Sandro Eggenberger

eBackup versus konventionelles Backup in Bezug auf das Datenwachstum eines Terrabyte bei einer Kostenreduktion von 50%

Bachelorarbeit

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de/ abrufbar.

Impressum:

Copyright © 2014 GRIN Verlag
ISBN: 9783656894995

Dieses Buch bei GRIN:

https://www.grin.com/document/289210

Sandro Eggenberger

eBackup versus konventionelles Backup in Bezug auf das Datenwachstum eines Terrabyte bei einer Kostenreduktion von 50%

GRIN Verlag

Gegenüberstellung eines eBackups versus eines konventionellen Backups in Bezug auf das Datenwachstum von einem Terrabyte bei einer Kostenreduktion von 50%

Semesterarbeit II Thesis zur Erlangung des akademischen Grades:

Bachelor of Science FH in Business Information Technology an der Kalaidos Fachhochschule Schweiz

Vorgelegt von: Sandro Eggenberger

Fachrichtung: Business Information Technology

Datum der Abgabe: 06.01.2014

Inhaltsverzeichnis

Vorwort

Es ist mir eine grosse Freude, dass Sie meine Arbeit in den Händen halten. Offenbar stösst die von mir gewählte Thematik auf Interesse: Elektronisches Backup oder Backup as a Service ist aktuell, jedoch bei den IT-Diskussionen noch nicht omnipräsent.

In den vergangen Monaten habe ich mich intensiv mit dem Thema befasst. Es war eine lehrreiche, spannende und inspirierende Zeit für mich

Ein besonderer Dank für die Unterstützung geht dabei an folgende Personen: Edith Hinder, Silvio Corti und an alle Beteiligten, die mir bei der Fertigstellung der Arbeit tatkräftig zur Seite standen.

Nun wünsche ich eine interessante und aufschlussreiche Lektüre.

Zürich, im Januar 2014

Sandro Eggenberger

Abstract

Der Stellenwert in Bezug auf die Datensicherheit hat sich in den letzten Jahren massiv verändert. Ein Datenverlust oder ein kompletter Datenverlust kann und will sich keine Firma mehr leisten.

Dank neuer Technologien wie Deduplizierung gibt es heute die Möglichkeit, Daten als elektronisches Backup beziehungsweise Backup as a Service zu beziehen. Bei einem Cloud Service muss neben der Datensicherheit auch das Datenschutzgesetz berücksichtigt werden. Ziel dieser Arbeit ist es, anhand ausgewählter Fachliteratur beantworten zu können, ob mittels eines Datensicherungskonzepts ein elektronischen Backup, eine Kostenreduktion von 50% gegenüber einem Konventionell Backup ermöglicht werden kann, dabei darf die Datenmenge von einem Terrabyte nicht übersteigt werden.

Laut Nelson ist ein elektronisches Backup nur unter einer Datenmenge von 200 Gigabyte oder einer Infrastruktur unter 30 Server aus finanzieller Sicht sinnvoll. Gemäss Winkler und Meine muss jedoch jede elektronische Backup Lösung noch eine lokale Lösung vor Ort haben, dies erhöht die Sicherheit, treibt jedoch die Kosten in die Höhe und ist somit nicht 50% günstiger als eine konventionelle Backup Lösung. Je nach Ausgangslage einer Firma kann ein elektronisches Backup Sicherheitsvorteile bringen, jedoch muss vorgängig ein sorgfältiges Datensicherungskonzept erstellt werden.

Bei einem Backup muss der Recovery Point Objective oder Recovery Time Objective genau definiert werden. Ob ein elektronisches Backup diese Zeiten erfüllen kann, wird sich je nach Ausgangslage zeigen. Ein elektronisches Backup wird in Zukunft vermehrt in der IT-Branche anzutreffen sein. Wichtig ist dabei, auch die entsprechenden Service-Level-Agreement und Operational-Level-Agreement zu überprüfen (vgl. Nelson 2011, Winkler & Meine, 2011).

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

BaaS	**B**ackup **a**s **a** **S**ervice
BC	**B**usiness **C**ontinuance
BSI	**B**undesamt für **S**icherheit in der **I**nformationstechnik
CD	**C**ompact **D**isc
CDP	**C**ontinuous **D**ata **P**rotection
CIFS	**C**ommon **I**nternet **F**ile **S**ystems
CPU	**C**entral **P**rocessing **U**nit
CRR	**C**ontinuou**s** **D**ata **R**eplication
CSP	**C**loud **S**ervice **P**rovider
DLT	**D**igital **L**inear **T**ape
DR	**D**isaster **R**ecovery
DSG	**D**aten**s**chutz**g**esetz
DVD	**D**igital **V**ideo **D**isc
EDÖB	**E**idgenössischer **D**atenschutz- und **Ö**ffentlichkeitsbeauftragter
GB	**G**iga**B**yte
HTTP	**H**yper**t**ext **T**ransfer **P**rotocol

I/O	**I**nput**/O**utput
IaaS	**I**nfrastructure **a**s **a** **S**ervice
LTO	**L**inear **T**ape-**O**pen
MB/s	**M**egabyte **P**er **S**econd
MD5	**M**essage **D**igest
NAS	**N**etwork **A**ttached **S**torage
NFS	**N**etwork **F**ile **S**ystem
OLA	**O**perational-**L**evel **A**greement
PaaS	**P**latform **a**s **a** **S**ervice
PC	**P**ersonal **C**omputer
REST	**R**epres**e**ntational **S**tate **T**ransfer
RPO	**R**ecovery **P**oint **O**bjective
RTO	**R**ecovery **T**ime **O**bjective
SaaS	**S**oftware **a**s **a** **S**ervice
SHA	**S**ecure **H**ash **A**lgorithm
SLA	**S**ervice-**L**evel **A**greement
SLM	**S**ervice-**L**evel **M**anagement
SLR	**S**ervice-**L**evel **R**equirements
TB	**T**era**B**yte
TCO	**T**otal **C**ost **o**f **O**wnership
UC	**U**nderpinning **C**ontracts
VDSG	**V**erordnung zum Bundesgesetz über den **D**aten**s**chutz
VM	**V**irtual **M**achine
WAN	**W**ide **A**rea **N**etwork

1 Einleitung

1.1 Kundenbedürfnisse

Bechtle Dübendorf ist ein eigenständiges Systemhaus innerhalb der Bechtle-Gruppe, das neben IT-Dienstleistungen auch seit mehr als einem Jahr eine eigene Cloud aufgebaut hat. Die Dienste der Cloud nehmen derzeit rund zwölf unterschiedliche Unternehmen in Anspruch. Einige KMUs von Bechtle haben bereits die Anfragen bezüglich elektronisches Backup (eBackup) beziehungsweise Backup as a Service (BaaS) gestellt. Diesen Service möchte Bechtle Dübendorf im Jahr 2014 realisieren. Ein hierfür passendes Konzept ist allerdings noch nicht vorhanden.

1.2 Problemstellung

Daten auf den Servern werden immer wichtiger, denn sie sind ein Gut, auf welches nur noch selten verzichtet werden kann. Das Verheerendste, das den meisten Firmen oder Privatpersonen passieren könnte, ist ein kompletter Datenverlust. Eine Datensicherung wird jedoch erst gebraucht, wenn ein Server oder Client nicht mehr funktioniert oder von Viren befallen ist. Wie der Tagesanzeiger-Artikel erläutert, hilft bereits ein einfaches Backup um sich vor Datenverlust zu schützen (Tages-Anzeiger, 2012). Ein anderes aktuelles Beispiel im Zusammenhang mit Daten ist der Tape-Diebstahl bei der Firma Swisscom. Der Inhalt der Backup-Tapes/Archivtapes befindet sich nun im Beitz der Neuen Zürcher Zeitung. Dieser Fall ist besonders gravierend, da der Inhalt dieser Bänder vertrauliche Informationen über diverse prominente Personen enthält. Dieses Beispiel zeigt auf welche Konsequenzen Datenverlust bzw. Datendiebstahl mit sich bringen kann. (Tages-Anzeiger, 2013a, S. 43)

Vor mehreren Jahren genügte es den meisten Unternehmen ein Bandlaufwerk auszulagern. Mit zunehmender Datenmenge stossen jedoch die meisten Firmen heutzutage an ihre Grenzen. Ein weiterer Faktor, welcher die IT zunehmend beeinflusst, sind die steigenden Kosten aufgrund der zunehmenden Komplexität der Branche. Die beiden oben genannten Punkte haben zur Folge, dass bei immer mehr Kunden das Bedürfnis entsteht, die Backup-Daten auszulagern, um fehlendes Know-how zu kompensieren und Platz und Kosten zu sparen. Gewisse Unternehmen bevorzugen sogar die physikalische Trennung ihrer Daten über mehrere Kilometer hinweg. Zu dieser Art von Kunden gehören meistens KMUs von 50 bis 250 Personen, welche in diversen Branchen angesiedelt sind. Die Anforderungen an ein Backup sind von Banken, Industriebetrieben bis zu Pharma-Konzernen sehr unterschiedlich. Derzeit besitzen allerdings die wenigsten Firmen eine IT-Sicherheitsstrategie, welche die Risiken und Gefahren im Allgemeinen aufzeigt und an die jeweiligen Bedürfnisse des Unternehmens

angepasst ist. Die Bedeutung der IT Sicherheit hat in den letzten Jahren massiv zugenommen. Auch das Datenvolumen hat sich im Vergleich zu den Vorjahren massiv vergrössert.

In Bezug auf die elektronische Datensicherung müssen einige Kriterien beachtet werden, wie beispielsweise der geografische Standort des Anbieters, die Sicherheit und die Verfügbarkeit. Ein wichtiger Aspekt bei der Auslagerung von Daten einer Firma ist die Berücksichtigung der Service Level Agreements (SLA). Die SLA definieren diverse Merkmale, wie zum Beispiel Recovery time objective (RTO) und Recovery point objective (RPO). Dies gehört zu den wichtigsten Merkmalen einer Backupstrategie. Auch das Thema ‚Big Data' wird in Bezug auf die Datensicherung einen immer grösseren Einfluss nehmen, da die grossen Datenmengen mit den klassischen Sicherungsmethoden nicht mehr bewältigt werden können.

Bei der Auslagerung von Daten müssen Gesetze wie das Datenschutzgesetz (DSG) und allenfalls andere Teile der schweizerischen Gesetzgebung berücksichtigt werden. Denn je nach Branche gibt es verschiedene Datenklassifizierungen, die besonders hohen Schutz verlangen. Somit ist es auch wichtig, wo die Daten gelagert werden, denn dieser Zugriff sollte durch Sicherheitsvorkehrungen geschützt sein. Die geographische Wahl eines Rechenzentrums ist je nach Firma unterschiedlich.

1.3 Hypothese

Die Hypothese lautet wie folgt: Die Kosten eines eBackups lassen sich im Gegensatz zu den Kosten eines konventionellen Backups durch ein sorgfältiges Datensicherungskonzept um 50% reduzieren, sofern die Datenmenge ein Terrabyte nicht übersteigt.

1.4 Ziel der geplanten Untersuchung

Folgendes soll bewiesen werden: Die Kosten eines eBackups lassen sich im Gegensatz zu den Kosten eines konventionellen Backups durch ein sorgfältiges Datensicherungskonzept um 50% reduzieren, sofern die Datenmenge ein Terrabyte nicht übersteigt. Des Weiteren soll die Aussage von Steven Nelson (*Pro Data Backup and Recovery*) verfochten werden. Dieser behauptet, dass die Initialkosten eines konventionellen Backups höher sind als diejenigen eines eBackups. Bedingt durch ein schnelles Datenwachstum können diese Kosten jedoch höher ausfallen als bei einer konventionellen lokalen Backuplösung beim Kunden vor Ort. Der Schwerpunkt der Untersuchung liegt auf den Äusserungen des Autors Steve Nelson, da seine wissenschaftlichen Erkenntnisse in den Bereichen Backup, traditioneller Datensicherung und zukunftsorientierten Backuplösungen zu den aktuellsten der Branche gehören (Nelson, 2011).

Es wird noch ein Ausblick in Bezug auf die physische Sicherheit der Räumlichkeiten gegeben wie z.B. Zutrittskontrollen oder Transport des Backups. Diese sollen mit denjenigen eines konventionellen Backups verglichen werden. Mit Hilfe der ausgewählten Fachliteratur soll die aufgestellte Hypothese wiederlegt oder bewiesen werden. Durch die Konklusion können die Ergebnisse in den Transfer übergeleitet werden. Ziel ist es, dass Bechtle Dübendorf seinen Kunden eine Empfehlung über die optimale Technologie, lokale Datensicherung versus eBackup aufzeigen kann. Die gewonnenen Erkenntnisse sollen zu einem späteren Zeitpunkt erfolgreich umgesetzt werden können.

1.5 Methodische Vorgehensweise

Anhand von unterschiedlicher Fachliteratur soll die Hypothese widerlegt bzw. bewiesen werden. Mittels Beweisführung soll die anfängliche Hypothese zu einer These überführt werden.

Die aufgestellte Hypothese leitet sich von der Aussage des Autors Steven Nelson (*Pro Data Backup and Recovery*) ab. Dieser behauptet, dass die Initialkosten eines konventionellen Backups zwar anfänglich höher sind als die eines eBackups, jedoch können die Kosten eines eBackups bei grösser werdender Datenmenge zu einem späteren Zeitpunkt auch wieder ansteigen. Gerade aus diesem Grund wurde die Datenmenge in der Hypothese auf ein Terrabyte festgelegt. Nur bis zu diesem Wert sollte eine Kostensenkung von 50% mittels einem sorgfältigen Datensicherungskonzept möglich sein. Das Schweizerische Datenschutzgesetz soll in dieser Arbeit auf die einzelnen Anforderungen in Bezug auf Datenhaltung und physische Sicherheit ‚gestreift' werden (Nelson, 2011).

