

Backup und Archivierung

Datensicherung: Verschiedene Stufen der Paranoia

Ein Unternehmen, das eine Gebäudebrandversicherung abschließt, muss sich fragen lassen, was es zum Schutz seiner Daten gegen Feuer unternommen hat. Offensichtlich ist ja die Wahrscheinlichkeit eines Feuers gegeben. Wenn man sich dieser Frage stellt, führt dies zwangsläufig dazu, dass man Daten einteilen muss zwischen unternehmenskritisch, schützenswert und weniger wichtig. Die nächste Frage ist: Wie wahrscheinlich ist die Katastrophe und wie teuer wäre der Verlust dieser Daten? Diese Einschätzung und die daraus resultierenden Maßnahmen werden im Folgenden als Level Of Paranoia (LoP) bezeichnet.

Risiken einstufen

Um zu einer Einschätzung gelangen, lassen sich zwei gegensätzliche Szenarien vergleichen. Die Variante eines sehr niedrigen LoP könnte so aussehen, dass die Daten auf einer ungespiegelten Festplatte liegen. Jede Nacht wird eine inkrementelle Sicherung auf Band durchgeführt und am Wochenende eine Vollsicherung. Für die Bereitstellung der Daten entstehen in diesem Szenario sehr geringe Kosten. Falls die Festplatte jedoch ausfällt, gehen alle Änderungen seit der letzten Sicherung verloren. Wenn das Risiko als gering und wenn die zu schützenden Daten als weniger wichtig eingestuft werden, ist die Lösung vertretbar. Anders im Fall eines sehr hohen LoP: hier werden die Daten auf einem hochverfügbaren Storage gehalten. Es wird eine lokale Spiegelung vorgenommen, um sicherzustellen, dass der Verlust einer Festplatte nicht zum Datenverlust führt. Um der Möglichkeit vorzubeugen, dass das gesamte Storage-System zerstört wird, werden die Daten nochmals in eine andere Lokation gespiegelt. Mit Hilfe von CDP (Continuous Data Protection) werden alle Änderungen in Echtzeit protokolliert, so dass bei einem logischen Fehler auf die aktuellste intakte Version zurückgegriffen werden kann. Für den Fall des Versagens aller Mechanismen wird eine Sicherungskopie auf Band erzeugt und in einem Tresor verwahrt. Damit auf die Daten unterbrechungsfrei zugegriffen werden kann, wird der zugehörige Server lokal geclustert. Beim Ausfall eines Servers kann so ein anderer automatisch einspringen. Zusätzlich findet ein Remote Clustering statt, um zu verhindern, dass der Ausfall einer kompletten Lokation den Zugriff auf die Daten verhindern könnte.

Das richtige Modell

Beide Modelle haben ihre Berechtigung. Wichtig ist zunächst eine Bewertung der Risiken, um eine Entscheidung für eine Absicherungsstrategie treffen zu können. Die wichtigste Kennzahl, die in diesem Zusammenhang für jede Kategorie von Daten gefunden werden muss, ist die Recovery Time Objective (RTO), die festlegt, wie lange es dauern darf, bis bestimmte Daten wiederhergestellt werden können. In der Praxis sind hier Werte von wenigen Minuten bis zu mehreren Tagen möglich. Eine weitere Kennzahl ist die Recovery Point Objective (RPO), die den Zeitraum zwischen zwei Datensicherungen definiert. Im Rahmen von File-Server-Daten sind 24 Stunden, also eine Sicherung pro Tag, üblich. Bei intensiver täglicher Datenverarbeitung ist dieser Zeitrahmen jedoch zu gering.

