Best Practice-Leitfaden

Business Continuity für wichtige Anwendungen

Best Practices für Störfälle im Rechenzentrum

Hewlett Packard
Enterprise

Stellen Sie sich vor, dass Sie mit dem Auto auf dem Weg zum Markt sind. Etwas in Ihrem Augenwinkel lenkt Sie für einen Moment ab. Genau in diesem Moment stoppt das Auto vor Ihnen und Sie verursachen einen Auffahrunfall. Das ist schon traumatisch genug, aber was wäre, wenn Sie weder eine Versicherung noch das Geld hätten, um die Schäden zu bezahlen? Was passiert als Nächstes?

Für jeden von uns ist es selbstverständlich, eine Versicherung zu haben, die uns bei Unfällen, Diebstählen oder anderen Verlusten absichert. Bei Unternehmen ist das ähnlich. Manche geben jährlich sehr hohe Summen aus, um Risiken zu beurteilen und zu minimieren. In vielen Branchen spielt für IT-Organisationen bei der Sicherung von Daten vor unberechtigtem Zugriff und Offenlegung das Geld keine Rolle. Bei der Sicherstellung einer kontinuierlichen Verfügbarkeit von Anwendungen, die Daten verarbeiten, oder bei der Vermeidung von Datenverlusten bei System- und Rechenzentrumsausfällen haben manche Unternehmen eine veraltete Vorstellung davon, welche Bereiche immer aktiv sein sollten. Aus diesem Grund zögern diese Unternehmen oft, wenn es darum geht, die erforderlichen Ressourcen für ein ausreichendes Maß an Business Continuity aufzuwenden.

Laut Schätzungen von IDC betragen die durchschnittlichen Kosten durch Ausfallzeiten in etwa 1,7 Mio. Dollar pro Stunde, manche Ausfällen verursachen sogar bis zu 10 Mio. Dollar pro Stunde. Im Durchschnitt beträgt die Ausfallzeit 90 Minuten, in manchen Fällen mehr als 24 Stunden.[3]

Die Risiken sind real

Stellen Sie sich folgende Frage: Welche Umsatzeinbußen ergeben sich aus dem Ausfall eines Systems auf Kundenseite? Welche Auswirkungen hat der Verlust von aktiven Kundentransaktionen auf das Unternehmen und Ihre Kunden? Welche Kosten muss Ihr Unternehmen durch Produktivitätsverluste hinnehmen, wenn Ihr E-Mail- oder UCC-System (Unified Communications and Collaboration) ausfällt? Was passiert, wenn ein Feuer im Rechenzentrum ausbricht, die Stromversorgung unterbrochen wird oder ein Erdbeben Schäden verursacht und das Rechenzentrum lahmlegt? Was passiert als Nächstes?

Diese Dinge passieren öfter, als Sie annehmen. 95 % der Unternehmen haben in den letzten 24 Monaten mindestens einen ungeplanten Rechenzentrumsausfall erlebt – nicht nur bei einzelnen Systemen, sondern bei ganzen Rechenzentren. Im Durchschnitt hat ein Unternehmen in der Finanzbranche in den letzten 24 Monaten 1,8 vollständige Rechenzentrumsausfälle erlebt. Im Gesundheitswesen liegt der Durchschnitt der letzten 24 Monate bei 3 Ausfällen.[1]

Laut Schätzungen von IDC betragen die durchschnittlichen Kosten durch Ausfallzeiten in etwa 1,7 Mio. USD pro Stunde, manche Ausfälle verursachen sogar Kosten von bis zu 10 Mio. USD pro Stunde.[2] Im Durchschnitt beträgt die Ausfallzeit 90 Minuten, in manchen Fällen mehr als 24 Stunden. Neben den Umsatzeinbußen führen Ausfallzeiten oft auch zu Imageverlusten, Vertrauensverlust beim Kunden, Rückgang der Kundenbindung, Verlust von Wettbewerbsvorteilen und möglicherweise auch Haftungsfällen durch Nichteinhaltung gesetzlicher Auflagen. In der heutigen vernetzten Gesellschaft dauert es nicht lange, bis sich die Nachrichten über den Ausfall verbreiten. Es kann Jahre dauern, bis der Schaden behoben ist.