1.6 Definition der Begriffe

Je nach Fachliteratur und Autor wird in dieser Arbeit von Backup, Datensicherung, Deduplizierung, deduplication gesprochen.

2 Theoretische Grundlagen eines Backups

Laut Hoppe und Priess werden nicht nur die Gesellschaften, sondern auch die Wirtschaft seit Anfang der Informations- und Kommunikationstechnologie stark beeinflusst. Informationen sind bei vielen Unternehmen heutzutage das Kerngeschäft der Unternehmung und beeinflussen die strategische Richtung des Unternehmens bzw. dessen Erfolgsfaktor. Diese Informationen können im Extremfall sogar das Überleben der einzelnen Unternehmung gefährden. Die Informationen müssen aktuell, konsistent und qualitativ für das Unternehmen sein und jederzeit zur Verfügung stehen. Mit Hilfe dieser Informationen können neue Märkte, Zielgruppen oder Produkte entwickelt und vertrieben werden. Auch werden Informationen für Entscheidungen benötigt. Nur die wenigsten Unternehmen können heutzutage ohne Informationssystem oder nur eingeschränkt auskommen (Hoppe & Priess, 2003, S. 16).

Hoppe und Priess fügen an, dass sich nicht nur die Anzahl von Informationen rasant entwickelt sondern auch die Gefahren in den letzten Jahren erheblich zugenommen haben. Die einzelnen Schwachstellen der Informationssysteme sind heutzutage sehr verbreitet. Informationen müssen geschützt und rund um die Uhr verfügbar sein. Welche Sicherheitsmassnahmen in Bezug auf die Datensicherung heutzutage wichtig sind, wird in den nächsten Kapiteln beschrieben (Hoppe & Priess, 2003, S. 16).

2.1 Ursachen und Konsequenzen eines Datenverlustes

Eine Datensicherung wird erstellt, weil es diverse Bedrohungen gibt. Was ein Datenverlust kostet wurde durch eine Studie in USA 1998 belegt. Tabelle davon einfügen (Wald, 2002, S. 26).

PCs in Benutzung (USA, 1998)	72‘000‘000
Ermittelt Ursachen für Datenverlust	
Hardwarefehler	1‘921‘300
Menschliches versagen	1‘397‘300
Softwarefehler	611‘300
Computervirus	305‘700
Diebstahl	234‘400
Hardwarezerstörung	131‘000

Gesamt:	4‘601000
Durchschnittliche Kosten pro Vorfall	
Kosten für Wiederherstellung(Technik)	$ 380
Verlorene Produktivität	$ 177
Wert der verlorenen Daten	$ 2‘000
Gesamt:	$ 2‘557
Gesamtkosten Datenverlust in den USA 1998	$ 11.8 Milliarden

Tabelle 1: Die Kosten durch Datenverslust in den USA 1998, Quelle: Darstellung entnommen aus Wald, 2002, S.26.

Müller sieht als Bedrohnungsgruppen z.B. Feuer, Flüssigkeit, technischer Defekt, Angriff von Hackern oder auch Terrorismus. Alle diese Bedrohungen können ein Unternehmen oder dessen Daten bedrohen. Diese Bedrohungen sollten in sogenannte Bedrohungsgruppen zusammengefasst werden. Somit können Massnahmen für die speziellen Bedrohungsgruppen gebildet werden (Müller, 2008, S. 132). Das Deutsche Bundesamt sieht als zusätzliche Bedrohung unkontrollierte Veränderung von gespeicherten Daten, Veränderung gespeicherter Daten oder versehentliches Löschen in Archivsystemen an. Auch interne oder externe Datendiebstähle können nie ausgeschlossen werden (BSI, 2004, S. 3).

Eine zentrale Frage, die sich jede Firma stellen muss ist: „Wie lange und wie oft kann ich es aushalten ‚down’ zu sein?“ (Wald, 2002, S. 28). Es gibt für keine Firma eine exakte Antwort auf diese Frage. Je nach Firmenbedürfnisse können diese Anforderungen unterschiedlich sein. Die einen Firmen haben bereits nach fünf Minuten Ausfallzeit einen erhöhten Schaden, während andere Firmen sich auch einen ganzen Ausfalltag leisten können. Eine nahezu 100%-ige Sicherheit können sich nur die wenigsten Firmen leisten. Jede Firma, die IT als wichtig unterstützender Prozess braucht, sollte sich die Frage stellen, wie lange die Ausfallzeit der einzelnen Systeme sein darf. Dazu wird häufig folgende Regel verwendet: Kosten eines Ausfalls pro Zeiteinheit. Die IT-Abteilungen und die Fachbereiche sollten eine gemeinsamen Lösung zusammen erarbeiten (Wald, 2002, S. 28). Auch das Deutsche Bundesamt für Sicherheit in der Informationstechnik weist darauf hin, dass bei Verlust von Anwendungsdaten oder Kundenstammdaten die Existenz des Unternehmen bedroht ist. (BSI, 2004, S. 3). Auch die Bundesbehörden der Schweizerischen Eidgenossenschaft für Melde- und Analysestelle Informationssicherung (MELANI) empfiehlt eine regelmässige Datensicherung. Somit wird das Risiko vermindert, dass eine Fehlmanipulation, technischer Defekt, Viren oder Würmer die Daten teilweise oder ganz zerstören können. Auch die Überprüfung des Ba-

ckups und dessen Lesbarkeit bzw. Wiederherstellung der Daten wird ausdrücklich empfohlen (Melani, 2013).

2.2 Notwendigkeit eines Backups

Die Autoren Brooks, McFarlance, Pott, Trcka & Tomaz haben bei der Notwendigkeit des Backups festgestellt, dass viele Firmen die eigenen Business-Anforderungen ihrer Firma nicht mehr verstehen, wenn es um das Thema Backup und Speicherkapazität geht. Ein Backup ist wie eine Versicherungspolice. Niemand will sie, weil sie viel kostet und nur selten gebraucht wird, wenn es jedoch zum Ernstfall kommt und in einem Haus eingebrochen wird oder das Auto einen Defekt aufweist, ist jeder dankbar, dass er eine Versicherung abgeschlossen hat, die Unterstützung garantiert. Das Gleiche gilt beim Backup, keiner sieht seine Notwendigkeit bis ein Fall eintritt, wo z.B. ein neuer Report beschädigt wird oder alle Festplatten abstürzen. Wenn kein Backup vorhanden wäre, könnte in einem solchen Fall auch kein Restore zurückgespielt werden. Falls ein Backup vorhanden ist und ein Restore in der gewünschten Zeit zurück gespielt werden kann, sind die entsprechenden Personen zufrieden (Brooks, McFarlance, Pott, Trcka & Tomaz, 2003, S. 62).

Bevor ein Datensicherungskonzept geschrieben wird, muss sich die Firma die Frage stellen, ob sie es sich leisten kann, Daten zu verlieren oder nicht. Wenn diese Frage mit JA beantwortet werden kann, muss kein Datensicherungskonzept geschrieben werden. Viele Firmen werden jedoch ein klares NEIN wiedergeben, da sich die wenigsten einen Datenverlust leisten können. Es ist sogar möglich, dass sich die Firmen nicht einmal einen Datenverlust für kurze Zeit leisten können. Somit ist es wichtig, auch einen Restore schnell und effizient wiederherzustellen zu können (Brooks, et al., 2003, S. 62).

Laut Nelson (2011) kostet ein Backup immer Geld, jedoch kostet es noch mehr wenn keines vorhanden ist und Daten verloren gehen:

> A statistic that is often quoted is that 70 percent of the businesses without a strong backup and recovery infrastructure that were affected by 9/11 never reopened. Of the remaining 30 percent, almost 90 percent of them failed within 6 months. While we all hope that 9/11 was a one-time event, there are any number of catastrophes (fires, floods, tornadoes, and so on) that can have similar impacts from a business perspective. It is a small cost to the business and the people that are employed there to simply have good backups. (S. 262)

Wie Nelson noch anfügt, sollte ein Backup nicht ein Ärgernis sein, das noch erledigt werden muss, da noch nie etwas geschehen ist. Ein gutes Backup bietet Schutz für jedes Unternehmen. Heutzutage sind Informationen eines der wichtigsten Güter eines Unternehmens und dieses Gut sollte auch entsprechend geschützt sein. Je besser die Information, desto besser kann sich auch ein Unternehmen am Wettbewerb behaupten. Falls diese Informationen ver-

loren gehen, ist es fragwürdig ob diese überhaupt wieder hergestellt werden können. „Backups provide a means by which all these functions can be completed, thus protecting critical assets within the organization“ (Nelson, 2011, S. 261).

2.3 Daten

Ohlhorst sagt, dass eine menschliche Gehirngrösse ungefähr 2.5 Pettabyte gross ist. Falls ein Gehirn so funktionieren würde wie eine Kamera im Fernsehen, dann wären 2.5 petabytes genug Platz um 3 Millionen Stunden Fernsehshows aufzunehmen. Um den ganzen Speicherplatz aufzubrauchen müsste der Fernseher die ganze Zeit laufen für mehr als 300 Jahre. Wenn man die heutige Technologie hierbei in Bezug auf Big Data vergleicht, wachsen diese Daten exponentiell (Ohlhorst, 2013, S.16).

Heutzutage werden auch im Business immer mehr Informationen bzw. Daten gebraucht. Diese Datenmengen wachsen ständig und verbrauchen immer mehr Speicherplatz. Die Daten sind immer schwieriger zu verwalten, welches ein sogenanntes Big Data erzeugt (Ohlhorst, 2013, S.16). Wie Ohlhorst (2013) feststellt haben auch diese Daten auf die Backup Strategie einen Einfluss:

> „The reasons vary for the need to record such massive amounts of information. Sometimes the reason is adherence to compliance regulations, at other times it is the need to preserve transactions, and in many cases it is simply part of a backup strategy.” (S.16)

2.4 Big Data

Wie Hurwitz, Nugent, Halper & Kaufman erwähnen nimmt das Verwalten und Analysieren von Daten bei jedem Unternehmen einen immer wichtigeren Standpunkt ein. Viele Unternehmen sind in den letzten Jahren gewachsen und ihr Produktportfolio hat sich dementsprechend vergrössert. Immer mehr Kundendaten werden von jedem Unternehmen gesammelt und analysiert. Die Herausforderung dieser enormen Datenmenge ist heute, wie die Daten der einzelnen Applikationen verknüpft werden, sodass eine grosse Datenmenge zentral analysiert werden kann. Big Data ist der neuste Trend, der in die Richtung grosse Datenmengen geht (Hurwitz, Nugent, Halper & Kaufman 2013, S. 9).

K. Davis und D. Patterson sehen dies noch von einem anderen Blickwinkel aus. „At this point you might be asking: ‚Why not just any data?’“(Davis & Patterson, 2012, S.4). Die Definition: Big Data wurde auf der Grundlage der Unterschiede zwischen den Fähigkeiten von den Legacy Datenbanktechnologien, den neuen Datenspeichern und dessen Verarbeitungstechniken definiert. Denn Big Data sind zu grosse Datenmengen, das es traditionelle Datenbank Protokolle wie SQL managen könnten. Aus diesem Blickwinkel ist es nur eine neue Form

von neuen Technologien. Das Volumen, die Vielfalt und die Geschwindigkeit erhöht die Komplexität exponentiell, denn für jedes Unternehmen ist es schwierig diese Daten zu verwalten bzw. zu analysieren. Das Resultat, das mit diesen neuen Daten herausgelesen werden kann, ist atemberaubend. Bereits heute leben wir in der Grösse von Exabytes und Zettabytes, bis 2025 wird erwartet, dass die Kapazität des Internet's die Gehirnkapazität von allen lebenden Menschen überschreiten wird (David & Patterson, 2012, S. 4).

Wie sich Big Data in Bezug auf das Datenwachstum auswirkt, wird im nächsten Kapitel erläutert. Laut Ohlhorst hat Big Data einen grossen Einfluss auf das Datenwachstum, denn durch die neuen Technologien und neuen Datenquellen wachsen diese exponentiell. Die Wachstumsrate der meistens Unternehmen ist nahezu unendlich weil die meisten Unternehmen im Big Data Bereich bei null gestartet haben (Ohlhorst, 2013, S.82).

2.5 Datenwachstum

Wie im Anhang die Studie von IDC/EMC 2011 wiederlegt, wird ein Datenwachstum bis im Jahr 2015: 7.9 Zettabyte ergeben und im Jahr 2020 gar 35 Zettabyte, wobei 1/3 dieser Daten in der Cloud sein werden. Von 2012 bis 2020 sind dies ca. 4300% laut CSC (CSC, 2011). Die Autoren Drakos und Paquet von Gartner berichtet im Jahr 2010, dass die Unternehmensdaten einen Wachstum von 650% in den nächsten 5 Jahren haben werden. 80% dieser Daten werden unstrukturierte Daten sein (Drakos & Paquet, 2009, S. 8). "Technology Trends You Can't Afford to Ignore" (Drakos & Paquet, 2009, S. 8).

Die aktuellste Studie von Dezember 2012 der IDC wiederlegt, dass das Datenwachstum von 2005 bis 2020 Faktor 300 zunimmt von 130 Exabytes bis 40'000 Exabytes. Dies entspricht rund 40 Zettabytes. Dies ergibt rund 5200 Gigabytes für jede Frau, Mann und Kind im Jahre 2020 (Gantz & Reinsel, 2012).

2.6 Traditionelle Datensicherung

Die klassische Methode seine Daten zu sichern, ist immer noch ein dediziertes Bandlaufwerk. Wald beschreibt schon 2002, dass die Tape Laufwerke einen grossen Vorteil mit sich bringen, neben der grossen Datenmenge, die auf den Magnetbänder gespeichert werden kann, lassen sich die Tapes einfach auslagern (Wald, 2002, S. 83).