Um die individuellen Sicherheitsanforderungen zu erfüllen, muss im Vorfeld die Anzahl der nötigen Sicherungen genau definiert werden. Auch die Datensicherungslösung selbst muss untersucht werden. Ressourcen, die für eventuelle Recoveries zur Verfügung stehen sollen, müssen eingeplant werden. Außerdem ist zu klären, ob für ein wichtiges Recovery gegebenenfalls ebenso wichtige Backups abgebrochen werden dürfen. Entscheidend ist, wie lange ein Unternehmen ohne Backup auskommen kann und so die Verfügbarkeit des Backup-Systems zu bestimmen. Hier spielt vor allem die Sicherung der Log-Dateien wichtiger Datenbanken eine Rolle. Wenn das Filesystem für die Log-Dateien voll läuft, bleibt die Datenbank stehen. Diese Überlegung führt zur Auslegung des Backup-Servers. In einigen Fällen ist es nötig, hierfür ein Cluster-System vorzusehen, um den Backup-Service für diese Dateien mit einer möglichst hohen Verfügbarkeit einzurichten. Die Auslegung der Datensicherungslösung hat auch Einfluss auf die Antwortzeiten der produktiven Systeme. Während des Backups werden Server-Ressourcen gebunden, die somit für operative Prozesse nicht zur Verfügung stehen, weshalb die Backups oftmals auf die Nachtzeiten verlegt werden. Da zunehmend Applikationen ganztägig im Einsatz sind, wird es immer schwieriger, freie Zeiträume für Backups zu bestimmen. All diese situationsabhängigen Faktoren sind ausschlaggebend für die Wahl des Backup-Modells.

Definition von SLAs

Nach der Untersuchung der Einflussfaktoren können Service Level Agreements (SLAs) vereinbart werden. In diesen wird festgeschrieben, welche Werte den einzelnen Kennzahlen zugewiesen werden. Oft werden dabei unterschiedliche Tier-Level festgelegt, in die dann die verschiedenen Anwendungen eingeteilt werden. Für jeden Tier-Level werden die entsprechenden Kennzahlen quantifiziert und somit bestimmt, wie schützenswert die Daten dieser Kategorie sind. Darüber hinaus können bei der Bestimmung der SLAs auch Szenarien wie das Rolling Disaster sowie die Bestimmung der Single Points of Failure (SPoF) berücksichtigt werden. Das Rolling Disaster ist ein Spezialfall des LoP. Hierbei wird davon ausgegangen, dass ein Disaster nicht zu einem bestimmten Zeitpunkt eintritt, wie etwa bei einem Headcrash, sondern dass ein Ereignis ein weiteres nach sich zieht. Erst das Ende der Ereigniskette begründet dann das eigentliche Disaster. Mögliche Beispiele wären ein Schreibfehler auf einer Komponente eines RAID, der sich auf das gesamte RAID auswirkt oder der Ausfall einer CPU in einem Mehrprozessorsystem, der zu einer Mehrbelastung der übrigen CPUs führt, die dann sukzessive ausfallen. Vor einem Rolling Disaster kann auch die Vermeidung von SPoF nicht schützen. Trotzdem ist die SPoF-Analyse ein wichtiger Baustein bei der Entwicklung von Disaster-Recovery-Lösungen. Dabei werden die wichtigen Komponenten eines Systems redundant ausgelegt. Durch die Vermeidung von SPoF kann die Notwendigkeit zur Durchführung von Disaster Recoveries deutlich reduziert werden.

Störungsquellen

Die Daten und ihre Verfügbarkeit sind unterschiedlichen Gefahren krimineller und natürlicher Ursachen ausgesetzt. Die einfachste natürliche Katastrophe ist der Stromausfall. Je nach LoP gibt es auch hier verschiedene Methoden sich dagegen zu schützen. Eine Unterbrechungsfreie Stromversorgung (USV) ermöglicht bei einem Stromausfall ein geordnetes Herunterfahren der Server. Eine weitere Variante ist die Einführung zwei getrennter Stromversorgungen im Rechenzentrum oder die Bereitstellung eines Notstromaggregates, das eine autarke Stromversorgung gewährleistet. Je nach SLA wird man sich für eine dieser Lösungen oder eine Kombination entscheiden.

