Cloud-native Webanwendung erstellen, die aus mehreren Microservices einschließlich eines persistenten Datenspeichers im Backend besteht.

Wahrscheinlich können die meisten Microservices direkt und vollständig von Cloud Foundry verwaltet werden, so dass sie diese Komponenten für eine Produktivitätsmaximierung bereitstellen möchten. Der Datenspeicher kann jedoch eine benutzerdefinierte Implementierung erfordern und sollte dann auch als eigenständiger Container über Kubernetes bereitgestellt werden. Zurück in der

Cloud-Foundry-Welt können die Entwickler dann den Datenspeicher über Service Broker in die Web-Anwendung integrieren und so die beiden Managementsysteme effektiv überbrücken.

Auf diese Weise ist die komplette Anwendung aus Cloud Foundry heraus zu verfolgen. Es entsteht ein unschlagbares Team, mit dem Unternehmen das Beste aus Microservices, Containern und Open Source bietet. Am Ende sparen Unternehmen nicht nur Zeit und Aufwand mit einer dynamischen Bereitstellung und automatisierten Skalierung, sondern sie können mit Kubernetes und modernen Cloud-Foundry-basierten Plattformen auch den Footprint ihrer IT verkleinern. So gelingt der Start in die moderne Welt von DevOps und Machine Learning schneller, leichter und mit weniger Aufwand.

■ *Michael Jores*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



## Software-Ausblick: Serverless, Funktionen und Unikernels

# Was kommt nach den Containern?

**Bei der Suche nach weiteren Einsparungspotentialen – gerne als Effizienzsteigerung verkauft – haben sich die Betreiber von Rechenzentren nun der Software zugewandt. Microservices und Container schicken sich nämlich an, das Versprechen der Virtualisierung wahr zu machen und Server maximal auszulasten. Wir blicken eine Runde weiter: Wo führt das hin, was folgt auf Container?**

Aus Microservices zusammengesetzte Container sind eine Technologie, die sich aktuell fast ausschließlich in Rechenzentren von Enterprise-Kunden finden. Sie dient zuallererst der leichteren Portierbarkeit von Anwendungen in Multicloud-Umgebungen, sorgen aber auch für eine höhere Auslastung von Server-Racks – wodurch unterm Strich einige Server eingespart und nicht mit Strom versorgt werden müssen.

Die Entwicklergemeinde dahinter ist jung, dynamisch und auf der rastlosen Suche nach Wegen, um die Effizienz noch weiter zu steigern. Wir haben uns bei Betreibern und Entwicklern umgehört, was für sie der nächste logische Software-Schritt ist.

### Funktionen statt Container

„Das nächste große Ding nach Container ist Serverless“, kommt es wie aus der Pistole geschossen vom CTO von Suse, Dr. Thomas Di Giacomo. „Serverless im Sinne von Funktionen, die einen noch kleineren Fußabdruck hinterlassen als Microservices.“

Denn Serverless tritt noch schmäler auf als Microservices, weil die Funktionen weder ein Betriebssystem wie bei Virtuellen Ma



Bild: gemeinfrei/geralt/Pixabay/CC0

**Wir blicken in die nahe Zukunft: Was folgt in der Rechenzentrums-IT auf Container?**

schinen, noch Code, Laufzeit, Systemwerkzeuge, Systembibliotheken und Einstellungen wie bei Microservices benötigen. Für Serverless sind weder Funktionsaufrufe noch Container noch Maschinen erforderlich.

Die Funktionen selbst sind, wie der Oracle-Entwicklungschef Chad Arimura es unlängst auf DataCenter-Insider formulierte, „Codeblöcke, idealerweise klein und zweckbestimmt“. Sie werden über eine Functions-as-a-Service (FaaS)-Plattform koordiniert und terminiert, Beispiele dafür wären AWS Lambda, Azure Functions oder Google Cloud Functions, aus der Open-Source-Ecke kommen Frameworks wie das Fn-Flow-Projekt von Oracle, OpenWhisk oder OpenFaas.

Arimura spricht von „Serverless“ als das „das nächste SOA-Ding“ (SOA = Service Oriented Architecture). Der SOA-Hype der 0er-Jahre dürfte Brancheninsidern noch in (un)guter Erinnerung sein.

## Serverless

Auch Jürgen Hoffmann, Teamleiter von Multi-Cloud-Entwicklern bei Pivotal, zeigt sich überzeugt, „dass wir in drei Jahren viel über Serverless sprechen werden.“ Serverless befreie die Funktion von der Anwendung – das komme Anwenden gerade beim Internet der Dinge (Internet of Things; IoT) zugute.