2.6.1 Sicherungsmethoden

Müller beschreibt, dass sich bei einem Datensicherungskonzept die folgenden vier Themen unterscheiden lassen:

„1. Sicherungsmethoden (komplette, differenzielle, inkrementelle, selektive Datensicherung)

2. Sicherungszeitpunkt

3. Aufbewahrungsort

4. Auslagerungsverfahren“ (Müller, 2008, S. 202)

Auch beim Backup lassen sich verschiedene Sicherungsmethoden unterscheiden. Bei der kompletten Datensicherung sichert man den ganzen Datenbestand. Es ist wichtig, dass der benötigte Zeitraum für die Datensicherung zu Verfügung steht. Desweitern sollten während dieser Zeit auch die Daten nicht verändert werden, um die Inkonsistenz zu vermeiden. Der Vorteil von diesem Verfahren ist das eine komplette Wiederherstellung gefahren werden kann. Diese Sicherung braucht jedoch einen hohen Zeitbedarf (Müller, 2008, S. 203). Beim Bundesamt Datensicherungskonzept wird jedoch noch erläutert, dass der Zeitraum zwischen den Sicherungen nicht lange auseinander sein sollte. Ein Wiederherstellen der Daten sei bei einer Volldatensicherung schnell und einfach ausgeführt (BSI, 2004, S. 4).

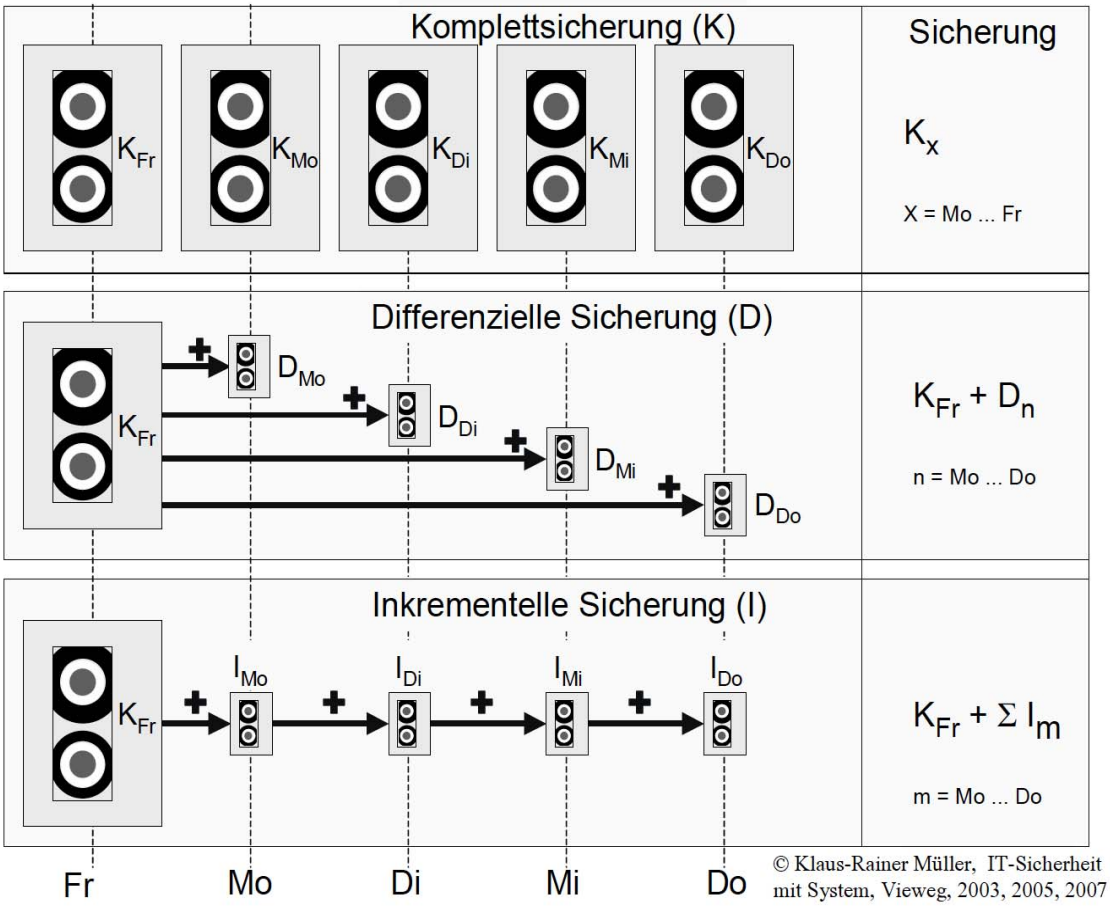


Abbildung 1: Datensicherungsmethoden, Quelle: Darstellung entnommen aus Müller, 2008, S. 203

Für die differenzielle Datensicherung braucht es als Erstes eine komplette Datensicherung, in einem zweiten Schritt werden nur die veränderten Daten gesichert, dies ist die sogenannte Differenz zur Anfangssicherung (Müller, 2008, S. 203). Steve Nelson sieht beim differentialen

Backup jedoch folgendes Problem. Ein Differentialbackup kann schnell wachsen, durch die Veränderung kann es schnell grösser sein als die komplette Datensicherung (Nelson, 2011, S. 4).

Steve Nelson (2011) gibt ein entsprechendes Beispiel:

> An environment has 20 TB of data to back up. Each day 5 percent or 1 TB of data changes in the environment. Assuming that this is a traditional backup environment, if a differential backup methodology is used, the first day 1TB of data is backed up (the first day's change rate against the previous full backup). The second day, 2 TB is backed up, and so on. By the end of 5 days, 5 TB of data is being backed up; by the end of 10 days, 10 TB might be being backed up; in 20 days, it could be a backup of 20 TB. However, this 20 TB is not necessarily representative of a full backup. If the change rate represents a mixture of new files added to the respective clients as well as changes to existing data, the cumulative data backed up will not capture data that has not changed, even though the total amount of a full backup would still only incrementally change in size.(S. 4)

Steven Nelson erläutert auch den Vorteil bei einem Restore eines Differentialbackups. Hier müssen nur zwei Restores gefahren werden, ein Full Backup und das neuste Differentialbackup. Denn es gibt nur eine Anzahl von Snaphosts, dass jedoch zwei Snapshots gleichzeitig defekt sind, hierfür ist die Wahrscheinlichkeit sehr klein (Nelson, 2011, S .4). Müller sieht den Vorteil beim Differentialbackup gegenüber dem Vollbackup, da die Folgesicherungen nach der kompletten Datensicherung um ein Vielfaches kürzer sind. Der Nachteil sieht Müller jedoch bei der Zweistufigkeit im Gegensatz zur kompletten Datensicherung (Müller, 2008, S. 203).

Laut Müller beginnt man bei der inkrementellen Datensicherung zuerst mit einer kompletten Datensicherung. Als zweiter Schritt bzw. bei der Folgesicherung werden nur die Daten gesichert, die sich bei der letzten Sicherung verändert haben. Auch die inkrementelle Datensicherung wird wie die differenzielle Datensicherung im Wochenrhytmus durchgeführt. Der Vorteil gegenüber einer kompletten Datensicherung zeichnet sich durch die kleinen Zeitfenster der Folgesicherungen aus. Der Nachteil sieht Müller bei der Rückspielung einer Datensicherung, da diese in der richtigen Reihenfolge geschehen muss. Hier muss auch die komplette Datensicherung als Grundlage genommen werden. Dies braucht somit viel Zeit und ist kompliziert, wenn es viele Tapes gibt (Müller, 2008, S. 204).

Nelson erläutert, dass die inkrementelle Datensicherung die meist genutzte Form der Datensicherung ist. Nelson vergleicht es mit dem obengenannten Beispiel. Als Erstes wird wieder eine komplette Datensicherung von 20TB durchgeführt. Am zweiten Tag wird eine weitere 1TB Datensicherung erstellt, sowie am Folgetag danach. Am achten Tag würde es wieder eine komplette Datensicherung geben. Auch Nelson sieht den Nachteil bei einer Wiederherstellung der Daten, da die komplette Datensicherung und eine inkrementelle Sicherung gebraucht werden. Steve Nelson (2011) schildert die Problematik anhand folgendem Beispiel:

> Suppose that four files need to be recovered: one has not changed since the full backup, one changed the first day, one changed the second day, and one changed the third day. To complete this restore, four images—the full backup and three different incremental images—are required to retrieve all the files needed (see Figure 1–3). In addition, because the images are relative to each other, some pieces of backup software will not allow parallel restores from the related sets of backup images, so the file recovers have to occur in serial fashion, each loading and unloading its own set of images for the restore. This can have a large impact on the time required for a restore. (S. 4)

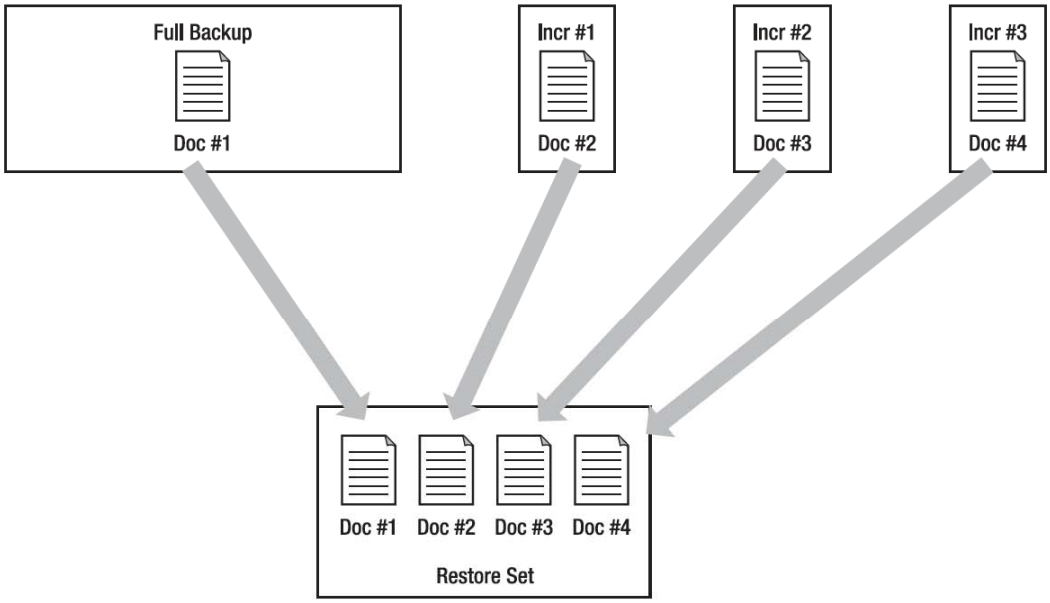


Figure 1–3. *Multiple file incremental restore*

Abbildung 2: Rückspielung von mehreren Incremental Files, Quelle: Darstellung entnommen aus Nelson, 2011, S. 5

Ein weiteres Problem bei Inkrementell ist die Wahrscheinlichkeit, dass die Backup-Software einzelne Daten nicht wiederherstellen kann, weil sie verloren gegangen oder beschädigt sind. Im Gegensatz zu Müller beschreibt Nelson noch eine Alternative bezüglich inkrementellem Backup. Hier wird als Erstes eine komplette Sicherung durchgeführt. Als zweiter Schritt werden die Änderungen gesichert, beim dritten Backup wird die Differenz vom zweiten Backup an das dritte Backup übertragen. Dies ist laut Nelson eine leistungsfähige Methode. Die Schwächen liegen jedoch, wie beim inkrementellen Backup, bei der Wiederherstellung. Steve Nelson (2011) schildert die Problematik mit folgendem Beispiel:

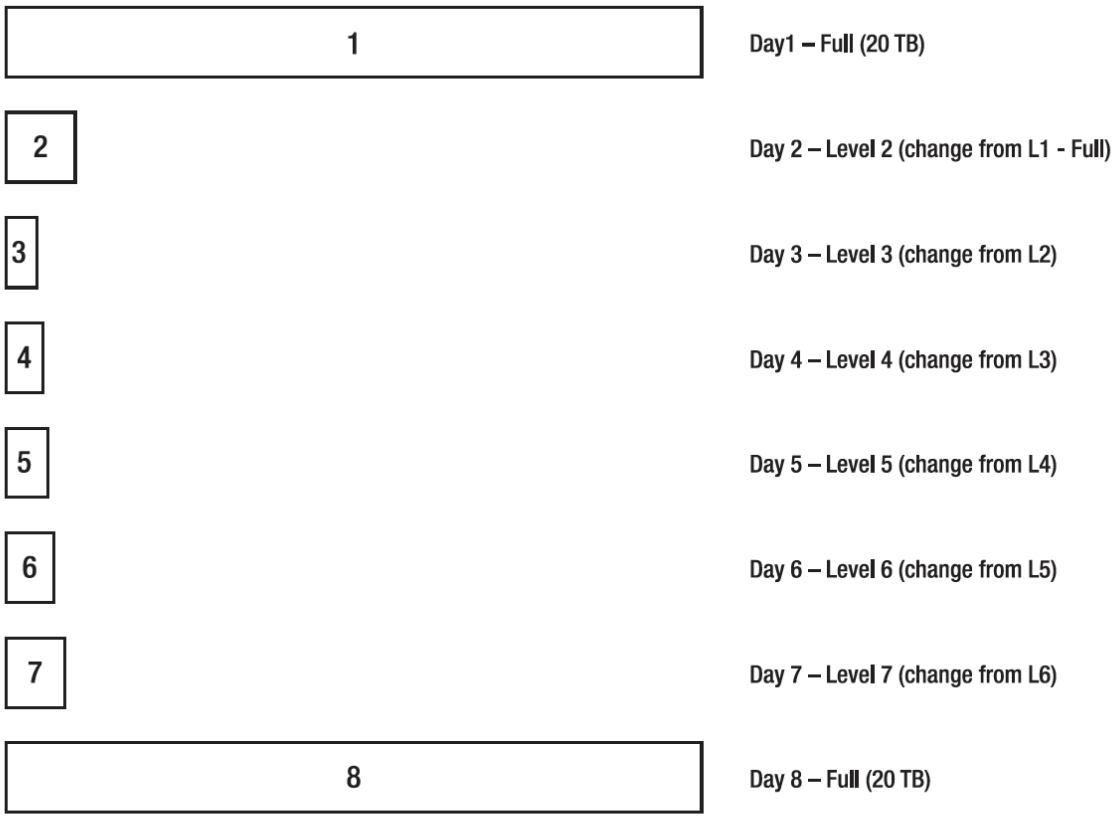


Figure 1–4. *Level-based backups*

Abbildung 3: Level-Based Backup anhand 8 Tagen, Quelle: Darstellung entnommen aus Nelson, 2011, S.6

> Consider the 20 TB example once again (see Figure 1–4). This time, a level 0 backup is executed against the 20 TB data set, capturing all 20 TB of data. The next backup executed is a level 1, which captures 1 TB of backup, representing the change since level 0 (the full backup). Now the next backup is a level 2 backup, which captures only the differences between the level 1 and the current backup—likely much smaller than a comparable level 1 backup. If another level 2 backup is executed as the next backup, the differences between the current data set and the last level 1 backup are captured. This type of backup can be much smaller than a comparable set of traditional incremental backups, resulting in short backup windows. However, the weaknesses of the incremental backup are also present in the level backup and need to be taken into consideration. (S. 5)

Durch die verschiedenen Backup-Level können mehrere Versionen der einzelnen Datensätze erkannt werden. Die Backup-Software kann anhand von statischen Informationen, wie z.B. der Dateiname, bei der Wiederherstellung mehrere Versionen der Datensätze anzeigen lassen (Nelso, 2011, S. 6).