Auch im Zusammenhang mit dem Headcrash lässt sich der Begriff LoP erläutern: Nahezu alle Server-Systeme sind heute durch RAID-Technologie vor dem Ausfall einer Komponente geschützt. Wie wahrscheinlich ist es wohl, dass in einem RAID-Verbund zwei Festplatten gleichzeitig ausfallen, und es damit zu einem Datenverlust kommt? Trotzdem existiert RAID6 am Markt, das genau vor diesem Fall schützt und zudem erschwinglich ist.

Ein Server-Ausfall ist ein weiteres Szenario, das man im Sinne der LoP bewerten und vor dem man sich gegebenenfalls schützen muss. Hier besteht die Möglichkeit, einen Cluster zu installieren, so dass beim Ausfall eines Servers automatisch der andere Server des Clusters dessen Funktion übernimmt. Ein Standby Server wäre die nächste Stufe der LoP-Pyramide. Hierbei wird ein kurzer Ausfall in Kauf genommen, um die Kosten der Cluster-Lösung zu sparen. Die unterste Stufe bestünde in

einem Rahmenvertrag mit einem Server-Hersteller: Durch die Vereinbarung garantierter Lieferzeiten kann eine überschaubare Ausfallzeit gewährleistet werden.

Durch Leitungsausfälle im WAN aber auch dem Ausfall einzelner Netzwerkkomponenten im LAN kann der Zugriff auf wichtige Daten behindert werden. Die einzige Möglichkeit, hier Vorsorge zu treffen, besteht darin, alle Verbindungen einer SPoF-Analyse zu unterziehen. Für kritische Daten ist dies oftmals die aufwändigste und kostspieligste Komponente in einer umfassenden Strategie zur Katastrophenvermeidung.

Auch Anwenderfehler können in Ausnahmefällen einen Ausfall hervorrufen. Beispiele sind das irrtümliche Löschen einer Server-Partition, das versehentliche Betätigen des Not-Aus-Schalters im Rechenzentrum oder das falsche Beschalten im Patch-Feld für die Netzwerkverbindungen. Es ist nahezu unmöglich, alle Fälle zu verhindern, doch wenn es tatsächlich zu einem Datenverlust gekommen ist, gibt es verschiedene Möglichkeiten, die verlorenen Daten wiederherzustellen.

Datensicherungsansätze

Das klassische Backup auf Bandlaufwerken ist eine Option, wenn die Struktur der zu sichernden Daten diesem Ansatz entgegenkommt. Bei einer großen Anzahl monolithischer Daten, die leicht parallelisiert werden können, ist die Performance von modernen Bandlaufwerken kaum zu schlagen. 25 LTO-4 Laufwerke, die optimal ausgelastet werden, erreichen einen Datendurchsatz von etwa 3GB/s. Damit können bis zu 10 TB in einer Stunde gesichert werden. Auch die Möglichkeit spezielle Daten in einen gesicherten Bereich auszulagern, wird oft als Argument für eine Bandlösung angeführt. Alternative Lösungen führen das Backup auf Disk-basiertem Storage durch. Der Reiz dieses Verfahrens besteht darin, dass bei der Wiederherstellung der Daten keine Rüst- und Positionierungszeiten anfallen. Recoveries sind meist bedeutend schneller und erfordern weniger administrativen Aufwand als Bandlösungen.

