Cloud Computing erfordert stabile und zuverlässige Lösungen für den Edge-Bereich

White Paper 256

Version 0

von Kevin Brown und Wendy Torell

Zusammenfassung

Immer mehr Unternehmen setzen heute auf Cloud Computing. Eine stärkere Abhängigkeit von Cloud-basierten Anwendungen bedeutet aber auch, dass Unternehmen die Redundanz der technischen Infrastruktur (Stromversorgung, Kühlung, Netzwerk) neu überdenken müssen, die vor Ort im Edge-Bereich installiert ist. In diesem Dokument beschreiben und kritisieren wir die heute gängigen Lösungen der technischen Infrastruktur, präsentieren eine Methode zur Analyse der erforderlichen Ausfallsicherheit und beschreiben Best Practices, die sicherstellen, dass Mitarbeiter ihre unternehmenskritischen Anwendungen ohne Unterbrechung nutzen können.

Einleitung

Das anhaltende Wachstum im Bereich Internet of Things (IoT), das steigende Volumen des Datenverkehrs und die zunehmende Verbreitung von cloud-basierten Anwendungen sind wichtige Technologietrends, die zu Veränderungen der Datacenter-Strukturen führen.

Kritische Anwendungen, die bisher in unternehmenseigenen Datacentern „on-premise“ liefen, wurden verlagert in große bis sehr große Cloud-Datacenter. Doch nicht alle Anwendungen laufen heute in der Cloud. Die Gründe dafür sind vielfältig – rechtliche Vorgaben, Unternehmenskultur, proprietäre Anwendungen und Latenz sind nur einige davon.

Dadurch ist eine so genannte „hybride“ Datacenter-Struktur entstanden. Diese besteht aus einem Mix von (1) zentralen Cloud-Datacentern, (2) mittleren bis großen regionalen Datacentern und (3) kleineren lokalen On-Premise-Datacentern. Siehe **Abbildung 1**. Wo früher ein 1-Megawatt-Datacenter vor Ort in einer Unternehmensniederlassung zum Einsatz kam, stehen heute ein paar Racks mit IT-Systemen für kritische Anwendungen und die Bereitstellung der Netzwerkanbindung an die Cloud. Auch wenn Stellflächen und Kapazität dieser neuen On-Premise-Datacenter wesentlich kleiner sind, sollte ihre Bedeutung für die Unternehmen nicht unterschätzt werden. Häufig sind die am Standort verbliebenen IT-Systeme wichtiger als je zuvor.

In diesem Dokument beschreiben wir die gängige Praxis in den zuvor genannten Datacenter-Varianten, erklären, wie sich die Verfügbarkeitsanforderungen geändert haben und präsentieren eine Methode zur Evaluierung der erforderlichen Ausfallsicherheit in Edge-Rechenzentren (On-Premise), die sicherstellt, dass die Unternehmensziele erreicht werden. Zudem beschreiben wir Best Practices für die Implementierung von Micro-Datacentern im Edge-Bereich.

Abbildung 1
3 Arten von Datacentern. In diesem Dokument konzentrieren wir uns auf lokale Edge-Datacenter.

Verschiedene Arten von Datacentern

Das zentrale Cloud-Datacenter wurde ursprünglich nur für bestimmte Arten von Anwendungen genutzt, z. B. E-Mail, Lohnbuchhaltung, Social Media usw., die nicht zeitkritisch sind. Doch mit der Verlagerung von kritischen Anwendungen in die Cloud wurde auch deutlich, dass Faktoren wie Latenz, Bandbreitenbeschränkungen, Sicherheit sowie gesetzliche Anforderungen berücksichtigt werden müssen. Denken Sie nur an Systeme für selbstfahrende Autos. Damit diese einwandfrei funktionieren können, ist eine hohe Rechenleistung erforderlich, und Latenzen würden unweigerlich zu Unfällen führen. Auch im Gesundheitswesen gibt es lebenswichtige Anwendungen: Sensoren, die Patientenzustände erfassen, oder medizinische Instrumente, die Chirurgen während der Operation mit Feedback in Echtzeit versorgen. Daher war es erforderlich, die Rechenleistung näher an die Orte zu bringen, wo sie benötigt wird.

Auch für die Verteilung von großen Datenvolumen ist es wichtig, dass die relevanten Daten in der Nähe der Nutzer vorgehalten werden. Das reduziert Kosten für die bandbreitenintensive Übertragung und verbessert die Streaming-Performance.

Viele Unternehmen müssen (oder wollen) unternehmenskritische Anwendungen direkt an ihrem Standort einsetzen. Dadurch haben sie eine bessere Kontrolle und erfüllen gesetzliche Vorgaben sowie Anforderungen an die Verfügbarkeit. In einigen Fällen werden diese Anwendungen in der Cloud repliziert, um Redundanz herzustellen.