Müller zeigt in der folgenden Tabelle die Unterschiede der Sicherungsmethoden auf:

Sicherungsumfang	Zeitfenster		Medien	Komplexität
	Sicherung	Rücksicherung		
Komplett	Gross	Klein	Hoch	Niedrig
Differenziell	Mittel	Mittel	Mittel	Mittel
Inkrementell	Klein	Gross	Niedrig	Hoch

Tabelle 2: Vor- und Nachteile der Sicherungsmethoden, Quelle: Darstellung entnommen aus Müller, 2008, S. 204

Müller erwähnt zusätzlich, dass ein Defekt eines Datenbandes unterschiedliche Auswirkungen auf die verschiedenen Sicherungsmethoden haben kann. Bei einer Komplettsicherung und inkrementellen Datensicherung kann ein ganzer Tag verloren gehen. Bei der inkrementellen Datensicherung kann es gar eine Woche sein. Müller geht noch auf das Vater-Sohn-Prinzip ein. Dabei lassen sich eine Tagessicherung: Montag bis Sonntag (Sohn) eine Wochensicherung (Vater), eine Monatssicherung (Grossvater) und eine Jahressicherung (Urgrossvater) sichern. Jedes Band erhält eine spezielle Kennzeichnung, was älter als ein Jahr ist. Bei der täglichen Sicherung kommt die inkrementelle Datensicherung zum Zuge, während bei der Wochen-, Monats-, Jahres-Sicherung eine Komplettsicherung durchgeführt wird. Die Wochensicherung wird immer an einem bestimmten Wochentag durchgeführt, wodurch nachher die inkrementelle Tagessicherungen wieder anfängt (Müller, 2008, S. 205).

Nelson geht nicht auf das Vater-Sohn-Prinzip ein, sondern erläutert den gesamten Platzbedarf bei 20 TB. Bei einem Backup von vier Wochen ist die Hochrechnung von Nelson ein Minimum von 150 TB. Pro Woche wird eine komplette Datensicherung und unter der Woche ein inkrementelle Datensicherung durchgeführt. Die Gesamtgrösse der Datensicherung ist nicht die treibende Kraft, sondern die inkrementellen Kopien, die wiederholt vorkommen. Wenn eine Datensicherung für eine sehr lange Zeit (mehrere Jahre) gespeichert werden soll, muss zwingend auf die Wiederherstellung geachtet werden. In diesen Zeitraum kann sich die Hardware, Backupsoftware, Betriebssystem ändern. Im schlimmsten Fall kann das Backup nicht zurückgespielt werden. Wenn dies der Fall ist, sollte man die Ausschau Richtung Archivierung halten (Nelson, 2011, S. 8). In Bezug auf Tape Laufwerke äussert sich Nelson wie folgt. Der Vorteil von Tape Laufwerken ist die günstige Preis per Medium in Bezug auf $ / TB. Die Einstiegskosten sind relativ gering und die Lagerung der Sicherungskopie kann einfach an ein anderes Ort Transportiert werden. Jedoch stehen auch diverse Nachteile eines Tape Laufwerks, durch das einfache Transportieren ist es möglich das unerlaubte dritte dies lesen können. Ab dem LTO-4 Standard ist es auch möglich die Tapes zu verschlüsseln. Des weiteren sollten die Bänder nicht in einem herkömmlichen Schrank gelagert sein, sie sollten in

einem Klimatisiertem Raum gelagert werden wo auch die Luftfeuchtigkeit überwacht wird und die sensiblen Tapes von Feuchtigkeit und Staub geschützt sind (Nelson, 2011, S. 48).

> But tape's biggest disadvantage lies in the reliability of the media and drives. Tape drives and media are highly complex, high-precision pieces of equipment. Tape drives are masters of mechanical complexity, with motors that move tape across spinning heads at very precise rates of speed, can stop near instantaneously, and then restart move tape media back to high speed within very short periods of time. (Nelson, 2011, S. 48)

Die Zuverlässigkeit von Bandlaufwerken nicht immer gegen ist und die Magnetspurt von Zeit zu Zeit gewartet werden sollte (Nelson, 2011, S.48). Wald erwähnt noch das die Manuele Bedingung eines Backup-Gerätes auch Risiken mit sich bringt, wenn dies nicht eine Geschulte Person durchgeführt wird (Wald, 2002, S. 82).

2.6.2 Backup oder Archive

Archivsysteme und Backup haben diverse Gemeinsamkeiten betont Wald, denn die Hard- und Software sind in vielen Beispielen gleich oder sehr ähnlich. Ein Archiv kann auch als Langzeitbackup benutzt werden und die Bänder werden an einem sicheren Ort aufbewahrt. Wenn die Daten auf Band ausgelagert und auf dem Speichersystem gelöscht werden, muss eine fehlerfreie Auslagerung garantiert werden. Dies kann auch durch mehrere Kopien erfolgen (Wald, 2002, S. 150).

> Eine wichtige Besonderheit im Zusammenhang mit Archivierung ist die über Jahre zu erhaltende Möglichkeit, die Bänder technisch noch lesen zu können. Die technologischen Fortschritte sind so rasant, dass teilweise eine Abwärtskompatibilität nicht mehr garantiert werden kann. Hier gilt es rechtzeitig das Umkopieren auf modernere Archivierungsmedien vorzunehmen. (Wald, 2002, S. 150)

Der offene Linear-Tape-Open-Gruppe (LTO) Standard wurde 1997 von der Gruppe IBM, HP und Seagate gegründet und diente als alternative zu Digital Linear Tape (DLT). Dieser Standard hat neue Ansätze im Bereich Bandlaufwerk entwickelt und sich als Weltweiterstandard durchgesetzt (Wald, 2002, S. 78). Auch Nelson wieder gibt drei Hersteller, LTO, DLT und noch T10K dieses wurde von Sun / StorageTek entwickelt welches jedoch nicht sehr weitverbreitet ist wie LTO. (Nelson, 2011, S. 38) Bei der LTO 6 lassen sich heutzutage 6.25 Terrabyte speichern bei einer Übertragung von 400 MB/s, jedoch kann ein LTO 6 Laufwerk meistens zwei Generation darunter noch lesen also bis LTO 4 (LTO, 2013).

2.7 Deduplication

Laut Nelson hat nur eine Art von Revolution die Backup-Welt verändert. Dank der Server-Virtualisierungstechnologien haben neue Methoden die Bereitstellung von Backups, die nicht

zur traditionellen Begriff Sicherung passen verändert. Die Duplizierung ist eines der meist gesprochen Themen bezüglich neuen Technologien (Nelson, 2011, S. 87).

Im Kapitel 2.5 wird von einem enormen Datenwachstum gesprochen. Die Autoren Osuna A., Balogh E., Carvalho A., Javier R. und Mann Z. sagen, dass die Geschäftsdaten in den kommenden Jahren rasant wachsen werden. Viele dieser Daten werden immer noch auf Festplatten gespeichert. Auch die Menge an Speichersystemen wächst stetig, der Fokus in Bezug auf die auf die Verbesserung der Storage Effizienz in der gesamten IT-Infrastruktur wird immer grösser. Anhand von neuen Technologien wie Deduplication (Duplizierung) können die Daten verkleinert werden, davon kann ein besser Total Cost of Owenership (TCO) für Speicherinfrastrukturen erzielt werden. Mit Hilfe von Deduplication können kleinere Netzwerk Bandbreiten überwunden werden, sowie auch eine erhebliche Datenreduktion (Osuna, Balogh, Galante de Carvalho, Javier & Mann, 2011, S. 3).

Wie Osuana A. et al. erläutert, wie bei der Abbildung 4 zuerkennen ist wird bei Deduplication eine Konsolidierung der redundanten Kopien einer Datei oder dessen Unterkomponenten konsolidiert. Die Neuen Daten werden in Brocken (engl. ‚Chunks') zerlegt und auf die Redundanz überprüft. Wenn ein sogenanntes Duplikate gefunden wurde, wird dies verschoben bzw. konsolidiert (Osuna et al., 2011, S. 5).

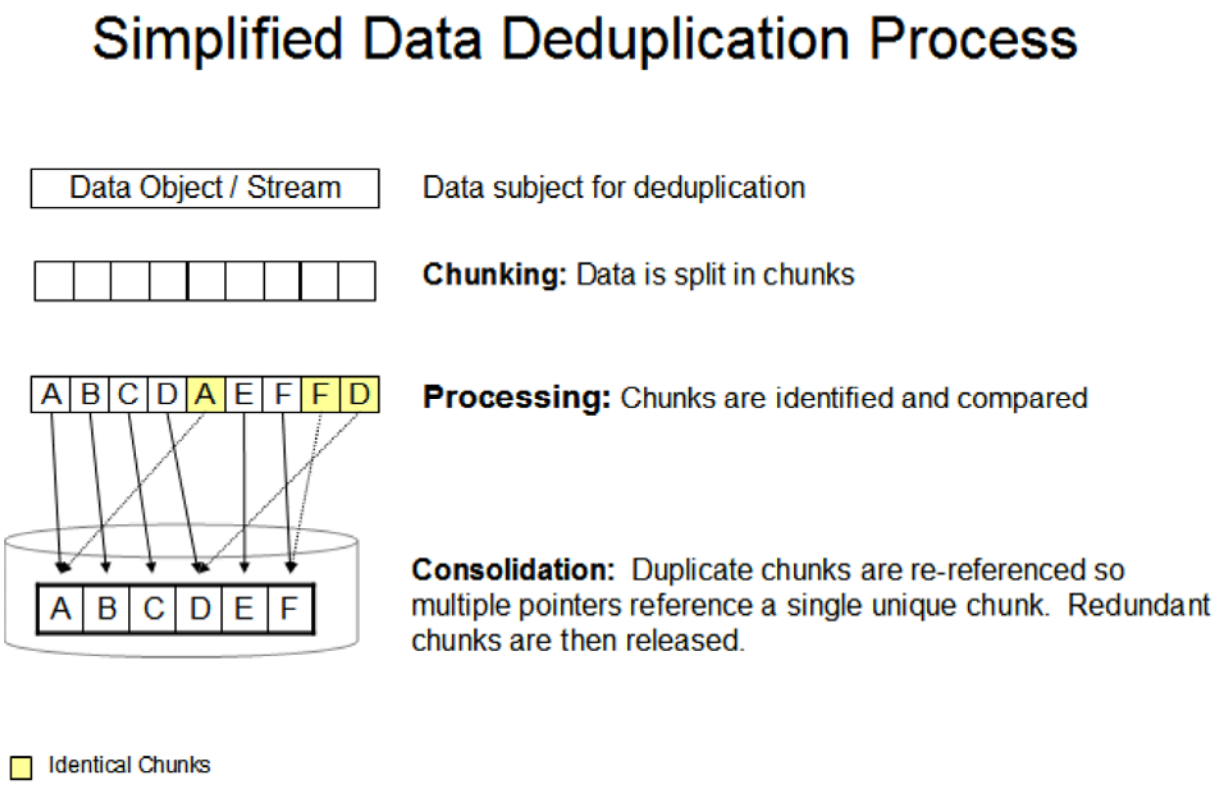


Abbildung 4: Deduplication Methode, Quelle: Darstellung entnommen aus Osuna et al., 2011, S. 5

Osuana A. et al. erläutern die vier unterschiedlichen Methoden wie File based, Block based, Format Aware und Format agnostic in Bezug auf Deduplication. Jedes dieser Methoden beeinflusst das Verhältnis der Deduplizierung.

File based Die File Based Methode weniger effizient als andere Dedpulication Methoden, da sie nur auf File ebene dupliziert.

Block based Bei Blockspeichergeräten wird meistens Block basierte Deduplizierung verwendet. Bei Block based Deduplication werden die Blöcke unabhängig vom Dateitype, Anwendung und Betriebssystem verglichen. Auch verschlüsselte Files können auf Block Level dedupliziert werden. Die Blockgrösse kann in feste oder Variable Grösse zerlegt werden. Die Block based Deduplication benötigt jedoch mehr Rechenleistung und einen grösseren Index Katalog um die einzelnen Stücke zu verfolgen (Osuna et al., 2011, S. 5).