Wiederherstellungsziele

Welcher Zeitrahmen ist für eine Wiederherstellung nach dem Ausfall eines Anwendungsservices akzeptabel? Diese Recovery Time Objectives (RTO) sind von Anwendung zu Anwendung unterschiedlich: fünf Sekunden, fünf Minuten oder fünf Tage. Die Kosten, die sich durch Ausfälle ergeben, sind meistens ausschlaggebend für RTOs. Manche Apps haben eine RTO von null, d. h. Kunden und Endbenutzer sollten den Ausfall gar nicht bemerken.

Welche Datenmenge darf bei einem Ausfall eines Anwendungsservices maximal verloren gehen? Diese Recovery Point Objective (RPO) kann anhand der Art und Wichtigkeit der Daten oder anhand ihres Werts für das Unternehmen (d. h. Kosten durch verlorene Umsätze und Haftungsansprüche) oder eine Kombination beider Faktoren festgelegt werden.

[1] „Fingers Crossed? Or What is Your Business Continuity Plan for the Inevitable," Gravic, Inc., 2015 (ursprüngliche Quelle: Ponemon Institute)

[2, 3] **„High-Value Business Applications on x86: The Need for True Fault-Tolerant Systems**," Peter Rutten, IDC, Mai 2015

RTO: die maximal akzeptierbare Wiederherstellungszeit nach einem Ausfall

RPO: das Maximum an Daten, das bei einem Ausfall verloren gehen darf

RPOs basieren oft auf dem durchschnittlichen Wert einer verlorenen Transaktion. Grundsätzlich mag das logisch erscheinen, betrachtet man es aber genauer, ergibt dieser Ansatz nicht immer Sinn. Ein Beispiel: Bei einem Finanzunternehmen beträgt der durchschnittliche elektronische Zahlungsverkehr 1.000 USD, eine einzelne Überweisung kann aber mehrere Millionen Dollar betragen.

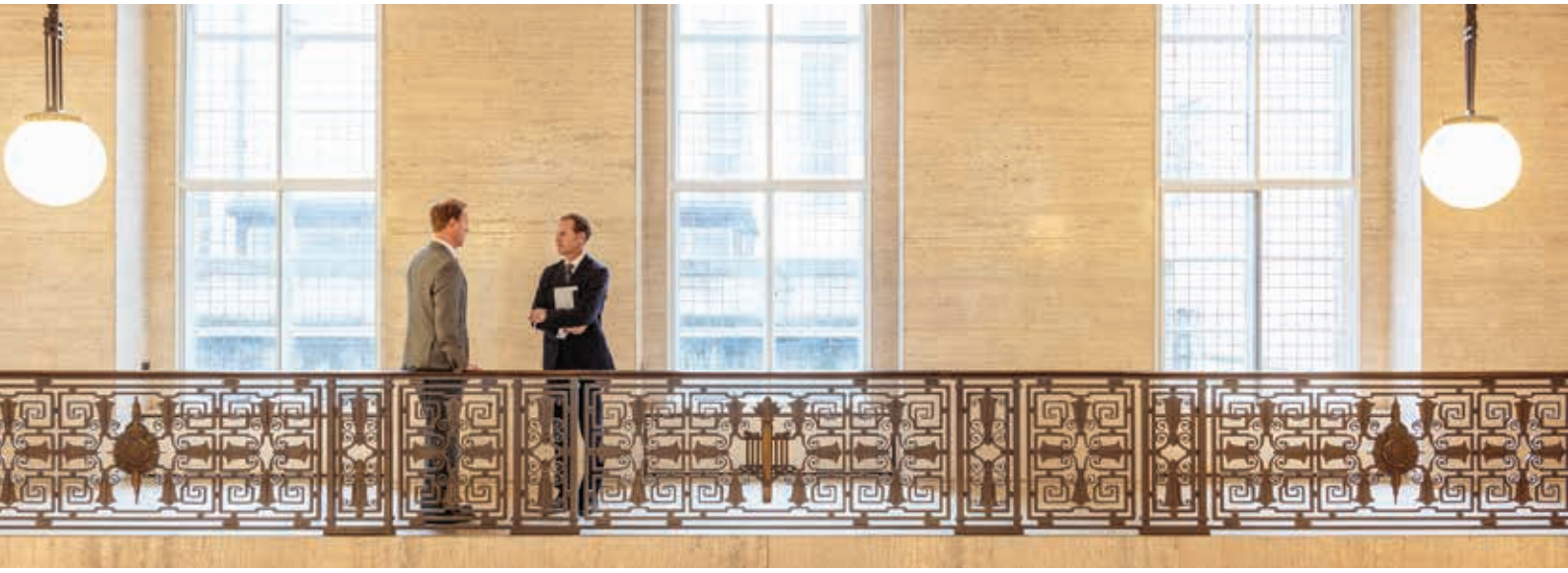
Da nicht vorhersehbar ist, bei welchen Transaktionen Probleme auftreten, entsprechen die tatsächlichen Kosten eines Ausfalls dem Wert der wertvollsten Daten, die verloren gehen können (z. B. die größte elektronissche Überweisung, der größte Auftrag, die Handelsgeschäfte mit dem größten Volumen, die höchsten Haftungsansprüche aufgrund verlorener Daten oder das größte Konto). Dieser Wert sollte die RPO-Ziele bestimmen.