Das Schneider Electric White Paper 226, Die treibende Kraft hinter Edge Computing und die Vorteile der Technologie, beschreibt, wie diese Art von Anwendungen zu einer Struktur führt, die mehr regionale und lokale Datacenter integriert. In diesem Abschnitt beschreiben wir die einzelnen Arten von Datacentern und erläutern die jeweils typischen Konzepte für die technische Infrastruktur.

Zentrale Datacenter

Große, zentrale Datacenter mit mehreren Megawatt – als Bestandteil einer Cloud oder im Besitz von Unternehmen – werden häufig als aufgabenkritische Installationen eingestuft und daher bereits mit hoher Ausfallsicherheit geplant. Dafür gibt es bewährte Best Practices, die bereits seit Jahren angewandt werden, um Ausfälle zu vermeiden. Für das zuständige Personal von IT-Abteilungen und Gebäudemanagement hat ein unterbrechungsfreier Betrieb dieser Datacenter höchste Priorität. Zudem sind sie häufig nach den Standards des Uptime Institute konzipiert und in einigen Fällen auch als Tier 3 oder Tier 4 zertifiziert. Colocation- und Cloud-Provider werben häufig mit diesem High-Availability-Design für ihre Datacenter.

Folgende Best Practices gehören zu den Merkmalen:

- **Redundante kritische Systeme** – kritische Stromversorgungs- und Kühlsysteme sind redundant ausgelegt (häufig mit 2N-Redundanz), um Betriebsunterbrechungen durch Systemausfälle oder Wartungsarbeiten zu vermeiden.
- **Hohe physische Sicherheit** – häufig zu finden sind biometrische Sensoren an Türen, Sicherheitsschleusen, Videoüberwachung und Wachdienste, die rund um die Uhr sicherstellen, dass nur autorisiertes Personal Zutritt hat.
- **Organisierte Racks und Reihen** – neben den Racks mit Verriegelung sorgt auch die geordnete Verlegung von Strom- und Netzwerkkabeln dafür, Handlingsfehler zu reduzieren (z. B. durch Herausziehen der falschen Kabel, Anschluss von zwei Stromversorgungen an einen Stromkreis, usw.). Die Luftverteilung ist optimiert und Komponenten wie Bürstenleisten und Blenden reduzieren die Gefahr von Hotspots (Wärmenestern).
- **Überwachung** – Sensoren und Messgeräte erfassen Informationen für Datacenter-Infrastrukturmanagement (DCIM) und Gebäudemanagement (BMS), die sämtliche Systeme im Datacenter managen und optimieren.

Abbildung 2 zeigt die Sicherheitsvorkehrungen in diesen Datacentern:


Biometrische Sensoren


Sicherheitsschleusen


Wachpersonal

Abbildung 2
Gängige Sicherheitsvorkehrungen in zentralen Cloud- und Colocation-Datacentern

Regionales Datacenter

Regionale Datacenter liegen näher an den Endpunkten (wo Daten generiert und genutzt werden) und sind kleiner als die großen, zentralen Datacenter. Wie zuvor beschrieben, laufen in diesen Installationen Anwendungen, die geringe Latenzen und hohe Bandbreiten erfordern, und daher eine größere räumliche Nähe zu den Nutzern benötigen. Die Lage dieser Datacenter ist so gewählt, dass große Datenvolumen problemlos übertragen werden können. Diese Datacenter dienen als eine Art Brücke zwischen den zentralen Datacentern und den lokalen On-Premise-Datacentern.

Auch die regionalen Datacenter werden in der Regel für hohe Sicherheits- und Verfügbarkeitsanforderungen konzipiert. Häufig erfüllen diese Einrichtungen auch den Standard Tier 3. In einigen Fällen werden vorkonfektionierte Lösungen verwendet. Referenzdesigns bieten dafür eine gute Grundlage (siehe Beispiel in **Abbildung 3**).

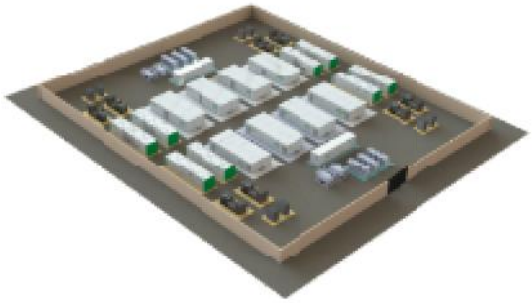

Abbildung 3
Referenzdesign für ein zentrales oder regionales Datacenter

Lokales Datacenter

Ein lokales Datacenter befindet sich am gleichen Standort wie die Nutzer. Es gibt verschiedene Bezeichnungen für diese Variante, unter anderem **On-Premise-Datacenter**, **Edge-Datacenter** oder **Micro-Datacenter**. Die Kapazität lokaler Datacenter reicht von 1 - 2 MW bis zu 10 - 20 kW. Da Unternehmen immer mehr Geschäftsanwendungen zu Cloud- oder Colocation-Anbietern auslagern, sind viele dieser lokalen Datacenter eher kleinere Installationen, die in einigen Fällen nur aus ein paar Racks in einem kleinen Raum bestehen.