Ein Sonderfall der Disk-basierenden Backup-Lösungen ist die Sicherung mit Hilfe von Snapshots. Snapshots verkürzen die Wiederherstellungszeit nochmals erheblich. Bei dieser Technik werden jedoch die produktiven Storage-Systeme durch das Backup belastet. Da sich der Snap immer aus dem Original und den dazugehörigen Änderungen zusammensetzt, werden die Originaldaten für den Restore benötigt. Um einen Restore durchzuführen, müssen sowohl die Originale, als auch die Snap-Daten gelesen werden. Wenn in den Originalen ein Defekt auftritt, ist keine Wiederherstellung aus einem Snap mehr möglich. Eine Weiterentwicklung der Snap-Technologie stellt die Continuous Data Protection (CDP) dar. Hierbei werden Änderungen, die auf den produktiven Storage geschrieben werden, in einem sekundären Storage-System fortgeschrieben. Diese Technik ermöglicht eine Wiederherstellung für jeden beliebigen Zeitpunkt, womit es möglich ist, sensible Dateisysteme mit sehr hohen SLAs effektiv abzusichern. Eine ganz andere Richtung schlägt ein neuer Backup-Ansatz ein. Der Fokus ist, möglichst effektiv mit dem Backup-Storage umzugehen und zusätzlich die weiteren Komponenten des Backup-Prozesses zu entlasten. Erreicht wird dies durch Deduplikation der zu sichernden Daten. Am Markt haben sich zwei Konzepte etabliert, die unterschiedliche Zielrichtungen aufweisen. Das erste ist die Target-based-Deduplication, bei der die Dateneduplizierung auf dem Backup-Medium stattfindet. Bei diesem System wird tatsächlich nur der Platz für die gesicherten Daten reduziert, indem redundante Blöcke erkannt und durch Referenzen auf bereits existierende Daten ersetzt werden. Bei der Source-based-Deduplication wird nahezu das gleiche Verfahren verwendet. Allerdings wird die Deduplizierung direkt an der Quelle vorgenommen, was zusätzlich zur Verringerung des Backup-Speichers eine Entlastung des Netzwerks, sowie eine Verringerung der CPU-Last auf den Servern bewirkt.

Disaster Recovery

Während die Datensicherung die Wiederherstellung der unternehmenskritischen Daten als Ziel hat, fokussiert das Disaster Recovery die Wiederherstellung der darunterliegenden Systeme.

In diesem Zusammenhang wird oft von Bare Metal Recovery (BMR) gesprochen. Dieser Begriff ist irreführend, da die exakte Wiederherstellung eines Systems nur auf identischer Hardware problemlos möglich ist. Oft ist der Systemausfall jedoch durch den Ausfall einer Komponente bedingt, die nicht mehr in der ursprünglichen Form verfügbar ist. Da viele BMR-Lösungen auf dem Prinzip eines Image-Backup basieren, in dem die System-Partition als Image gespeichert wird, ist es verhältnismäßig aufwändig, das alte Image an die neue Umgebung anzupassen. Alternativ dazu existiert die Möglichkeit einer Profilsicherung. Bei diesem Verfahren werden wichtige Kennzahlen des Systems in einer Datenbank verwaltet, so dass sie beim Wiederherstellen individuell auf das neue System angewendet werden können. Diese Lösung besteht aus zwei Komponenten: einem Image des Betriebssystems, das in Hardware-unabhängiger Form vorliegt, sowie einem Profil aller relevanten Parameter, die das wiederherzustellende System ausmachen. Tritt beim Disaster Recovery eine Hardware-Änderung auf, werden die Parameter, die für das neue System nicht mehr zutreffen, einfach nicht mehr zurückgespielt.

Zusammenfassung

Wie man sich im Rahmen der LoP auf ein konkretes Disaster vorbereitet, ist ein sehr individueller Prozess, dessen Ergebnisse nicht für alle Unternehmensbereiche gleich ausfallen können. Selbst innerhalb eines Unternehmensbereiches wird es verschiedene Kategorien geben, die vollkommen unterschiedlich zu bewerten sind. Wichtig sind drei Kernfragen: Wie wahrscheinlich ist ein spezielles Disaster-Szenario? Wie teuer wäre es, wenn das Szenario eintritt? Was kostet es, sich vor diesem Disaster zu schützen? Die weiteren Schritte ergeben sich aus den Antworten. Um zu ihnen zu gelangen, müssen diejenigen befragt werden, die von einem eventuellen Disaster betroffen wären. Dies ist ein langwieriger Weg. Doch am Ende steht die Sicherheit, sich für jede Situation gewappnet zu haben.



EMC Deutschland GmbH
Am Kronberger Hang 2a
65824 Schwalbach/Taunus
06196 / 4728 0
www.emc2.de

Machen Sie den nächsten Schritt:

Die Vorbereitung auf ein konkretes Disaster im Rahmen der LoP ist ein sehr individueller Prozess. Informieren Sie sich, wie EMC Ihnen dabei helfen kann.