Nelson hingegen unterscheidend im Gegensatz zu Osuana A. et al. die Block Deduplication in zwei Formen. Die sogenannte Fix Block deduplizierung und die Variable-Block Deduplication. Bei der Fix-Blocked dedpulication wird ein Fixer Bit wert definiert dieser ist meistens 4 Bits gross. Die alternative zu Fix-Blocked dedpulication ist die Variable variante. Hier kann die Block grösse von 2 Bit bis zu 6 Bits variieren, denn er sucht die Markierungen bei den Blöcken. Dies ist eine sehr gute deduplizierung Methode, jedoch steckt auch hinter dieser Methode ein kompliziertes Mathemathische Struktur die durch verschiedene patente Geschütz ist (Nelson, 2011, S. 93).

Format aware: Bei der Form aware werden expliziten die Dateiformate und Datenobjekt Stücke entsprechend zerlegt. Eine Powerpoint Präsentation kann z.B. in separate Sides zerlegt werden.

Format agnostic: Der Format agnostic basiert auf einem algorithumus der ähnliche elemente in einem Objekt sucht (Osuna et al., 2011, S. 5).

Wie Osuana A. et al. erläutert gibt es in Bezug auf die Differenzierung eines Brockens drei Methoden, jede dieser Methoden beeinflusst die Leistung.

Hasing: Beim Hashing wird die Prüfsumme (MD-5, SHA-1, SHA-2) für jedes Datensegment berechnet und vergleicht dieser mit den vorhanden Daten. Ein identischer Hash bedeutet, das die Daten wahrscheinlich identisch sind. Bei einer Hash Kollision ist der Hash Wert gleich jedoch die Daten nicht, um diese Datenkorruption zu vermeiden wird eine zweite Variante wie z.B. Metadaten-Vergleich, Binär-Vergleich oder einen zweiten Hash-Vergleich durchgeführt.

Binary comparison: Bei dieser Methode werden alle Bytes oder ähnliche Brocken verglichen

Delta differencing: Bei der letzten Methode wird das Delta zwischen zwei ähnlichen Brocken verglichen. Da die Differenz einzigartig ist kann es keine Kollision geben. Für eine Rekonstruktion des Brocken werden die Deltas mit der Baseline zusammengefügt (Osuna et al., 2011, S. 6).

Bezüglich Hash-Based deduplication äussert sich Osuna, Cecchetti, Franz und Mencarelli wie folgt: „The size of the index may also impact scalability, as the index space is required to increase. Assuming an 8 kilobyte (KB) data chunk, processing 10 terabytes (TB) of data may require 1,250,000,000 accesses to an index“ (Osuna, Cecchetti, Franz & Mencarelli, 2010, S. 6).

Osuna et al. sprechen von den folgenden drei Arten wie Hash Based, Content Aware und Hyper Factor in Bezug deduplicaiton. Bei der Content Aware Deduplication Methode werden gemeinsame Muster oder Strukturen der Daten von Applikationen genutzt. Bei dieser Deduplizierung ist der Beste Kandidat ein Objekt mit den gleichen Eigenschaften wie z.B. der selbe Dateiname. Wenn diese Datei gefunden worden ist, wird bit für bit Verglichen und nur die Veränderungen werden gespeichert (Osuna et al., 2010, S. 7). „The index is file size dependent. An average file size of one megabyte (MB) would require 10,000,000 accesses to an index to process 10 TB of data“ (Osuna et al., 2010, S. 7). Die Hyper Factor Deduplication ist eine IBM Technologie die bei einem Storage bis zu 25 mal mehr Speicherplatz ermöglicht. Die Hyperfactor Methode vergleicht auch die Daten und macht einen byte level vergleich (Osuna et al., 2010, S. 7).

Laut Osuana A. et al. gibt es zwei Möglichkeiten wie die Datenströme verarbeitet werden. Bei der Inline Deduplication werden die Daten, bevor sie auf die Festplatte geschrieben werden dedupliziert, dies bedeutet das der Hash-Vergleich zuerst geschehen muss und erst danach auf die Disk geschrieben werden kann. Inline deduplication ist im generellen nicht empfehlenswert. Der Vorteil jedoch von inline Deduplication ist, das die nicht deduplizierten Brocken nie auf die Disk System geschrieben werden (Osuna et al., 2011, S. 8). Nelson argumentiert jedoch das es beim Inline Deduplication verfahren nur ein absolutes Minimum an Storage Anforderung für die Lösung braucht. Nelson argumentiert das inline Deduplication langsamer ist als post-process deduplication (Nelson, 2011 S.95).

Beim Postprocess Deduplication werden die Daten zuerst auf die Disk geschrieben und erst in einem zweiten Schritt dedupliziert (Osuna et al., 2011, S. 8). „This allows more control over the performance impact of deduplication to avoid peaks or conflicts with other data I/O processes such as backups or DR replication“ (Osuna et al., 2011, S. 8). Bei dieser dedupli-

zierung braucht es Temporär mehr Speicherplatz auf dem Storage bis die deduplizierung abgeschlossen ist. Dies ist auch die einzige Möglichkeiten Daten zu deduplizieren, wenn diese bereits auf einem Storage System liegen (Osuna et al., 2011, S. 8). Nelson (2011) argumentiert hier wie folgt:

> This has the advantage of ingesting the backup stream as the fastest rate possible, but also can require substantial amounts of disk space, over that required for the deduplicated storage, as the landing area needs to be large enough to hold all the inbound data long enough to be processed and moved to the deduplicated target storage. (S. 95)

Osuana A. et al. zeigen auf das es drei Arbeiten in Bezug auf die Rechenleistung von Deduplizierungen gibt. Bei der Client Based Dedulication wir die ganze Last von der CPU und Disk I/O Client seitig genutzt. Bei der Server Base Deduplication werden die ganzen Ressourcen der CPU und Disk I/O vom Server genutzt. Bei der Storage System Deduplciation wird die ganze I/O und CPU vom Storage System selber genutzt und belastet somit den Server und Client nicht. Die Vorteile von Deduplication erwähnen die Autoren Osuana A. et al. wie folgt (Osuna et al., 2011, S. 7).

Osuana A. et al. (2011) sehen den Vorteil von Deduplication ganz klar beim Total Cost of Ownership (TCO):

- Storage capacity by allowing you to store more data per physical storage system
- Energy by using less data/disk, thereby requiring less energy
- The amount of data that must be sent across a network to primary storage, for backup replication, and for disaster recovery
  In case you are using disks for backups, the more efficient use of disk space also allows for longer disk retention periods, which provides better recovery time objectives (RTO) for a longer time. (S. 8)

Nelson (2011) sieht die Komprimierung wie folgt:

> Typically, good deduplication will achieve ratios of 10:1 and will in many cases regularly reach over 15:1. A word here about deduplication ratios (or data reduction ratios, as they are also known): from a mathematics standpoint, the dedup ratio is simply the inverse of the percentage of data reduction, as shown here: (S. 87)

$$DD\ ratio = \frac{1}{1 - Data\ reduction\ \%}$$

For example, if a backup is reduced in size by 85 percent through the use of deduplication, the resulting deduplication ratio would be the following:

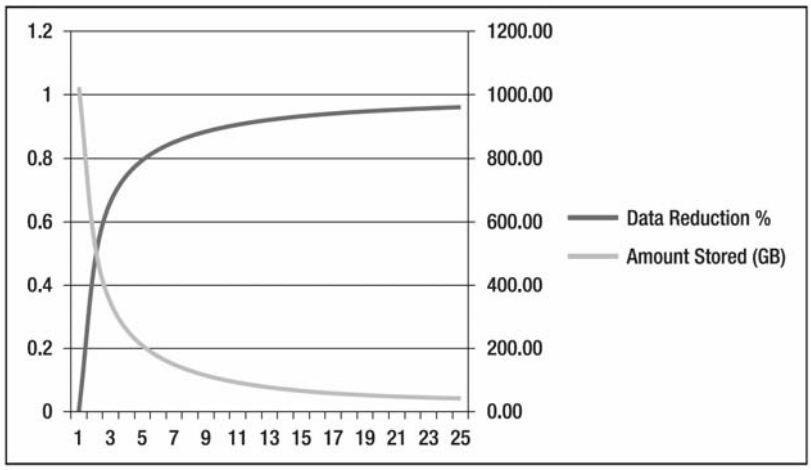
$$DD\ ratio = \frac{1}{1 - .85}$$

$$DD\ ratio = \frac{1}{.15}$$

$$DD\ ratio = 6.66\ or\ 6{:}1$$

Abbildung 5: Deduplication Methode, Quelle: Darstellung entnommen aus Nelson, 2011, S. 87

Nelson erläutert, das wie viele Statistiken auch Deduplication falsch interpretiert oder falsch verstanden werden kann. Viele Produkte werben mit Dedpulication und ihrer Deduplizierung von 10:1, 15:1, 20;1 oder gar noch höherer (Nelson, 2011 S. 88). „The ever increasing values of the deduplication ratio seem to imply that the difference between a 10:1 deduplication ratio and a 20:1 ratio would represent a very large decrease in the amount of data actually stored on a particular device“ (Nelson, 2011 S. 88).



De-dup ratio	Data Reduction %	Amount Stored (GB)	De-dup ratio	Data Reduction %	Amount Stored (GB)
1:1	0.00%	1024.00	13:1	92.31%	78.77
2:1	50.00%	512.00	14:1	92.86%	73.14
3:1	66.67%	341.33	15:1	93.33%	68.27
4:1	75.00%	256.00	16:1	93.75%	64.00
5:1	80.00%	204.80	17:1	94.12%	60.24
6:1	83.33%	170.67	18:1	94.44%	56.89
7:1	85.71%	146.29	19:1	94.74%	53.89
8:1	87.50%	128.00	20:1	95.00%	51.20
9:1	88.89%	113.78	21:1	95.24%	48.76
10:1	90.00%	102.40	22:1	95.45%	46.55
11:1	90.91%	93.09	23:1	95.65%	44.52
12:1	91.67%	85.33	24:1	95.83%	42.67
			25:1	96.00%	40.96

Abbildung 6: Deduplication Verhältnis bei der Gegenüberstllung von einem Terrabyte, Quelle: Darstellung entnommen aus Nelson, 2011, S. 89

Wie Nelson erläutert zeigt die Abbildung 6 die Beziehung zwischen der Deduplication und dessen Datenmenge die tatsächlich gespeichert wird. Hier ist der Vergleich bei einem Terrabyte Backup in Prozent wie es verkleinert wird (Nelson, 2011 S. 89).

„What is interesting is that the difference between the amount of data stored between two adjacent ratios gets ever smaller as the ratios increase“ (Nelson, 2011, S. 89).

> For instance, if Product A advertises a 20:1 deduplication ratio, and Product B advertises a 25:1 ratio, for a 1 TB backup, the difference in the amount of data actually stored is only 10 G of storage—a difference of less than 1 percent of the original 1 TB backup (Nelson, 2011, S. 90)

Laut Nelson hat Deduplication auch Nachteile, denn es gibt BAD Deduplication Kandidaten die sogenannten 3 Ms: Medien, Mutter Natur und Mystery / Verschlüsslung. Medien wie

.mp3, .jpg und andere Formate weisen nur kleine Wiederholungsmuster auf. Diese dieser Art von Daten bringt die Deduplizierung nur eine kleine Anzahl an Vorteilen. In Zusammenhang mit Mutter Natur gibt Bilddaten oder grosse Mengen von Statistischen Daten wie seismologische Studien die sich nur schwer deduplizieren lassen. Als letztes noch die Mystery oder auch Verschlüsslung die sich auch nicht deduplizieren lassen. Bei der Verschlüsselung gibt es keine Wiederholung der Datenblöcke somit ist die Deduplizierung fast 1:1 ohne Daten Komprimierung. Die Daten werden jedoch nach der Deduplizierung verschlüsselt gespeichert, denn die auf dem Backup sind nur die Unikate gespeichert. Denn alle anderen Blöcke sind im index gespeichert und können nur bei einer Entschlüsselung gelesen werden (Nelson, 2011, S. 94).

> When confronted with a security challenge on a deduplicated backup, offer the challenge to reconstruct any particular file, given only the set of unique blocks to work with—a task that is virtually impossible without a map with which to place the blocks in the correct order. (Nelson, 2011, S. 94)

2.7.1 Deduplication für Virtuelle Server

Laut Nelson ist die Verwendung von Source-based Deduplication für virtuelle Maschinen (VM) Hypervisor sehr effektiv. Da es nur eine Handvoll Betriebssysteme gibt, kann hier die Deduplizierung sehr effizient genützt werden. Es ist möglichen, den ganzen Hypervisor zu sichern oder nur eine einzelne VM im Hypervisor. Der Backup Client führt die Überprüfung auf Block-level durch, dies kann eine Belastung aller VM's hervorheben (Nelson, 2011, S. 235).

Nelson (2011) erwähnt noch, dass eine Primäre Technologie für eine Backup Lösung ist:

> By using standard vendor-provided expansion systems, as well as having support models that cover both the hardware and software aspects of the appliance (something that the software solutions do not offer), the appliance solution has a superior support model. This further justifies the recommendation for its use as the primary media storage for backups. (S.235)

2.7.2 Continuous Data Protection / Remote Replication

Laut Nelson gibt es auch noch die Möglichkeit Continuous Data Protection zu verwenden. Dies funktioniert mit sogenannten Snapshots, die zu gewünschter Zeit erstellt und aufbewahrt werden. Somit ist es möglich Backups in wenigen Minuten zu erstellen und auf einem entfernten Standort zu speichern. Die Methode wird CDP genannt Continuous Data Protection und CRR Continuous data repliction genannt. CDP wird für Lokale Datensicherung verwendet, wobei CRR für ein Remote Backup verwendet wird. CDP funktioniert mit einer speziellen Software oder gar mit einer Appliance auf beiden Seiten, indem es die einzelnen I/O Vorgänge auf der Festplattenebene teilt und diese übermittelt (Nelson, 2011, S. 105).