Viele dieser Downsized-Datacenter sind heute so konzipiert, dass sie einem Tier 1 Standard entsprechen. Redundanz und Ausfallsicherheit spielen dabei häufig eine untergeordnete Rolle. Nachfolgend beschreiben wir einige Szenarien, die nicht selten zur gängigen Praxis in kleinen On-Premise-Datacentern gehören:

- **Mangelnde Sicherheit** – die Räume sind häufig ungesichert und Racks sind offen (ohne Türen).
- **Unorganisiert Racks** – fehlendes Kabelmanagement; die Folgen sind Kabelsalat und ungenügende Luftverteilung im Rack sowie häufige Handlingsfehler bei Ergänzungen oder Änderungen der Konfiguration. Siehe **Abbildung 4**.
- **Keine Redundanz** – Stromversorgungssysteme (USV, Verteilung) sind oft nur einfach vorhanden (1 N), wodurch sich die Ausfallsicherheit und die Möglichkeiten für die Wartung bei laufendem Betrieb verringern.
- **Keine dedizierte Kühlung** - die kleinen Räume sind vielfach nur an die Gebäudeklimaanlagen angeschlossen, sodass es zur Überhitzung von Systemen kommen kann.
- **Kein DCIM Monitoring** – für diese Räume gibt es häufig kein zuständiges Personal oder Softwarelösungen für das Management der Systeme und die Optimierung der Ausfallsicherheit.

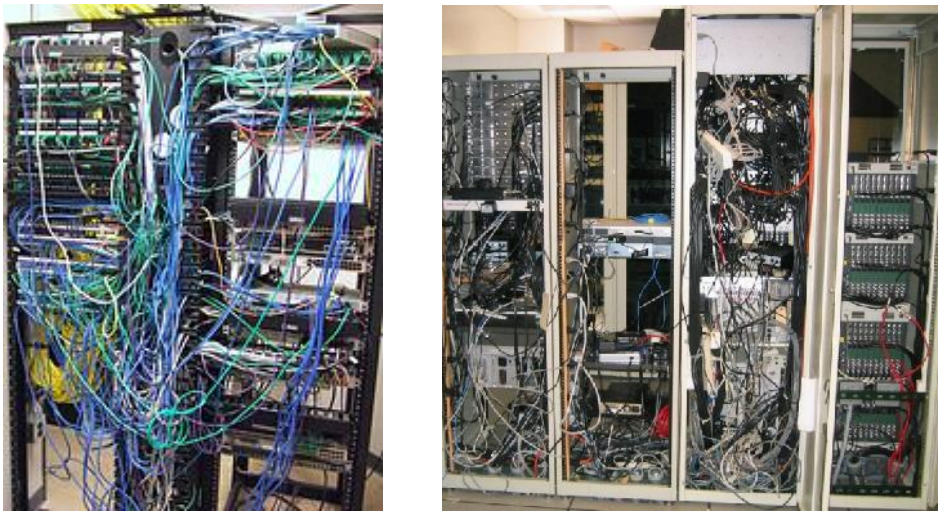

Abbildung 4
Beispiele für kleine On-Premise-Datacenter mit fehlendem Kabelmanagement und schlechten Sicherheitsvorkehrungen.

In vielen Fällen sehen die IT-Räume so aus wie auf diesen Bildern. Wenn die Unternehmen ihre Anwendungen in die Cloud oder in Colocation-Center verlagern, werden die wenigen verbleibenden Racks oft vernachlässigt. Der Fokus liegt dann häufig auf der Ausfallsicherheit der größeren Datacenter. Doch das kann zu Problemen führen, weil die verbleibenden Racks in der Regel ebenso wichtig oder sogar noch wichtiger für die Unternehmen sind.

Im Allgemeinen verbleiben die folgenden Komponenten am Unternehmensstandort (on-premise): (1) proprietäre, unternehmenskritische Anwendungen und (2) Systeme, die eine Netzwerkanbinduing zur Cloud sicherstellen. Die entscheidende Frage lautet: Welche Auswirkungen hat es auf die Produktivität meines Unternehmens, wenn ein Zugriff auf meine Anwendungen nicht möglich ist? Wenn wir davon ausgehen, dass die gleiche Anzahl von Menschen weiterhin am gleichen Ort arbeitet, wo nur noch wenige Racks verblieben sind, **erhöht dies die Bedeutung jedes dieser verbliebenen Racks**. Die lokalen Systeme sind wichtig für die Verbindung zu den täglich genutzten Geschäftsanwendungen. Angesichts der zunehmenden Nutzung der Cloud führt ein Ausfall der Systeme für die Netzwerkverbindungen zu einem beträchtlichen Produktivitätsverlust.