Ein typisches Einsatzszenario aus der Kfz-Branche sähe laut Pivotal so aus: Sensoren und Echtzeitanalysen informieren Mechaniker über Defekte und erforderliche Wartungsarbeiten. Vorausschauende Analysen sagen den Zielort voraus, berechnen den Kraftstoffverbrauch beziehungsweise den Aktionsradius elektrischer Fahrzeuge und lotsen sie im Bedarfsfall zur nächsten Tankstelle oder Ladestation.

„Durch Serverless kann ich auf einen Event im IoT flexibel reagieren, weil die Funktion dafür zwar laufen, aber keine Ressourcen verbrauchen, und somit immer bereitsteht. Im Falle von sehr vielen Sensoren, etwa in Autos auf der Straße, müssen Sie die horizontal skalieren, etwa durch Plattformen wie Google Cloud Functions - aktuell als Betas einsehbar -, eine ereignisgesteuerte Server-lose Computing-Plattform. Diese sorgt dafür, dass die Funktionen nicht nur im Rechenzentrum laufen, sondern durch Cloud-Busting auf Cloud-Level vor Ort gebracht werden können. Wir selbst können mit unserer Cloud-nativen Plattform sowohl Anwendungen, als auch Container und Funktionen beliebig verschieben.“

## Unikernels

Chip Childers, Mitbegründer und aktueller CTO der Cloud Foundry Foundation, sieht Unikernels als den nächsten logischen Schritt nach Containern. „Container erhöhen die Auslastung von Servern, weil sie weniger OS benötigen. Diese Entwicklung wird von Unikernels auf die Spitze getrieben.“

Denn bei Unikernels würden nur die Funktionen geladen, die für eine Anwendung benötigt werden. Die Applikation werde dadurch buchstäblich bis auf den nackten Kern ausgezogen und lasse sich als ‚Image‘ auf einem Hypervisor oder direkt auf der Hardware ausführen, so Childers,

## Heute: Diskussionen auf dem DataCenter Day

Container, Serverless und Unikernels stehen auch auf dem heutigen DataCenter Day in Würzburg auf dem Programm; denn das diesjährige Motto »IT-Strategy drives Infrastructure« greift die gerade abgehandelte Frage direkt auf und spiegelt die Dynamik auf den Rechenzentrumsmarkts zurück. Der Fachkongress bringt Rechenzentrums-Betreiber, -Administratoren und -Verantwortliche mit Energie- und System-Managern sowie mit Technologie-Anbietern und Co-Locator zusammen, um sich über die wichtigsten Themen und Entwicklungen auszutauschen.

■ *Dr. Dietmar Müller*



Alles über Data Center täglich aktuell –

abonnieren Sie jetzt den Newsletter von DataCenter-Insider!



**Hier klicken**

# Überblick über das Angebot an Tauchbädern

## Server gehen immer öfter baden

**Steigerung der Effizienz, Effizienz, Effizienz – das ist das tägliche Brot der Rechenzentrumsbetreiber. Der zur Verfügung stehende Raum im Rechenzentrum muss immer besser ausgenutzt werden. Die Gerätedichte nimmt zu, zeitgleich steigen Stromverbrauch und Temperatur. In den Server-Räumen ist Wärmeverteilung und Kühlung ein Kostenproblem.**

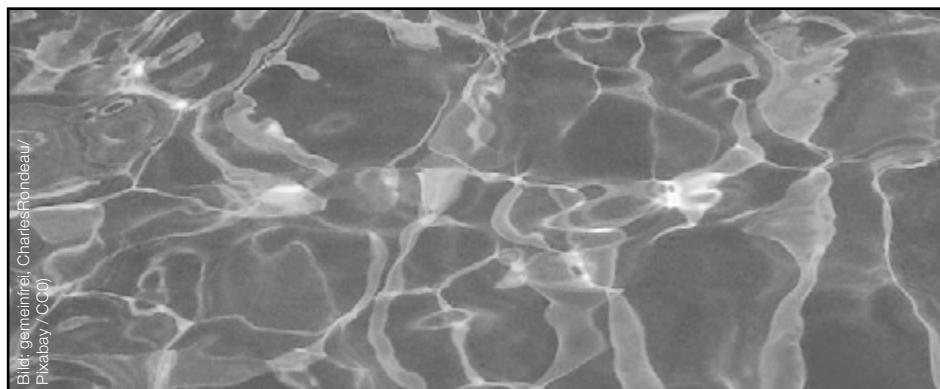
Der Kühlung von Rechenzentren kommt damit eine entscheidende Bedeutung zu, zumal sich mit einer effektiven Klimatisierung des Rechenzentrums die Energiekosten spürbar senken lassen. Bei herkömmlicher Kühlung mit Ventilatoren macht der Stromverbrauch für die Klimatisierung fast die Hälfte des Gesamtstromverbrauchs aus.