2.8 Definition von RTO und RPO

Wenn eine Backuplösung designet wird, sollte auf drei wichtige Massnahmen geachtet werden:

- Recovery Time Objective (RTO)
- Recovery Point Objective (RPO)
- Service Level Agreement (SLA) (verweis auf Kapitel 2.14)

Der RTO steht für die maximale Zeit, wie lange ein System ausfallen darf oder wie lange es geht bis die Endbenutzer ihre Daten wieder herstellt haben wollen. Es handelt sich um die Zeit des Schadens bis zur kompletten Wiederherstellung aller Daten. Nelson spricht noch die Problematik des RTO an. Der Benutzer und der Systemadministrator haben zwei verschiedene Sichten von RTO, denn der Endbenutzer kann erst wieder arbeiten, wenn die Daten auch in der Applikation sind. Der Systemadministrator hat seinen RTO erfüllt, wenn die Daten wieder im System sind (Nelson, 2011, S.14). Müller definiert den RTO als maximale Wiederanlaufzeit, dies ist die Zeit, die benötigt wird, um die Schutzobjekte bzw. IT-Systeme und deren Applikationen nach einem Ausfall wieder betreiben zu können. (Müller, 2008, S. 448)

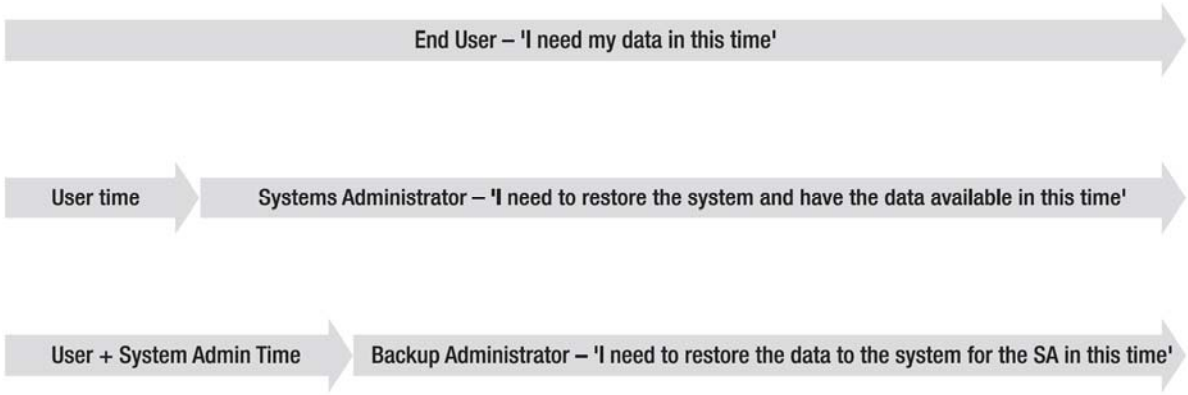


Abbildung 7: Mögliche Sichtweise eines RTO, Quelle: Darstellung entnommen aus Nelson, 2011, S. 14

„Note that each of these viewpoints also contains an implied definition of the end time, or the time at which the data is released to the end user“ (Nelson, 2011, S. 14). Müller erwähnt, dass der Recovery Point Objective den maximal tolerierbaren Datenverlust bis zum Wiederherstellungszeitpunkt angibt. Wenn jeder Tag ein Backup durchgeführt wird, ist der maximale Datenverlust ein Tag. Der RPO legt den maximal tolerierbaren Zeitlichen Abstand zwischen zwei Sicherungen vor (Müller, 2008, S. 455).

> Für korrekte arbeitstägliche Datensicherungen beträgt der Wiederherstellungszeitpunkt dementsprechend einen Arbeitstag. Die Wahl des RPO hängt davon ab, welche Daten, z. B. durch Nacherfassung, wiederhergestellt werden können, welche Kosten dadurch

> entstehen und wie sich die Nutzung des Schutzobjekts und des Geschäftsprozesses dadurch verzögert. (Müller, 2008, S. 110)

Nelson erwähnt, dass der RPO die maximale Datenmange ist, die seit der letzten Datensicherung verloren gegangen werden kann. Ein Datenschutz-Event muss nicht zwingend ein Backup sein, dies können auch andere Arten in Bezug auf den Schutz der Daten sein wie z.B. Snapshots, log dumps oder auch Replikationen. Diese Datenschutz-Events können von diversen Methoden kontrolliert werden, die entsprechende Backupsoftware, die diese Kriterien beinhaltet, vereinfacht das ganze Verfahren. Das primäre Problem beim RPO ist auch, aus welcher Sicht der RPO angeschaut wird, ein System Administrator misst die Zeit anders als der Backup Administrator, dies ist jedoch die Aufgabe der Organisation, dies zu klären bzw. zu definieren. Der RPO oder RTO kann von unterschiedlichen Perspektiven angeschaut bzw. gemessen werden (Nelson, 2011, S. 14).

Nelson (2011) erläutert noch die Abbildung 8 wie folgt:

> For the backup administrator, it will represent the largest amount of time that can elapse between backups (or controlled snapshots) in order to ensure that the data age is appropriately protected: “I need to complete a backup every 4 hours to ensure that only 3 hours of data is lost”. From the perspective of the data owner, this might represent a number of transactions, an amount of data that can be lost, or a particular age of data that can be regenerated: “The organization can afford to lose only the last 30 transactions” (S.14, 15)

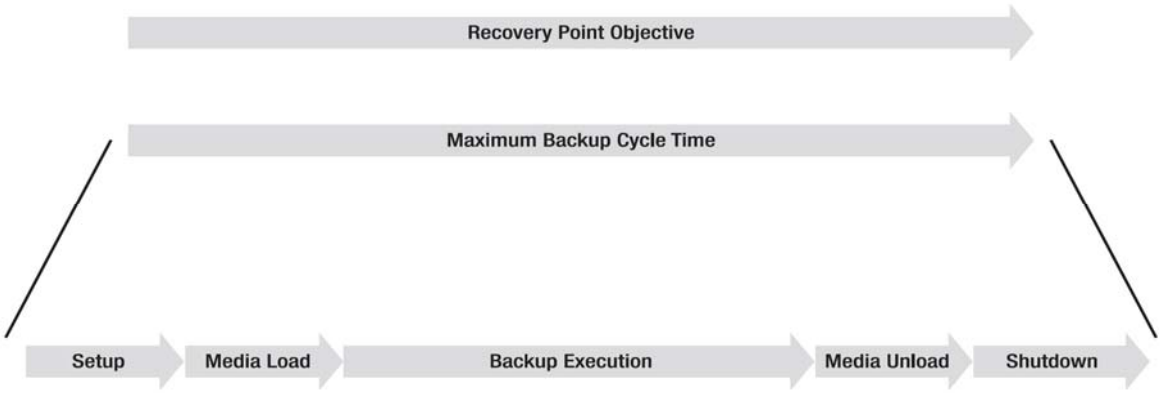


Abbildung 8: RPO Aktivität, Quelle: Darstellung entnommen aus Nelson, 2011, S.14

Nelson fasst nochmals kurz zusammen und erläutert, dass der RTO und RPO technisch nicht verwandt sind. Denn ein RPO kann eine Anzahl von Transaktionen haben, die innerhalb von einer gewissen Periode (RPO) zu schützen sind. Wie lange diese ausfallen darf bis die Daten wieder hergestellt sind, wird somit der RTO bestimmen. In der Praxis ist dies leider nicht der Fall, Nelson (2011) erläutert wie folgt:

> „RTOs tend to be proportionally as short as RPOs. Put another way, if the data is important enough to define an RPO, the RTO will tend to be as short as or shorter than the RPO:
>
> RPO <= RTO

> Although this is not always the case, it is a generalization to keep in mind if an RPO is specified, but an RTO is not.“ (S. 15.)

2.9 eBackup - Backup in der Cloud

2.9.1 Allgemein über die Cloud

Laut Borges und Schwenk werden die Cloud Services immer in drei Bereiche Unterteilt: Bei der Public Cloud sind dies grosse Plattformen wie Azure, Office 365 oder Amazone, die über das Internet genutzt werden können. Bei der Private Cloud werden die IT-Dienstleistungen aus eigenen Rechenzentren bezogen, jedoch kann die Cloud auch von Dritten betrieben werden und der Zugang erfolgt verschlüsselt über das Internet. Bei der Hybrid Cloud ist es eine Mischform zwischen Public und Private Cloud. Die drei Hauptservices sind meistens Infrastructure as a Service (IaaS), welche Rechenleistung und Speicherplatz zur Verfügung stellen. Danach kommt der Platform as a Service (PaaS), welche als Entwicklerplattform genutzt wird. Zum Schluss kommt noch der Software as a Service (SaaS), welche die Nutzung von einer Software über das Internet bereitstellt (Borges & Schwenk, 2012, S. 98). Auch Krutz und Vinces heben den ‚Pay as you go’ Cloud-Ansatz nochmals hervor. Denn die Investitionskosten sind geringer und es muss nur das bezahlt werden, was auch gebraucht wird. Bei diesen Kosten sind die Lizenzen, Hardware, Strom, Personal usw. bereits inbegriffen. Auch die Ressourcen können jederzeit erweitert werden (Krutz & Vines, 2010, S. 21).

2.9.2 Cloud Storage

Nelson betont, dass sich der Cloud Storage von anderen Storages wie IP Storage oder Network File Systemen (NFS) oder Common Internet File Systems (CIFS) unterscheidet. Cloud Storage bietet nicht eine Verzeichnisstruktur wie NFS oder CIFS an. Denn die Daten werden als ‚Generic Blobs’ gespeichert und nicht als spezifische File Typen. Das Protokoll für die Kommunikation ist nicht ein Ressourcen Intensives Protokoll. Das Protokoll, das am meisten verbreitet ist, ist das REST Protokoll, dieses basiert auf HTTP. Somit kann ein grosser Datenverkehr auf das Netzwerk ausgeschlossen werden. Cloud Storage bietet keine verschlüsselte Übertragung an. Diese muss zusätzlich installiert werden, falls dies nötig ist. In der Cloud gibt es keine feste Grösse in Bezug auf Cloud Storage. Der Benutzer zahlt nur den Storage denn er auch tatsächlich braucht. Cloud Storage ist designed für viele Schreib- und Lese-Operationen (Nelson, 2011, S. 109). „So where does cloud storage fit into backup?“ (Nelson, 2011, S. 109)

Der Cloud Storage kann für verschiedene Backup Lösungen gebraucht werden; für eine langfristige Datenablage von Backups oder für eine traditionelle Backup Lösungen.

> Cloud storage is typically cheaper than the physical storage of cartridges in a vault, and infinitely more reliable. A second use of cloud storage is for backups destined for business continuance (BC) and DR use. Since cloud storage is typically IP- and Internet-based, it is effectively accessible from anywhere. If a BC/DR event occurs, restoration from cloud-based backups can be to any server, anywhere in the world—even to virtual cloud servers. (Nelson, 2011, S. 109)
>
> Finally, cloud storage abstracts the media target to the ultimate level. There is not any associated physical entity on which the backups are stored—it is simply a network target that exists ‚somewhere' on the network (the physical locality is largely irrelevant). This abstraction of locality, once fully developed into a mature technology, will again revolutionize the backup world by potentially eliminating local storage entirely in favor of a generic target that has no physical construct. (Nelson, 2011, S. 110)

2.9.3 Backup in der Cloud

Nelson erwähnt Folgendes: Durch all die erwähnten Technologien wie Deduplication, Cloud Storage, Snapshot Technologien usw., ist es heutzutage möglich, ein Backup as a Service (BaaS) zu nutzen. Ein Backup ausserhalb eines Standortes zu nutzen, gibt es jedoch schon lange. Einen BaaS als dedizierten Service zu beziehen und das interne Backup vollständige auszulagern ist jedoch neu. Anhand von Cloud Storage müssen nur die benötigen Ressourcen bezahlt werden und via Internet kann die ganze Infrastruktur gesichert werden. Diese Dienstleistungen sind voll funktionsfähig und ein on-demand Backup und Restore ist jederzeit möglich, das Einzige, was benötigt wird, ist eine Backup Software für die Cloud (Nelson, 2011, S. 203).

> Great! So, why not replace all the backup software with a BaaS and call the backup problem solved? There are a number of issues with BaaS services. The primary one is that of scale. While backup services can handle a number of clients, the ability to ingest backups is largely gated by the connection to the Internet. (Nelson, 2011, S. 203)
>
> Regardless of whether you are backing up data for a cloud or in a cloud, you should at a minimum retain two copies of a backup. At least one of those copies should be located where it is not subject to destruction at the same time that your other copy is located. At minimum, keep it in another room, or better yet store it off-site. (Winkler & Meine, 2011, S. 127

Das Problem ist die Backup Zeit, am Abend kann die meiste Bandbreite des Internets einer Firma genutzt werden, jedoch kann dies eine enorme Datenmenge sein, die übertragen werden muss. Wenn die Kunden BaaS bei den Computern einsetzen, wird eine Art Block-Level Tracking, Real Time File Tagging oder Deduplication genutzt. Diese braucht einen Teil der CPU-Auslastung, die jeder Client bereitstellen muss. Bei der meisten BaaS Software ist es jedoch möglich die CPU-Auslastung zu drosseln. Eine wichtige Erkenntnis, die Nelson anspricht, ist die Internet Verbindung; diese sollte redundant ausgelegt und schnell sein bzw. im Verhältnis zum Backup, ansonsten könnte es mit dem BaaS Probleme geben. Wenn mehrere Files gleichzeitig zurückgespielt werden müssen und die Internet Verbindung am Tag langsam ist, kann es zu Problemen führen. Wenn die Internet Verbindung nicht funktio-

niert, ist es auch nicht möglich ein File zu restoren. Die Services Provider bieten hier auch eine CD / DVD von den gewünschten Files an, wichtig ist jedoch, dass genau das Verzeichnis bzw. File beschrieben wird. Darum ist ein Datensicherungskonzept oder ein Elektronisches File wichtig, was gesichert wird und wann es gesichert wird (Nelson, 2011, S. 203).