Daher ist ein neuer Ansatz für die Konzeption dieser kleinen On-Premise-Datacenter gefragt. Es reicht nicht aus, wenn wir uns auf die Ausfallsicherheit der großen zentralen und regionalen Datacenter konzentrieren. Die lokalen Installationen sind ebenso wichtig, weil sie aktuell das schwächste Glied in der Kette darstellen. In einem späteren Abschnitt dieses Dokuments erläutern wir Best Practices für diese Installationen, die unterbrechungsfreie Netzwerkverbindungen und hohe Ausfallsicherheit für maximale Produktivität sicherstellen.

Umfassende Evaluierung der Ausfallsicherheit

In diesen vernetzten, hybriden Datacenter-Strukturen müssen wir einen neuen Ansatz für die Definition von Kritikalität und Redundanz finden. Die Tools, die heute in der Datacenter-Branche eingesetzt werden, konzentrieren sich lediglich darauf, die Ausfallsicherheit eines einzelnen Datacenters zu maximieren. Standardvorgaben für verschiedene Tier ermöglichen die Konfiguration für bestimmte Verfügbarkeits-Level (z. B. 99,99%). Ein Ausfall ist definiert als Betriebsunterbrechung von IT-Systemen innerhalb eines Datacenter.

Die Tools und Parameter berücksichtigen weder Abhängigkeiten zwischen mehreren Datacentern, die Anzahl der von einem Ausfall betroffenen Nutzer noch die Kritikalität der betroffenen Geschäftsfunktionen oder das Failover von Anwendungen (Software). Doch genau das muss sich ändern.

Die neuen Verfügbarkeitsanforderungen

Die Erwartungen der Mitarbeiter heute unterscheiden sich von denen früherer Generationen. Mit den demografischen Veränderungen und dem zunehmenden Anteil von Millennials in der Arbeitswelt verändern sich auch die Anforderungen. Die neue Generation ist bereits mit „Always on“ aufgewachsen, sie ist es gewohnt, dass IT-Systeme dauerhaft mit dem Netz verbunden sind und unterbrechungsfrei funktionieren. Die Toleranzschwelle für Ausfälle ist sehr gering. Technologie spielt eine wichtige Rolle im Alltag dieser Generation und natürlich auch bei der Arbeit. So gaben auch 82% der Millennials an, dass die Technologie am Arbeitsplatz ein Kriterium für die Auswahl des Arbeitgebers ist.[1]

Wenn wir davon ausgehen, dass sich dieser Trend fortsetzt, bedarf es einer ganzheitlichen Perspektive auf die Bewertung der Zuverlässigkeit von Datacentern, um die richtigen Entscheidungen für das Datacenter-Design treffen zu können. „Man kann nicht steuern, was man nicht messen kann“, sagt ein altes Sprichwort. Daher ist es erforderlich, dass wir die Parameter für die Zuverlässigkeit an die aktuellen Anforderungen von Unternehmen anpassen.

Ein neuer Ansatz

Eine neue Perspektive zur Bewertung der Ausfallsicherheit verändert auch die Maßnahmen. **Tabelle 1** zeigt den Vergleich des bisherigen (alten) Szenarios mit dem neuen Szenario.

Tabelle 1
Alte und neue Szenarien für Datacenter-Ausfälle

Altes Szenario	Neues Szenario
Fokus auf zentralem Datacenter	Fokus auf der hybriden Struktur
Ausfalldefinition: IT-Systeme in einem Rack sind betroffen	Ausfalldefinition: Benutzer sind betroffen
Remote-Standorte oder Personen/Funktionen werden nicht berücksichtigt	Kritikalität ist abhängig von der Anzahl der betroffenen Mitarbeiter und Funktionen

Denken wir nur an einen Stromversorger und seine Bewertung von Ausfallsicherheit. Dabei werden nicht nur die Kraftwerke und Hochspannungsleitungen (in unserer Struktur: das zentrale Datacenter) betrachtet. Das Unternehmen beschneidet Bäume, wartet Transformatoren usw. – all diese Maßnahmen tragen zum Erfolg des Unternehmens bei. Und der Erfolg wird daran gemessen, wie zuverlässig der Strom für die Kunden (in unserer Struktur: die Edge-Datacenter) bereitgestellt werden kann. Die Datacenter-Branche muss lernen, ähnlich zu denken wie Versorgungsunternehmen, denn der Edge-Bereich ist ebenso wichtig wie die zentralen Datacenter.