Deutschlandweit betrachtet verursacht der Betrieb von Kälte-Anlagen in Deutschland rund 14 Prozent des gesamten Strombedarfs. Der Anreiz für Effizienzmaßnahmen ist deshalb auch im Rechenzentrum groß: Aktuell versprechen Tauchbäder großes Energie-Einsparpotenzial.

### Wasserkühlung ist effektiver als Luft

Die Luftkühlung über Doppelböden ist ab Wärmelasten von 4.000 Watt pro Quadratmeter aufwärts nicht mehr möglich. Flüssigkeitskühlung bietet sich als Alternative an, zumal die Wärme-Abgabe an Wasser etwa 100mal höher ist als an Luft. Mit vergleichsweise kleinen wassergefüllten Rohren kann also vergleichsweise viel Wärme abtransportiert werden.

Laut dem Report „Data Center Liquid Cooling Market“ von Gartner Insights ist daher in den kommenden Jahren auch mit deutlich



**Beim Thema Flüssigkeitskühlung gerät in letzter Zeit verstärkt die Immersions- beziehungsweise Tauchkühlung in den Fokus.**

steigenden Umsätzen im Markt für Flüssigkeitskühlung zu rechnen. Zu den Wettbewerbern zählen Asetek, Rittal, Fujitsu, Submer, Data Aire, CoolIT, IBM, Schneider Electric, Stulz, Clima-veneta, 3M, Horizon, Emerson, GRC, Allied Control, Green Data Center und Midas Green Technologies.

## **Tauchkühlung immer beliebter**

Beim Thema Flüssigkeitskühlung gerät in letzter Zeit verstärkt die Immersions- beziehungsweise Tauchkühlung in den Fokus. Dabei werden ganze Server in einer elektrisch nicht leitfähigen Flüssigkeit gehalten; man spricht auch von „Liquid Submerged Server“.

Der komplette Server inklusive aller Komponenten wird dabei umströmt. Die warme Flüssigkeit wird aus dem Gebäude zu Rückkühlern geführt und dann wieder zurückgeleitet.

Solche Konzepte versprechen eine Reduktion der Kühlkosten weit über 50 Prozent – sind aber in der Anschaffung nicht billig. Da Wasser eher nicht an elektrische geladene Komponenten geraten sollte, kommen dabei verschiedene Kühlmittel zum Einsatz. Nach Wasser denkt man in unserem Sprachgebrauch an Öl als Kühlmittel.

## **Immersive Kühlung mit Öl**

Ein Beispiel für die immersive Kühlung mit Öl liefert Asperitas. Das niederländische Unternehmen hat im vergangenen Jahr mit „AIC24“ ein geschlossenes, wassergekühltes Öl-Tauchsystem vorgestellt, das für die Zirkulation der dielektrischen Flüssigkeit natürliche Konvektion beziehungsweise thermische Energie nutzt.

Das bedeutet, dass der primäre Kühlkreislauf keine Pumpen benötigt, um die Wärme an den Wärmetauscher zu transportieren. Dort gibt das Öl seine Hitze an Wasser ab – ein schöner Nebeneffekt, der genutzt werden will: Asperitas erläutert, dass so bis zu 98 Prozent der Abwärme aus den Server-Racks weiterverwendet werden können.

## **Zweiphasen-Tauchkühlung mit Novec**

Es geht aber noch raffinierter: Ein Pionier auf dem Gebiet der Tauchbäder ist 3M und seine dielektrische „Novec“-Flüssigkeit. Dielektrische Flüssigkeiten sind nichtleitend und daher selbststreichend besser als Wasser, wenn mit elektronischen Geräten gearbeitet wird. Zudem sind sie nicht-brennbar, nicht-explosiv, nicht-

korrosiv und über einen breiten Bereich an Betriebstemperaturen hin thermisch stabil.

97 Prozent soll der Einsatz von Novec im Vergleich zu herkömmlichen luftgekühlten Verfahren einsparen. Dabei sei es egal, so der Hersteller, ob eine Einphasen-Tauchkühlung - wie bei Asperitas – oder eine 2-Phasen-Tauch-Siedekühlung gewählt wird. Letztere verfügt aufgrund des Phasenüberganges von flüssig auf gasförmig angeblich über einen bis zu 4000mal (!) besseren Wirkungsgrad als die Luftkühlung.

Ein anschauliches Beispiel für die Zweiphasen-Immersionskühlung fand sich im Innovationslabor des RZ-Betreibers Deltalis. Dort wurde mit der 3M Novec 7100 High-Tech Flüssigkeit, ein Hydrofluorether, ein Wärme-Management-Medium eingesetzt, das schon bei 61 Grad Celsius verdampft. Deltalis will aber – ohne Angabe der Gründe - zum Jahresende den RZ-Betrieb komplett einstellen.