Business Continuity

Zu Beginn steht meistens eine Diskussion über verlorene Transaktionen, verlorene Umsätze oder Produktivitätsverluste. Diese Probleme veranlassen Unternehmen dazu, Business Continuity in ihren strategischen Plan einzubeziehen. Die Frage ist nicht, ob Systeme ausfallen werden oder ob ein Katastrophenereignis eintritt, sondern wann. Dann stellt sich die Frage: Was passiert als Nächstes? Durch Business Continuity wird sichergestellt, dass es in Ihrem Unternehmen in keiner Situation zu Betriebsstörungen kommt.

In der Praxis sind für Business Continuity die zulässigen RTOs und RPOs für jedes System, auf dem wichtige Anwendungen und ihre Daten gespeichert oder unterstützt werden, zu überprüfen (gemäß den Best Practices mit einer erneuten Überprüfung und Neubewertung alle ein bis zwei Jahre). In manchen Fällen kann die Festlegung sehr einfach sein. Ein System mit Patientenakten muss beispielsweise ständig verfügbar sein und Datenverluste sind nicht tragbar. Das Leben der Patienten hängt davon ab. Bei einem Ausfall muss das System sofort wiederhergestellt werden können, sodass die Benutzer den Ausfall gar nicht wahrnehmen.

In anderen Fällen kann die Festlegung von zulässigen RTOs und RPOs etwas schwieriger sein. Welchen Wert haben beispielsweise Ihr Online-Store, Ihr VoIP-Telefon oder CRM-System? Was geschieht bei einem Ausfall? Ist es unternehmenskritisch oder geschäftskritisch? Könnte ein Großteil Ihres Unternehmens weiterarbeiten, wenn diese Systeme ausfallen? Welcher Zeitrahmen für die Wiederherstellung von Daten wäre tolerierbar? Welche Auswirkungen hätte es, wenn die Kunden Sie nicht erreichen könnten? Dieses Art der Überprüfung zeigt oft, dass viele Systeme wichtiger für Unternehmen sind, als Sie denken. Diese Systeme erfordern mehr Unterstützung für Business Continuity mit strengeren RTOs und RPOs. Genauer betrachtet, können sich diese Anwendungen oder Services sogar als unternehmenskritisch herausstellen. Das heißt. dass jeder Ausfall starke Auswirkungen haben wird. Bedenken Sie, dass Unwichtiges morgen schon unternehmenskritisch sein kann.