Winkler und Meine erläutert, dass der Cloud Service Provider (CSP) vorsichtig ausgewählt wird. Desweitern sollte der Kunde von Anfang an darauf achten, dass es keinen Lock-in gibt, bzw. dass es möglich ist, in eine andere Cloud zu verschieben. Auch auf die Verschlüsselung von eine CSP sollte genau geachtet werden und wie lange die Backup Daten gespeichert werden und wann diese auch wieder gelöscht werden. Winkler und Meine zeigt ein folgendes Beispiel bei der Abbildung 9 eines Backup as a Service Modells (Winkler & Meine, 2011, S. 148).

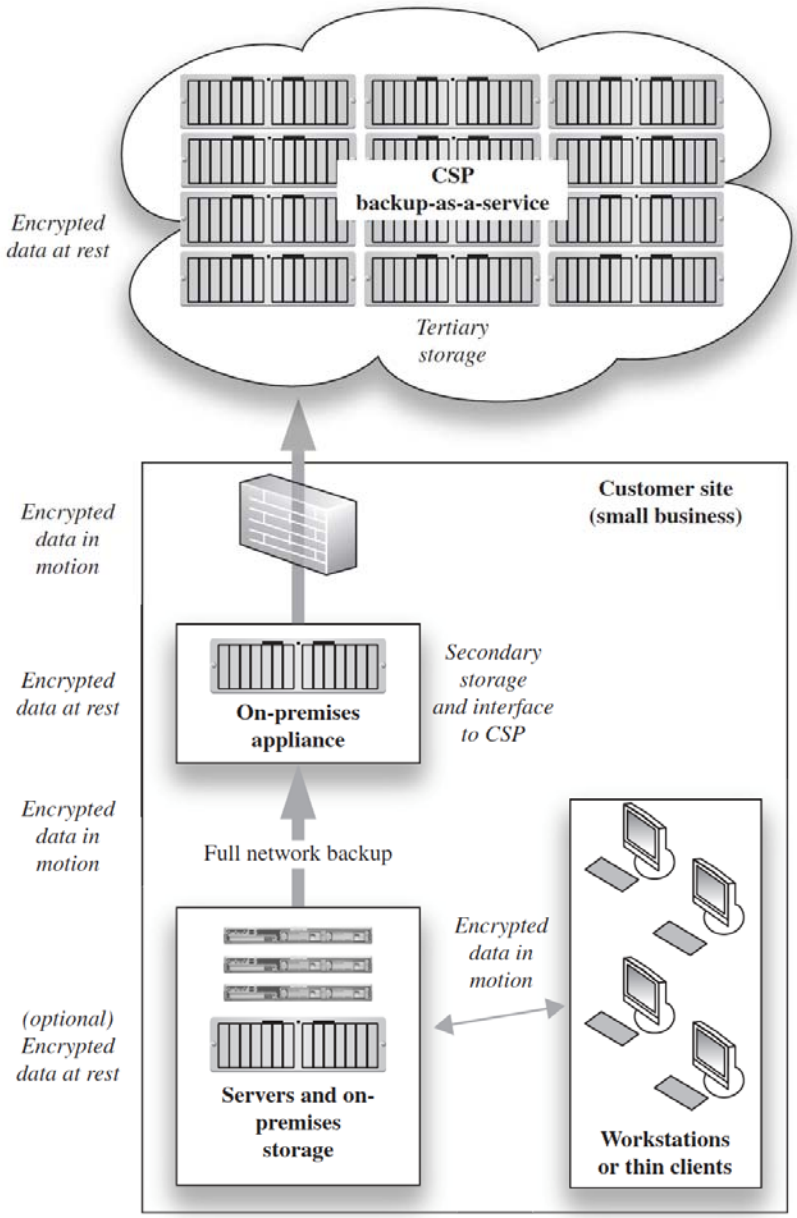


Abbildung 9: Backup as a Service Modell, Quelle: Darstellung entnommen aus Winkler & Meine, 2011, S. 147

2.9.4 Security und BaaS

Nelson erwähnt, dass sich mit dem Internetzugang schnell die Frage bezüglich Sicherheit stellt. Hier ist es wichtig bei der Auswahl eines BaaS Anbieter auf die Sicherheitsstandards zu achten: wie Datenübertragung, Verbindung mit der Software und Backup Access Layers. Der BaaS sollte ein starkes, signiertes Authentifizierung Zertifikat aufweisen. Somit ist es nur dem Service Provider möglich von Aussen auf die Server / Clints eine Verbindung herzustellen. Das Backup sollte natürlich verschlüsselt zum Service Provider übertragen werden, dies verhindert böswillige Attacken von internen oder externen Mitarbeitern, die den Backup-

stream manipulieren könnten. Schlussendlich sollte der Backup Service eine role-based access aufweisen, die einem einzelnen Benutzer die Möglichkeit geben, eine Wiederherstellung selber durchzuführen. Dies ermöglicht dem Backup Administrator die Kontrolle über seine Umgebung, jedoch bietet er dem einzelnen Benutzer die Möglichkeit die Daten selber zu verwalten und bei jeder Anfrage ein administrativer Eingriff ist bei jeder Anfrage nötig (Nelson, 2011, S. 204).

2.9.5 BaaS Kosten

Wie Nelson (2011) erwähnt sind das Wichtigste beim BaaS die laufenden Kosten.

> While, from an accounting perspective, such costs may be able to be categorized as an operational expense (which is generally better to report than a capital expense), these costs can grow without bound and can quickly become more expensive over time than just buying a small backup infrastructure. (S. 204)

Anbieter wie Mozy, Carbonit und Backblaze bieten einen privates oder ein geschäftliches Business Modell an, welches dem Privaten meist eine sogenannte flat-rat (Pauschalpreis) anbietet, weil 90 Prozent der Home User einen einzigen PC benutzt. Jeder weitere PC kostet zusätzlich, es werden jedoch keine Art von Network Attached Storage (NAS) unterstützt. Es muss auch die Art der Daten unterschieden werden, welche geschützt bzw. verschlüsselt werden müssen oder welche nicht (Nelson, 2011, S. 204).

> However, each of these BaaS providers has a pro- or business-level product offering. Depending on the provider, these can be a flat rate per server, per month; a small flat fee plus a charge per GB protected; or simply a GB metered rate. While initially these rates can be very reasonable, it is important to watch these costs, especially as the number of servers and the quantity of data grows because there is not a cap applied to the amount charged, which can lead to some unpleasant monthly bills. (Nelson, 2011, S. 204)

BaaS Advantages	BaaS Disadvantages
• Does not require infrastructure purchases	• Does not scale well over time—each additional client consumes Internet bandwidth
• Complete remote management of Backups	• Can impact client performance during backup
• Can be deployed on desktops and servers	• Ongoing monthly costs may outweigh infrastructure and administrative costs over time.
• Backups can happen from	• Restores can be slow, especially if

anywhere—great for "road warriors"	multiple parallel restores consume available Internet access bandwidth

Tabelle 3: BaaS Advantages and Disadvantages, Quelle: Darstellung entnommen aus Nelson, 2011, S. 205

Mahmood fügt noch hinzu, dass bei einem BaaS es allenfalls mit der WAN Übertragung technische Probleme geben kann. Somit kann auch der RPO oder RTO nicht eingehalten werden (Mahmood, 2013, S.216). Laut Winkler und Meine ist es unabhängig ob die Daten in der Cloud sind oder nicht, es sollte immer eine zweite Backup Kopie angelegt werden. Diese sollte sich jedoch auch an einem anderen Standort befinden. (Winkler & Meine, 2011, S. 127)

> But even a backup can fail when you need it the most, so an even better practice would be to use a cloud-based backup service in addition to your on-site backups. The cost and ease of using such cloud services makes their use very practical if you have reliable network connectivity. Many of these services support encryption of your data before it is sent to the cloud backup service, greatly reducing concern over using such a facility for any but your most sensitive personal information. (Winkler & Meine, 2011, S. 127)

Laut Nelson (2011) gibt ab einer gewissen grösse eine Problem bezüglich Baas Service:

> BaaS starts to become an issue when 200 GB of data or 30 servers, whichever comes first, need to be protected. At this point, the price of storing the data (especially if you are on a metered plan) becomes equivalent to simply buying some backup infrastructure and performing the backups yourself. Of course, this does add overhead to the organization, which for smaller organizations can offset the rising cost of BaaS. But if the BaaS is being used as a backup solution for remote locations, the additional infrastructure cost is negligible as an existing backup solution can be extended to meet the need, as will be described following. From a technical perspective, the network bandwidth required to perform backups for BaaS also starts to become problematic as well, all of which can be absorbed by local resources versus Internet uplinks. (S. 205)
>
> The single backup server is a good solution for small environments that typically run over the 200- GB/30-client constraint discussed for BaaS solutions. It also just so happens that this also represents a good point at which a single backup server environment is best utilized, given a number of factors. (Nelson, 2011, S. 206)

Wie Nelson noch anfügt sind BaaS hervorragende Lösung für Unternehmen die an verschieden Standorten platziert sind so wie auch für Rechenzentren. Diese Lösungen sin bereits ohne die Bereitstellung von zusätzlicher Hardware möglich. Das einzige das Benötigt wird ist eine Internetleitung. Bei einer herkömmlichen Lösung ist eine direkte Verbindung nötig. (Nelson, 2011, S. 206)

> The downside, as previously discussed, is the requirement to manage a second backup solution and the potential ongoing costs associated with storing backups over time. But for many RO environments, this should be considered as the primary solution, given a cost analysis of the alternatives, even with the necessity to manage a second backup solution. (Nelson, 2011, S. 206)

Auch Wald argumentiert durch den hohen Stellenwert der Unternehmensdaten wird viel Aufwand für ein sicheres Backup betrieben. Der Investitionsaufwand zeichnet sich durch die hohen Kosten in Bezug auf Hard- und Software kosten ab. Nicht zu vernachlässigen ist der Schulungen und Trainingsaufwand für die Backup-Systemadministratoren (Wald, 2002, S. 150).

2.10 Datenschutzgesetzt

Wie Patak erläutert gibt es kein Gesetz zur Datensicherung (Backup). Durch die diversen gesetzlichen Anforderungen ergibt sich dies jedoch direkt oder indirekt (Patak, 2010, S. 2).

> In zivilrechtlicher Hinsicht ergehen Anforderungen aus Aufbewahrungspflichten (vor allem Art. 962, Abs. 1 OR[ii], Art. 6 ff. GeBüV[iii]) sowie Sicherungspflichten bezüglich Persönlichkeitsschutz und Geheimhaltung (Art. 28 ZGB[iv] generell, Art. 328b OR für Arbeitnehmer, vertragsrechtliche Geheimhaltungs- und Sorgfaltspflichten). Im Weiteren gelten grundsätzlich für die Durchsetzung jedes zivilrechtlichen Anspruches Art. 8 ZGB, wonach derjenige das Vorhandensein einer behaupteten Tatsache zu beweisen hat, der aus ihr Rechte ableitet. Fehlende Beweisbarkeit bedeutet Rechtsverlust. (Patak, 2010, S. 2)

Laut dem eidgenössischen Datenschutz- und Öffentlichkeitsbeauftragten unterliegen personenbezogene Daten in der Cloud, die durch Dritte bearbeitet werden, dem Datenschutz Gesetz Schweiz (DSG) DSG Art. 10a. Es ist möglich, Personendaten durch Vereinbarung einem Dritten bzw. Cloud Service Anbieter zu übertragen. Diese Daten dürfen aber nur so bearbeitet werden wie der Auftraggeber bzw. Cloud-Nutzer dies selbst erledigen würde, solange es keine gesetzliche oder vertragliche Geheimhaltungspflicht dies verbietet. Der Auftraggeber oder auch Cloud-Nutzer muss sich vergewissern, dass die Datensicherheit durch den Dritten bzw. Cloud-Anbieter gewährleistet ist. Jeder Cloud Service Anbieter muss sich demnach verpflichten, dass er die schweizerischen Datenschutzbestimmen einhält, dies gilt natürlich auch für allfällige Subunternehmen die vom Anbieter einbezogen werden (EDÖB, 2011, S. 3).

> Die Umsetzung dieses Erfordernisses bereitet in der Praxis jedoch Schwierigkeiten, da bei den Cloud-Computing-Anwendungen, die Unterauftragsverhältnisse des Cloud-Service-Anbieters, für den Cloud-Nutzer oft nicht transparent sind. Weiter muss sich der Cloud-Nutzer vergewissern, dass der Cloud Service Anbieter als Dritter die Datensicherheit im Sinne von Art. 7 DSG und Art. 8 ff. bzw. 20 ff. VDSG gewährleistet. Das heisst, die Personendaten müssen durch angemessene technische und organisatorische Massnahmen gegen unbefugtes Bearbeiten geschützt werden. Es muss für Vertraulichkeit, Verfügbarkeit und Integrität der Daten gesorgt sein. (EDÖB, 2011, S. 3)

Dies bedeutet, dass sich der Cloud Service Anbieter gegen Risiken wie Diebstahl, unbefugtes ändern, kopieren, zugreifen oder andere Bearbeitungen schützen muss. Die Umsetzung dieser Schutzmassnahmen hängt von der Art der Daten ab, jedoch ist eine periodische

Überprüfung der Schutzmassnahmen notwendig. Es gibt natürlich auch Unterschiede, ob es sich um eine Private oder Public Cloud handelt (EDÖB, 2011, S. 3).