Nach wie vor in Betrieb ist das von Allied Control / Bitfury Group gebauten 40+-Megawatt-Rechenzentrum in der Republik Georgien. Es enthält das mutmaßlich weltweit leistungsstärkste Beispiel für das Zwei-Phasen Tauch-Siedekühlungsverfahrens („2PIC“). Auch dort kommt die 3M Novec 7100-Flüssigkeit als Tauchkühlmittel zum Einsatz.

„Die PUE (Power Usage Effectiveness), die wir in unserem Rechenzentrum in Georgien erzielen können, führt uns zu einem neuen Niveau der Kühlung von Rechenzentren für Hochleistungs-Computing“, so Valery Vavilov, CEO von Bitfury. Die Phase 1 dieses Projekts umfasste eine 40 Megawatt IT-Last mit einem PUE von 1,02, aktuell wird das Rechenzentrum erweitert und umgebaut.

## **Fluorinerts versus Hydrofluorether**

Es gibt aber noch weitere dielektrische Flüssigkeiten - eine andere, proprietäre, ungiftige und nicht-leitende Kühlflüssigkeit für Immersionskühlösungen stammt etwa von GRC (Green Revolution Cooling) und trägt die Bezeichnung „ElectroSafe“. Es soll die 1200fache Wärmekapazität von Luft bieten und Kühlleistung von bis zu 100 Kilowatt pro Rack offerieren.

F2 Chemicals hat „Flutec“ entwickelt. Auch 3M bietet Alternativen an, etwa „Fluorinert“. Dieses kommt auch im „Primergy Liquid Immersion Cooling-System von Fujitsu zum Einsatz.

Perfluorierte Kohlenstoffe wie Fluorinert kommen nicht zuletzt in militärischen Systemen für die Wärme-Ableitung zum Einsatz. Nachteil: Sie verharren sehr lange in der Atmosphäre und gelten daher als Ursache für das so genannte Ozon-Loch. 3M hat deswegen Hydrofluorether (HFE) als Alternative entwickelt. Diese Verbindungen verbleiben nur bis zu vier Jahren in der Atmosphäre.

## Vorgefertigter Tauchbadserver

Im November dieses Jahres will Fujitsu mit dem Vertrieb seines Servers Primergy Liquid Immersion Cooling System in Japan beginnen, dessen Prototyp auf dem Fujitsu Forum 2016 vorgestellt worden war. Es soll den Stromverbrauch im Vergleich zu herkömmlichen Luft-Kühlsystemen um etwa 40 Prozent reduzieren. Da der Server keine internen Ventilatoren benötigt, offeriert er eine PUE von 1,07.

Das Rechenzentrum für Biologie und Bioinformatik am Indian Institute of Technology Delhi nutzt die Liquid-Cooling-Lösung bereits, wo es für seinen geringen Stromverbrauch und den leisen Betrieb durch den Wegfall der Lüfter gelobt wurde. Sie ist auch schon bei Flexiscale Technologies in Nottingham im Einsatz, es gibt ein Video davon.

## Zukunftsansichten

Getauchte Server finden sich aktuell noch vorrangig in High Performance Computing-Regionen. Mit einem Ausbruch aus den akademischen Gefilden ist aber in Bälde zu rechnen, die „Data Center Liquid Cooling Market“-Studie von Garner Insights sagt dies ja auch voraus.

Da ein in Kühlflüssigkeit eingetauchter Server nicht der Außenluft ausgesetzt ist, kann er in einer Vielzahl von Umgebungen installiert werden, die dafür eigentlich gar nicht geeignet sind. Zu denken wäre etwa an hohe Temperaturen, hohe Feuchtigkeit oder Produktionsstätten, die ölig oder staubig sind. Auch in großer Höhe oder in Küstengebieten, in denen es sehr salzig ist, könnten solche abgekapselten Server tätig sein.

Unmittelbar denkt man da an die immer weiter verbreiteten Edge-Rechenzentren, die am Rande von Internet of Things (IoT)-Netzwerken operieren. Mit großer Wahrscheinlichkeit werden wir daher künftig viele weitere Systeme dieser Art sehen.

■ *Dr. Dietmar Müller*

# CIO BRIEFING

Technology-Update für IT-Manager

Regelmäßig  
kostenlos lesen?

+++ Jetzt anmelden auf +++



[www.bigdata-insider.de/cio](http://www.bigdata-insider.de/cio)



[www.cloudcomputing-insider.de/cio](http://www.cloudcomputing-insider.de/cio)



[www.datacenter-insider.de/cio](http://www.datacenter-insider.de/cio)



[www.dev-insider.de/cio](http://www.dev-insider.de/cio)



[www.ip-insider.de/cio](http://www.ip-insider.de/cio)



[www.security-insider.de/cio](http://www.security-insider.de/cio)



[www.storage-insider.de/cio](http://www.storage-insider.de/cio)