Was ist „kritisch“?

Wir haben festgestellt, dass „geschäftskritische“ Anwendungen und Daten für einen reibungslosen Betrieb des Unternehmens wichtig sind, während „unternehmenskritische“ Anwendungen und Daten so wertvoll sind, dass ein Ausfall einer Katastrophe gleichkommt. Bedenken Sie Folgendes: Welche Auswirkungen hätte der Ausfall unternehmenskritischer Systeme oder der Verlust von Transaktionsdaten? Welche Auswirkungen hat dies für Ihre Kunden? Und auf den Umsatz Ihres Unternehmens? Ergeben sich aus dem Verlust von Daten mögliche rechtliche oder Compliance-Probleme?

Anwendungen, für die ein Ausfall entweder nicht tragbar ist oder die so schnell wieder funktionieren müssen, dass der Ausfall niemandem auffällt, werden als unternehmenskritisch angesehen.

Die Unterstützung für Business Continuity muss eine Reihe unternehmenskritischer und grundlegender Anwendungen und Daten umfassen. Die Bewertung Ihrer Anwendungen und Systeme auf dieser Skala sollte auf den Anforderungen Ihrer Kunden und Ihren Umsatzzielen basieren. Eine Anwendung mit einer geringen Ausfalltoleranz (RTO von null oder fast null) oder einer geringen Toleranz bezüglich der Menge verlorener Daten (RPO von null oder fast null) sollte als unternehmenskritisch gesehen werden.

Beispiele für wichtige Anwendungen und Services sind:

- **Finanzdienstleistungen:** Zahlungsabwicklung, Betrugsprävention, Hochfrequenzhandel
- **Telekommunikation:** Verwaltung mobiler Netzwerke; Machine-to-Machine, Kundenservice in Echtzeit
- **Einzelhandel:** Point-of-Sale, E-Commerce, Online-Transaktionen und Auftragsbearbeitung
- **Fertigung:** Prozesse für die kontinuierliche Produktionssteuerung und Vertrieb über mehrere Kanäle
- **Gesundheitswesen:** Patienten- und Labordaten in Echtzeit; Informationsgewinnung vom Anbieter
- **Transportwesen:** Reservierungen; Kartenverkauf; Disposition

In einer Welt, in der ständige Verfügbarkeit wichtig ist, sind Ausfallzeiten bei komplexen, vernetzten, kundenorientierten Workloads üblicherweise keine Option.

Ansätze für Business Continuity

Echte Business Continuity erfordert ein bestimmtes Maß an geografischer Verteilung (oder Distanz), um sowohl lokale Ereignisse (z. B. Feuer im Rechenzentrum) als auch regionale Ausfälle (z. B. regionale Stromausfälle) zu überstehen. Es gibt drei grundlegende Ansätze, um eine geografisch verteilte Business Continuity-Infrastruktur zu erstellen, jeweils mit unterschiedlichen RTO- und RPO-Profilen:

- **Asynchron Aktiv/Passiv:** Hierbei handelt es sich um ein klassisches Disaster Recovery (DR)-Szenario, bei dem alle Transaktionen auf einem aktiven System ausgeführt und die Daten asynchron auf einen passiven Sicherungsknoten repliziert werden. Bei einem Ausfall müssen die Anwendungen auf dem Sicherungsknoten gestartet werden, was zu Verzögerungen und einer längeren RTO führen kann. Failover-Prozesse zum Start dieser Anwendungen sind oft schwierig zu testen oder auszuführen und das Fehlerpotenzial ist hoch.
- **Asynchron Aktiv/Fast Aktiv:** Wird auch als Sizzling-Hot-Takeover (SZT) oder Sizzling-Hot-Standby bezeichnet. Dieser Ansatz ähnelt einer Aktiv/Passiv-Architektur und nutzt die Replikation. Der Unterschied ist, dass der Sicherungsknoten sofort bereit für die Verarbeitung von Transaktionen ist, da die lokale Kopie der Anwendungsdatenbank bereits mit Lese-/Schreibberechtigung geöffnet ist. Im Grunde ist es eine Aktiv/Aktiv-Architektur, nur werden hier alle Benutzertransaktionen auf den primären Knoten geleitet. So erhöhen sich die Erfolgschancen im Falle eines Failovers und die RTO wird im Vergleich zur klassischen DR optimiert und wiederholbar.
- **Asynchron Aktiv/Aktiv:** In einer ausfallsicheren Architektur ist die Produktionsabwicklung auf mehrere Knoten verteilt. Jeder Knoten verfügt über eine Kopie der Datenbank, die über die bidirektionale Datenbankreplikation synchronisiert wird. Wenn ein Knoten ausfällt, wird der Datenverkehr automatisch auf andere aktive Knoten umgeleitet. Benutzer, die mit den übrigen Knoten verbunden sind, bemerken den Ausfall nicht. Failover wird zu einem einfachen Prozess, der leicht getestet und ausgeführt werden kann, da bekannt ist, dass alle Knoten ständig funktionieren.

Für unternehmenskritische Anwendungen ist eine ausfallsichere Architektur mit mehreren Knoten die beste Alternative und bietet die beste RTO und RPO. Natürlich gibt es eine Reihe von Technologien, die bei der Erstellung von Business Continuity-Lösungen helfen: von der softwarebasierten Replikation von Transaktionsdaten bis hin zu hardwarebasierten Clustering- und RAID-Technologien. Aber jede hat ihre Grenzen in Bezug auf die RTO- und RPO-Funktionen.

Eine weitere Überlegung wäre Atomarität, eines der vier AKID-Konzepte (Atomarität, Konsistenz, Isolation und Dauerhaftigkeit) für die Datenbankenentwicklung und die Anwendungsarchitektur. Atomarität bei Transaktionen bedeutet alles oder nichts: Eine Transaktion beginnt zu einem bestimmten Zeitpunkt und wenn vor dem vollständigen Abschluss der Transaktion ein Fehler auftritt, wird alles zurückgesetzt, als hätte die Transaktion nie stattgefunden. Business Continuity-Ansätze sollten mit diesem Prinzip im Hinterkopf entwickelt werden, vor allem für unternehmenskritische Anwendungen.

Gesamtbetriebskosten

Sollte auch nur ein Problem, ein Ereignis oder eine Katastrophe auftreten, die zu einem längeren Ausfall führen, als es für Ihre Kunden akzeptabel ist, sollten Sie über eine ausfalltolerante Architektur verfügen, damit Ihnen keine Umsätze entgehen und Ihr Ruf nicht leidet. Im Zeitalter der sozialen Medien können Kunden ihren Ärger sofort kundtun, und innerhalb von kürzester Zeit wissen alle über das Problem Bescheid. In Unternehmen, die strenge Richtlinien in Bezug auf Ausfallzeiten haben, kann eine ausfallsichere Architektur die gesetzlichen Bußgelder und Berichte eingrenzen. Wenn ein System beispielsweise erst nach drei Stunden wieder vollständig hergestellt ist, wären es unter Annahme der durchschnittlichen Ausfallzeitkosten laut der bereits erwähnten IDC-Studie mehr als 5 Mio. USD. In einer gut strukturierten Business Continuity-Umgebung kann der gleiche Ausfall nur wenige Sekunden dauern oder sogar unbemerkt bleiben und dadurch weit weniger Kosten verursachen.

Bei klassischen Aktiv/Passiv-DR-Architekturen sind die Gesamtbetriebskosten hoch, da durch Ausfallzeiten hohe Kosten entstehen.
Allgemein gilt:

- Je höher die Verfügbarkeit, desto höher die Komplexität und Implementierungskosten
- Je höher die Verfügbarkeit, desto geringer die Kosten durch Ausfälle

Während die Implementierungskosten steigen, sinken die Gesamtbetriebskosten – und das erheblich schneller.[4] Anders gesagt, mit dem Geld, das ein einziger Ausfall kostet, könnten Sie eine umfassende Fehlertoleranz erwerben.