Die Verfügbarkeit von zwei in Serie geschalteten Systemen lässt sich in eine Formel fassen:

$$\text{Verfügbarkeit}_{\text{System}} = \text{Verfügbarkeit}_1 * \text{Verfügbarkeit}_2$$

[1] http://www.dell.com/learn/us/en/uscorp1/press-releases/2016-07-18-future-workforce-study-provides-key-insights (zuletzt aufgerufen am 31.10.2016)

Nehmen wir an, ein Mitarbeiter ist auf das lokale On-Premise-Datacenter *und* das zentrale Datacenter gleichermaßen angewiesen, um seine Aufgaben zu erfüllen. Um die Datacenter-Verfügbarkeit aus dieser Perspektive zu berechnen, verwenden wir diese Formel. Wenn beispielsweise das zentrale Datacenter eine Verfügbarkeit von 99,98% hat (Tier 3 Datacenter, mit 1,6 Stunden Ausfallzeit), und das On-Premise-Datacenter eine Verfügbarkeit von 99,67% (Tier 1 Datacenter, mit 28,8 Stunden Ausfallzeit), ergibt sich aus Sicht des Mitarbeiters insgesamt eine Verfügbarkeit von 99,98%*99,67% = 99,65% (entspricht 30,7 Stunden Ausfallzeit).

Wie bewertet nun ein IT-Leiter (CIO) die Auswirkungen der gesamten Datacenter-Struktur auf die Produktivität und die Connectivity? Nicht jedes Datacenter ist von jedem anderen Datacenter abhängig, um funktionierende Anwendungen für Mitarbeiter bereitzustellen. Beispiel: Eine Niederlassung in London ist nicht abhängig von einer Niederlassung in Kalifornien, aber beide sind abhängig von einem zentralen Datacenter in New York.

Nicht alle Datacenter haben die gleiche Bedeutung für ein Unternehmen. Ein entscheidender Faktor ist die Anzahl der betroffenen Mitarbeiter. So ist beispielsweise ein On-Premise-Datacenter für 1000 Mitarbeiter wichtiger als eines für 10 Mitarbeiter. **Tabelle 2** zeigt die Anzahl der durch Ausfälle verlorenen Mannstunden in einer Struktur mit einem zentralen Tier 3 Datacenter und 10 lokalen Tier 1 Datacentern für jeweils 100 Mitarbeiter. Die Tabelle macht deutlich, dass die Tier 1 Edge-Datacenter der entscheidende Faktor für die gesamten Auswirkungen von Ausfällen sind. Je größer die Anzahl der Edge-Standorte, desto geringer ist die Anzahl der Stunden, in denen kein Standort Ausfälle verzeichnet.

Tabelle 2

Verfügbarkeit von 10 Edge Datacentern und 1 zentralen Datacenter unter Berücksichtigung der

Data Center Availability						
Description	**Availability**	**Downtime (hrs)**	**# Sites**	**# people/site**	**Total people impacted**	**People-hours of downtime/yr**
Tier 1 edge data centers	99.67%	28.82	10	100	1,000	28,820
Tier 3 central datacenter	99.98%	1.58	1	0	1,000	1,580
					Total people-hours of downtime/yr	30,400
					Availability	99.65%

Tabelle 2 zeigt ein einfaches Szenario mit zwei unterschiedlichen Datacenter-Ebenen, wobei insgesamt 1000 Mitarbeiter durch Ausfälle in beiden Ebenen betroffen sind. In Szenarien mit mehreren Datacentern (mit unterschiedlichen Verfügbarkeitseigenschaften) und unterschiedlicher Mitarbeiterzahl ist die Berechnung komplexer. Außerdem ist das genannte Beispiel unvollständig, weil es die Geschäftsfunktionen an den einzelnen Standorten unberücksichtigt lässt. Ein Standort, der für den Kundenservice oder die Fertigung genutzt wird, hat eine kritischere Bedeutung als ein Standort mit Administratoren, die ebenso remote arbeiten könnten, wenn ihr Netzwerk ausfällt.

Für einen ganzheitlichen Ansatz zur Bewertung aller Standorte schlagen wir ein Scorecard-System vor (siehe **Tabelle 3**). Mit diesem System können CIOs und Datacenter-Manager die Standorte mit der höchsten Priorität identifizieren, um entsprechende Maßnahmen einzuleiten. Die Scorecard enthält Angaben zu Verfügbarkeit und Ausfallzeiten jedes Standorts in der Datacenter-Struktur (möglichst basierend auf Messungen) sowie den jeweiligen Score unter Berücksichtigung der Kritikalität für jeden Standort.

Kritikalitätsanalyse

„Qualitative Criticality Analysis“ ist eine bewährte Methode zur Bewertung von Risiken und Priorisierung von Korrekturmaßnahmen (auch bezeichnet als Failure Modes, Effects and Criticality Analysis (FMECA)). Diese Methode ist in Fachpublikationen für Zuverlässigkeitstechnik umfassend dokumentiert. Die Analyse berücksichtigt auch den Schwergrad der Auswirkungen in Form einer Risk Priority Number (RPN). Die RPN basiert auf 3 Faktoren: (1) Schadensausmaß, (2) Auftretenswahrscheinlichkeit und (3) Entdeckungswahrscheinlichkeit.[2]

Siehe **Infokasten** links für mehr Informationen zur Kritikalitätsanalyse[2]. Für Datacenter basiert das „Schadensausmaß“ an jedem Standort auf den folgenden Faktoren:

- Anzahl der betroffenen Mitarbeiter
- Betroffene Funktionen

Gängig ist eine Skala von 1 bis 5, wobei 1 für die geringsten Auswirkungen bei einem Ausfall steht und 5 für die größten Auswirkungen. Auch wenn es sich hier um ein qualitatives Bewertungssystem handelt, bietet es einen systematischen Ansatz für die Evaluation aller Standorte der gesamten Datacenter-Struktur. Bitte beachten Sie, dass jedes Unternehmen andere Präferenzen für die genannten Bewertungen hat. Wichtig ist dabei lediglich, dass eine konsistente Methode für die Bewertung aller Standorte verwendet wird.

In diesem Beispiel wurden fünf Datacenter in einer hypothetischen Struktur betrachtet. Die jährliche Ausfallzeit wird mit dem jeweiligen Wert für das Schadensausmaß multipliziert, um das gewichtete Ergebnis zu erhalten.

Dann lassen sich die Standorte einfach nach dem erreichten Score sortieren. Der Standort mit dem höchsten Wert sollte Priorität haben, wenn es um die Durchführung von Verbesserungsmaßnahmen geht. Der Score kann zudem in Prozent für jeden Standort angegeben werden (siehe Beispiel), wobei die Standorte mit dem höchsten Prozentwert die höchste Priorität erhalten.

Tabelle 3

Beispiel eines Scorecard-Systems für die Priorisierung von Verbesserungsmaßnahmen im

Data Center Scorecard					
Site Name	**Availability**	**Annual Downtime (hours)**	**Severity of Effects of Failure (1-5)***	**Score (weighted for criticality)**	**Site impact on Score**
1	99.98%	1.752	2	3.5	0.4%
2	99.20%	70.08	4	280.3	30.0%
3	99.60%	35.04	1	35.0	3.7%
4	98.60%	122.64	5	613.2	**65.5%**
5	99.98%	1.752	2	3.5	0.4%
			Overall criticality score:	*935.6*	

Wir raten hier zu einem schrittweisen Vorgehen. Nachdem die Verfügbarkeit von Standort 4 in diesem Beispiel verbessert wurde, rückt ein anderer Standort auf Rang 1 dieser Liste vor. Durch einen kontinuierlichen Verbesserungsprozess werden jeweils die Standorte optimiert, an denen der größte Effekt erzielt wird.

Mit einer passenden Methode für das Availability Reporting sollte schnell deutlich werden, wo Optimierungen erforderlich sind, um die größten Verbesserungen bei Produktivität und Return-on-Investment zu erzielen. **In der Mehrzahl der Fälle wird sich zeigen, dass die Edge-Datacenter mit ihrer geringeren Ausfallsicherheit das größere Schadenspotenzial für die Unternehmen haben.**

[2] http://www.weibull.com/hotwire/issue46/relbasics46.htm (Zuletzt aufgerufen am 31.10.2016)

Best Practices für Edge-Datacenter

Verwendet man die richtigen Bewertungsmethoden, wird schnell deutlich, dass Optimierungen der Datacenter-Systeme im Edge-Bereich unumgänglich sind. Die bisherige Praxis in den Edge-Datacentern (wie zuvor beschrieben) ist angesichts der unternehmenskritischen Aufgaben dieser Installationen nicht mehr angemessen. Verbesserungen müssen in folgenden Bereichen erfolgen:

- Physische Sicherheit
- Monitoring (DCIM), Betriebsabläufe, Remote Monitoring
- Redundante Stromversorgung und Kühlung
- Doppelte Netzeinspeisung

In den nachfolgenden Abschnitten beschreiben wir die wichtigsten Best Practices Edge-Datacenter. Das Schneider Electric White Paper 174, *Practical Options for Deploying Small Server Rooms and Micro Data Centers*, beschreibt detailliert effektive Verbesserungsmaßnahmen in den Bereichen Stromversorgung, Kühlung, Racks, physische Sicherheit und Monitoring in kleinen Serverräumen und Niederlassungen mit einer Last von bis zu 10 kW.