Business Continuity-Angebote

HPE Services
Hewlett Packard Enterprise hilft Kunden mit geschäftskritischen Rechenzentren durch qualifizierte und erfahrene Experten, die eine umfangreiche Reihe von Services in den Bereichen Beratung, Entwicklung, Bereitstellung und Verwaltung von ausfallsicheren Architekturen bieten.

[4] „Fingers Crossed? Or What is Your Business Continuity Plan for the Inevitable," Gravic, Inc., 2015 (ursprüngliche Quelle: Ponemon Institute)

HPE Integrity NonStop X

HPE Integrity NonStop X-Systeme wurden speziell für Branchen entwickelt, in denen rund um die Uhr gearbeitet wird und die über das höchste Maß an Verfügbarkeit, systemweite Sicherheit, enorme Skalierbarkeit und die branchenweit niedrigsten Gesamtbetriebskosten verfügen. Das höchste Availability Level 4 (AL 4) von IDC ist eine Kombination aus mehreren Hardware- und Softwarekomponenten, die ein nahezu umgehendes Failover auf alternative Ressourcen ermöglichen, damit die Geschäftsprozesse ohne Unterbrechung weiterlaufen.[5] Die Verbindung von HPE Integrity NonStop X und HPE NonStop Shadowbase Software ermöglicht eine AL 4 Fehlertoleranz über geografische Standorte hinweg und somit eine einzigartige konti-nuierliche Verfügbarkeit ohne geplante oder ungeplante Ausfallzeiten. So können Anwendungsausfälle quasi ausgeschlossen werden. HPE NonStop X bietet eine neue Art von unternehmenskritischem Computing, das speziell auf eine maximale Verfügbarkeit, Skalierbarkeit und Dateninte-grität ausgelegt ist.

HPE XP Storage

HPE XP7 Storage wurde für hybriden Flash-Speicher entwickelt und ist ideal für unternehmenskritische Anwendungen, die eine kontinuierliche Datenverfügbarkeit, Skalierbarkeit und Leistung erfordern. Eine arraybasierte Virtualisierungstechnologie ermöglicht die Virtualisierung, Replikation und Verwaltung über mehrere Standorte und Arrays hinweg, um die Verfügbarkeit zu steigern, Katastrophen zu vermeiden und die Ressourcennutzung zu optimieren, indem Speichersilos vermieden werden. HPE XP7 Storage ist das SAN-Array mit der höchsten Verfügbarkeit, das HPE mit einer Reihe von Softwarelösungen anbietet, durch die die anspruchsvollsten Wiederherstellungsziele sowie umfassende Fernreplikation und DR-Funktionen erreicht werden können.

[5] „Worldwide and U.S. High-Availability Server 2014–2018 Forecast and Analysis," IDC, Doc #250565

Fazit

Die Frage ist nicht, ob eine Katastrophe unternehmenskritische Systeme beeinträchtigen wird, sondern wann es passieren wird. Wenn der Katastrophenfall eintritt, was dann? Eine Business Continuity-Lösung, die auf kontinuierliche Ausfalltoleranz ausgerichtet ist, hilft dabei, den Schaden zu minimieren und bietet die besten RTOs und RPOs bei Gesamtbetriebskosten, die Ihrem Unternehmen entsprechen.[6]

Weitere Informationen unter
hpe.com/info/nonstop
hpe.com/storage/xp
hpe.com/info/dcm

[6] „**High-Value Business Applications on x86: The Need for True Fault-Tolerant Systems**," Peter Rutten, IDC, Mai 2015

Melden Sie sich noch heute an.

4AA6-5326DEE, Mai 2016

Hewlett Packard Enterprise