Sichere Installationen

Kleine lokale Datacenter sind häufig in Räumen ohne Zugangsbeschränkungen (z. B. in Büroräumen) untergebracht. Häufig gibt es keine dedizierten, separaten IT-Räume, daher sind offene Racks nicht gesichert. Dadurch entstehen Risiken durch böswillige oder fahrlässige Handlungen.

Wir empfehlen folgende Best Practices zur Risikominimierung:

- Verlagern Sie die Systeme in einen abschließbaren Raum oder verwenden Sie verriegelbare Gehäuse.
- Installieren sie biometrische oder andere Zugangskontrollsysteme.
- In industriellen Standorten müssen die Systeme in einem Gehäuse untergebracht werden, das Schutz gegen Feuer, Wassereinbrüche, Feuchtigkeit, Vandalismus und elektromagnetische Störungen bietet.
- Installieren Sie Systeme zur Sicherheits- und Raumüberwachung (24 x 7) sowie Videoüberwachung

Beispiele für sichere Gehäuse sind in **Abbildung 5** dargestellt. Diese sind häufig bereits ab Werk vorkonfektioniert und enthalten alle erforderlichen Infrastrukturkomponenten.

Abbildung 5
Beispiele für Micro-Datacenter von Schneider Electric

Datacenter-Management

Management und Betriebsprotokolle unterscheiden sich häufig von einem Edge-Datacenter zum anderen (soweit Protokolle vorhanden sind). Das Management von Hunderten oder Tausenden Edge-Standorten kann kostenintensiv und zeitaufwändig sein. Zudem ist die Verfügbarkeit an vielen Standorten abhängig von gemeinsam genutzten Infrastruktursystemen wie Generatoren, Schaltanlagen, und Kaltwassersätzen.

Wir empfehlen folgende Best Practices zur Risikominimierung:

- Analysieren Sie vorhandene Management-Methoden und Systeme.
- Konsolidieren Sie diese in einer zentralen Monitoring-Plattform für sämtliche Systeme an allen Standorten.
- Nutzen Sie Remote-Monitoring, wenn Sie nur über eingeschränkte Ressourcen verfügen. Im White Paper 237, Digital Remote Monitoring and How it Changes Data Center Operations and Maintenance, finden Sie weitere Informationen über den Einsatz von Remote-Monitoring für die Reduzierung von Ausfallzeiten.

Stromversorgung und Kühlung

Infrastruktursysteme für Stromversorgung und Kühlung (z. B. USV- und Klimaanlagen) werden an Edge-Standorten häufig ohne Redundanz installiert. Dadurch entstehen so genannte Single Points of Failure. Außerdem kann eine Systemwartung nicht ohne Betriebsunterbrechung erfolgen. In einigen Fällen sind keine dedizierten Kühlsysteme vorhanden, sodass es zu einer Überhitzung von IT-Komponenten kommen kann. Die Infrastruktursysteme werden häufig gemeinsam mit anderen Anlagen in einem Mehrzweckgebäude genutzt, daher ist die Verfügbarkeit des Edge-Datacenters abhängig von der Verfügbarkeit dieser gemeinsam genutzten Komponenten.

Wir empfehlen folgende Best Practices zur Risikominimierung:

- Messen Sie die Temperatur und Luftfeuchtigkeit, um die Art der erforderlichen Kühung zu bestimmen (z. B. passiver Luftstrom, aktiver Luftstrom oder dedizierte Kühlung).
- Installieren Sie gegebenenfalls redundante Stromkreise, um eine Wartung ohne Betriebsunterbrechung an kritischen Standorten zu ermöglichen.
- Stellen Sie sicher, dass kritische Stromkreise über einen Notfallgenerator versorgt werden.

Abbildung 6 zeigt ein Beispiel für ein Tier 3 Micro-Datacenter, bestehend aus einer vorkonfektionierten, integrierten Lösung in einem 42-HE-Gehäuse mit redundanten USV-Systemen und Energieverteilung.

Vorderseite Rückseite

Abbildung 6
Beispiel für ein Micro-Datacenter mit einem Rack und integrierter Redundanz

Netzwerkkonnektivität

Wie bereits zuvor erwähnt, ist die Netzwerkanbindung an die Cloud ein entscheidender Faktor für Edge-Datacenter. Doch in vielen Fällen wird diese Anbindung mit nur einem Internet-Provider realisiert. Auch das ist ein Single Point of Failure. Zudem ist eine unorganisierte Kabelverlegung eine mögliche Ursache für Handlingsfehler.

Wir empfehlen folgende Best Practices zur Risikominimierung:

- Nutzen Sie einen zusätzlichen Internet-Provider für kritische Standorte.
- Sorgen Sie für eine organisierte Verlegung der Netzwerkkabel mit Komponenten für das Kabelmanagement (Kabelkanäle, Kabelführungen, Befestigungen usw.).
- Die Etikettierung und Kodierung von Netzwerkkabeln verhindert Handlingsfehler.

Fazit

Die zunehmende Verbreitung des Cloud Computing führt dazu, dass immer mehr Unternehmen auf hybride Datacenter-Strukturen – eine Kombination aus Cloud- und On-Premise-Datacentern (im Edge-Bereich) – setzen. Auch wenn die On-Premise-Datacenter häufig nur noch wenig Stellfläche belegen, sind ihre Systeme häufig umso kritischer für die Unternehmen. Die Gründe dafür liegen auf der Hand:

- Weil immer mehr Anwendungen in die Cloud verlagert wurden, ist die Netzwerkanbindung für viele Unternehmen Voraussetzung für einen unterbrechungsfreien Geschäftsbetrieb.
- Darüber hinaus sind moderne Technologien und sichere Netzwerkverbindungen für viele junge Mitarbeiter ein wichtiges Kriterium bei der Wahl ihres Arbeitgebers.

Leider sind viele der Datacenter im Edge-Bereich heute schlecht geplant, sodass teure Ausfallzeiten vorprogrammiert sind. Ein systematischer Ansatz für die Bewertung der Ausfallsicherheit sämtlicher Datacenter in einer Hybrid-Struktur ist unverzichtbar, um sicherzustellen, dass Investitionen dort getätigt werden, wo sie den größten Effekt haben.

Das vorgestellte Scorecard-System ermöglicht den Verantwortlichen eine realistische und ganzheitliche Bewertung aller Faktoren, die auch die Anzahl der Mitarbeiter und die Geschäftsfunktionen der einzelnen Datacenter berücksichtigt. So lassen sich Standorte identifizieren, die vorrangig optimiert werden müssen.

Vorkonfektionierte Micro-Datacenter sind eine unkomplizierte Option für die Bereitstellung sicherer, hochverfügbarer Installationen im Edge-Bereich. Zu den Best Practices gehören auch redundante USV-Systeme, ein gesichertes und organisiertes Rack, gutes Kabelmanagement und optimale Luftverteilung sowie Remote Monitoring und redundante Netzwerkverbindungen. Nur so können Sie sicherstellen, dass diese kritischen Installationen die erforderliche hohe Ausfallsicherheit erreichen.

Über die Autoren

Kevin Brown ist Chief Technology Officer der Data Center Division bei Schneider Electric. Er hat sein Studium an der Cornell University mit einem BS-Abschluss in Maschinenbau abgeschlossen. Vor seinem Wechsel zu Schneider Electric war Kevin Brown Director of Market Development bei Airxchange, einem Hersteller von Energierückgewinnungs-Produkten und -Bauteilen im Bereich Heizung, Klima und Lüftung.
Davor hatte er zahlreiche leitende Management-Funktionen bei Schneider Electric inne, unter anderem als Director Software Development Group und Senior Vice President of Data Center Solutions.

Wendy Torell ist Senior Research Analyst im Schneider Electric Data Center Science Center. In dieser Funktion forscht sie zu Best Practices bei Auslegung und Betrieb von Datacentern, veröffentlicht White Paper und Artikel und entwickelt TradeOff Tools, um Kunden bei der Optimierung der Verfügbarkeit, Effizienz und Kosten ihrer Datacenter-Infrastruktur zu helfen. Darüber hinaus berät sie Kunden zu Ansätzen der Verfügbarkeitskunde und Auslegungstechnik, um sie beim Erreichen der Leistungsziele ihrer Datacenter zu unterstützen. Sie hat ihr Studium am Union College in Schenectady, NY als Bachelor of Mechanical Engineering sowie ein MBA-Studium an der University of Rhode Island abgeschlossen. Wendy Torell ist ASQ Certified Reliability Engineer.

Ressourcen

Cost Advantages of Micro Data Centers
White Paper 223

Die treibende Kraft hinter Edge Computing und die Vorteile der Technologie
White Paper226

Digital Remote Monitoring and How it Changes Data Center Operations and Maintenance
White Paper 237

Alle White Paper anzeigen
whitepapers.apc.com

Alle TradeOff Tools™ anzeigen
tools.apc.com

Kontaktieren Sie uns

Rückmeldungen und Anmerkungen zum Inhalt dieses White Paper:

Data Center Science Center
dcsc@schneider-electric.com

Falls Sie Kunde sind und Fragen zu Ihrem spezifischen Datacenter-Projekt haben:

Wenden Sie sich an Ihre Schneider Electric-Vertretung unter
www.apc.com/support/contact/index.cfm