> Als Grundregel gilt: Je vertraulicher, geheimer, wichtiger (weil geschäftskritisch) oder sensitiver (weil besonders schützenswert) die Daten sind, umso eher ist von einer Auslagerung der Daten in die Cloud, insbesondere eine ausländische Cloud, abzusehen, und desto strikter und umfassender müssen die (Datenschutz-) Sicherheitsvorkehrungen und deren Kontrolle sein. (EDÖB, 2011, S. 3)

Daten ins Ausland zu lagern ist mit Vorsicht zu geniessen, denn Personen -Daten dürfen nicht ausgelagert werden, wenn die Persönlichkeit der betroffene Personen schwerwiegend gefährdet ist, da die Rechtssicherheit im Ausland meistens fehlt und keinen Schutz bietet. (Art. 6 Abs. 1 DSG) (EDÖB, 2011, S. 4). Auch in der Schweiz müssen die Personen kategorisiert bzw. klassifiziert werden, denn es gibt besonders schützenswerte Personen Daten. Im DSG Art. 3. Abs. c. geht es um die besonders schützenswerte Personen Daten, die wie folgt aufgeteilt wurden: „1. die religiösen, weltanschaulichen, politischen oder gewerkschaftlichen Ansichten oder Tätigkeiten, 2. die Gesundheit, die Intimsphäre oder die Rassenzugehörigkeit, 3. Massnahmen der sozialen Hilfe, 4. administrative oder strafrechtliche Verfolgungen und Sanktionen“ (DSG, 2011, S. 2).

„Schliesslich ist der Cloud-Nutzer auch dafür verantwortlich, dass das Auskunftsrecht nach Art. 8 DSG und das Recht auf Löschung und Berichtigung nach Art. 5 DSG jederzeit gewährleistet sind und entsprechend den datenschutzrechtlichen Vorgaben umgesetzt werden“ (EDÖB, 2011, S. 4).

Auch einer der wenigen Bundesgerichtsentscheide in Bezug auf das Datenschutzgesetz fügt folgendes hinzu:

> Gegenstand der Massnahmen ist nicht nur die EDV-geschützte, automatisierte, sondern auch jegliche Form der manuellen Datenbearbeitung. Bezüglich Art und Umfang der zu treffenden Massnahmen gilt das Verhältnismässigkeitsprinzip: Zufolge Art. 8 Abs. 2 VDSG sind insbesondere Zweck, Art und Umfang der Datenbearbeitung sowie der gegenwärtige Stand der Technik zu berücksichtigen und eine Einschätzung der möglichen Risiken für die betroffenen Personen vorzunehmen. (EDÖB, 2012, S. 28)

Laut EDÖB sollte sich der Endbenutzer gut überlegen, welche Daten er am eigenen Standort haben möchte und welche er in die Cloud auslagern möchte. Im Vorfeld wird eine genaue Prüfung des Cloud Anbieters vorgeschlagen (EDÖB, 2011, S. 3).

Broges und Schwenk fügen noch hinzu, dass in den meisten Cloud-Angeboten das rückstandfreie Löschen nicht gewährleistet ist. Auch wenn der Kunde diese Daten löscht, ist es nicht sicher, dass diese schlussendlich gelöscht werden. Denn die Daten können auf verschieden Server und Backup System gespeichert sein. Diese Daten können noch für eine unbestimmte Zeit gelagert werden. Der Lösungsansatz liegt in der Verschlüsselung: Die Daten, die auf der Cloud gespeichert werden, sollten bereits lokal verschlüsselt werden sowie

auch die Übertragung in die Cloud. Die vollständige Verschlüsslung der meisten Daten ist heutzutage jedoch in vielen Fällen nicht möglich. Wenn Daten eines besonderen Amts oder dem Berufsgeheimnis unterliegen, ist besonders hohe Vorsicht geboten, denn beim Verarbeiten der Daten liegen diese meistens im Hauptspeicher als Klartext vor (Borges & Schwenk, 2012, S. 90).

Broges und Schwenk haben sich auch mit dem Rechtsregime des Cloud Service Prodivers auseinander gesetzt. Denn meisten Kunden ist es nicht bewusst, dass ein Rechtsystem des Cloud Service Providers bestimmte Zugriffsbefugnisse der Daten vergeben kann. Anhand der nationalen Sicherheit kann ein Zugriff gewährleistet werden die betrifft meistens lese Zugriff in Bezug auf die rechtlichen Normen (Borges & Schwenk, 2012, S. 91). „Dass Daten von Cloud-Anwendern auf Anfrage der USA gemäss dem Patriot Act herausgegeben werden müssen, betrifft nicht nur US-Firmen als Cloud Service Provider, sondern auch alle deutschen Firmen mit einer US-Niederlassung“ (Borges & Schwenk, 2012, S. 91). Auch die Neue Zürcher Zeitung schreibt folgendes im Juli 2012:

> Dass das Geschäft mit Nutzern, die ausserhalb der USA leben, aus Unternehmenssicht durch den Patriot Act erschwert wird, ist seit längerem bekannt. Unter anderem haben Google und Microsoft ihre Nutzer darauf hingewiesen, dass amerikanische Ermittler gemäss dem Anti-Terror-Gesetz auf die gespeicherten Daten Zugriff haben. Nutzer erfuhren davon nichts, liessen die Unternehmen verlauten. Der eidgenössische Datenschutzbeauftragte Hanspeter Thür bemängelte dies bereits 2006 in einem Referat am Europainstitut Zürich. (Neue Zürcher Zeitung, 2012)

Auch Thür oberster Schweizer Datenschützer nimmt nochmals bezüglich Patriot Act Stellung: „Seit dem Patriot Act können US-Geheimdienste auf private Firmen zugreifen, wenn sie das Gefühl haben, sie bräuchten Informationen. Wer sich mit dem Thema beschäftigt, weiss über Informationsweitergaben von Kreditkarten-, Telekommunikations- oder Transportunternehmen, von Bibliotheken, von sozialen Netzwerken“ (Tages-Anzeiger, 2013b).

2.11 Datensicherungskonzept

Laut Müller werden in den meisten Unternehmen die Datensicherungs- oder Datenschutzkonzepte jedes Mal neu erfunden. Die Qualität der jeweiligen Ergebnisse ist immer unterschiedlich. Die vorhandenen Konzepte werden nicht übernommen oder nur teilweise abstrahiert. Somit gehen Unternehmen Standards verloren, desweitern spielen die Vorkenntnisse von jedem Mitarbeiter eine starke Rolle, ob dieser aus der Unix oder Windows Welt kommt. Je nach Herkunft weisen diese Personen unterschiedliche Sicherheitsniveaus auf. Auch Standards können teure Fehlinvestitionen sein, wenn je nach Betriebssystem das Sicherheitsniveau den Unternehmensanforderungen nicht entspricht. Durch die Vorgaben und Richtlinien eines Unternehmens kann die Qualität und die Sicherheit massiv erhöht werden. Durch die Effizienzsteigerung können auch Kosten gesenkt werden (Müller, 2008, S. 8).

Gadatsch und Mayer teilen die IT Kosten in zwei Kategorien direkte Kosten ein, welche die Beschaffung von Hardware und Software, Schulung der Mitarbeiter, Wartung usw. einbeziehen. Indirekten Kosten können durch Ausfallzeiten oder Fehlfunktionen entstehen (Gadatsch & Meyer, 2010, S. 109). Gadatsch und Meyer (2010) fügen noch folgendes Beispiel hinzu:

> Weitere Beispiele für indirekte Kosten sind Opportunitätsverluste durch Nichtnutzung von technologischen Möglichkeiten (z. B. Datensicherungskonzept, Laufwerke im Netz), deren Nichtnutzung höhere Kosten verursacht, als ihr konsequenter Einsatz. Ein fehlendes Datensicherungskonzept kann zu einem Datenverlust führen, wenn ein Mitarbeiter Unternehmensdaten auf einem Laptop aufbewahrt und diesen verliert. (S. 109)

Gadatsch und Meyer erwähnen auch das Zentrale IT-System wie E-Mailserver, Virenangriff auf das Unternehmensnetz enorme Arbeitsausfälle oder gar Folgekosten durch nicht erfasste Aufträge für die Firma entstehen kann. Wie bei der Abbildung 10 lassen sich diese Kosten mit einem Schiff vergleichen, dessen Rumpf die indirekten Kosten sind - unter der Wasserlinie und nicht sichtbar (Gadatsch & Meyer, 2010, S. 110). „Der Anteil der direkten Kosten erreicht etwa 45% der Gesamtkosten, während die nicht durch das Management beeinflussbaren Kosten bis zu 55 % betragen können“ (Gadatsch & Meyer, 2010, S. 110).

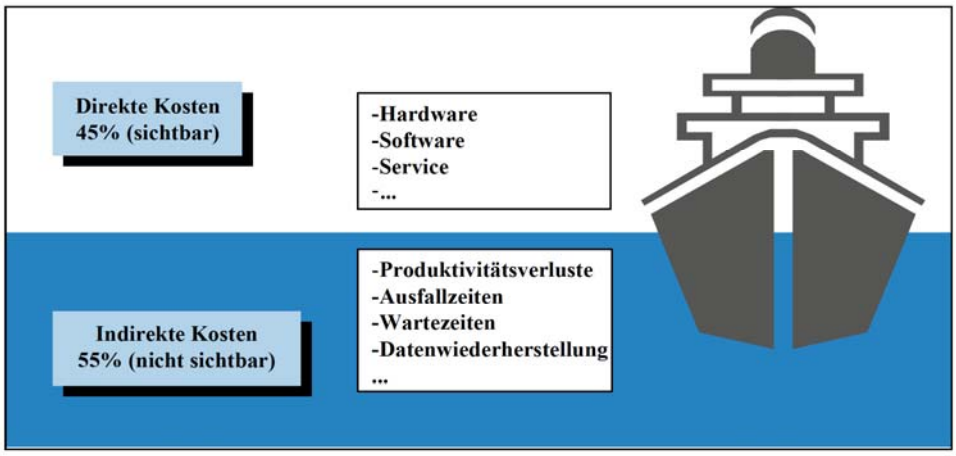


Abbildung 10: Direkte versus Indirekte IT-Kosten, Quelle: Darstellung entnommen aus Gadatsch & Meyer, 2010, S. 110

Laut Wald sollte auch jemand verantwortlich für die Dokumentation gemacht werden und diese auch stets auf dem aktuellsten Stand halten. Des Weiteren sollte diese Dokumentation auch noch ausserhalb des Rechenzentrums aufbewahrt werden (Wald, 2002, S. 175).

Auch Müller beschreibt im Datensicherungskonzept die kompletten Richtlinien. Die Sicherungsmethode und in welchem Rhythmus die Daten gespeichert werden sind zentrale Bestandteile davon. Auch der Lagerort, wo sich die Daten befinden, ist im Datensicherungskonzept wieder zu finden (Müller, 2008, S. 354).

Auch das Bundesamt für Sicherheit in der Informationstechnik (BSI) erwähnt, dass der IT-Leiter oder IT-Sicherheitsbeauftrage für das Datensicherungskonzept verantwortlich ist. „Im Datensicherungskonzept gilt es, eine Lösung zu finden, die diese Faktoren berücksichtigt und gleichzeitig unter Kostengesichtspunkten wirtschaftlich vertretbar ist“ (BSI, 2013).

Beim Datensicherungskonzept wird als Erstes die Gefahrenlage analysiert und aufgezeigt. In einem zweiten Schritt, werden das Datenvolumen und die Sicherungsarten wie Vollbackup und Inkrementell ausgewählt und definiert (wie im Kapitel 2.6.1 beschreiben wurde). Als nächster Schritt muss die Festlegung der Art der Daten bestimmt werden, mit welchem Medium gesichert wird. Die Festlegung der Vorgehensweise und dessen Zyklus wird auch im Datensicherungskonzept niedergeschrieben. Am Schluss müssen die Mitarbeiter bestimmt werden, welche für die Datensicherung verantwortlich sind und die Datensicherung auch regelmässig überprüfen (BSI, 2013).

„Which is the right backup strategy for your virtual machines? We hate to say it, but unfortunately this is a case for the consultant’s answer: it depends“ (Finn, Lownds, Luescher & Flynn, 2013, S. 409). Laut Finn, Luescher, Lownds und Flynn gibt es bezüglich Backup Strategie in Bezug auf die virtuellen Server die Public oder Privat Cloud. Dank den Hypervisor ist die Technology für Backup to Cloud bereits hier. Des Weiteren können die Kunden auch auswählen, welche Daten sie in die Cloud sichern, denn es wird nur das bezahlt, was auch benötigt wird (Finn et al., 2013, S. 409).

Laut Nelson ist das Backup und Recovery Konzept ein Thema, das für viele Menschen auf dem ersten Blick verwirrend ist. Denn Backup und Archivierung werden als Synonyme verwendet, welche eine Art Datenschutz sind, die eine Sicherung von einem Zeitraum sichert und von Zeit zu Zeit zu erneuern sind. Die meisten Firmen haben bei einem Datensicherungskonzept nur den Fokus auf die Sicherung der Daten und nicht auf die einzelnen Funktionen (Nelson 2011, S. 2).

Laut Müller müssen auch Transporte der Bänder gut und genau organisiert werden. „Durch Sicherheitslücken beim Transport gehen einem bekannten amerikanischen Paketdienst Computerbänder einer amerikanischen Grossbank mit Informationen über 3,9 Millionen Kunden verloren. Die Daten umfassen Namen, Sozialversicherungsnummern, Kontonummern und Zahlungsverhalten. Mehrere ähnliche Fälle hatte es in Amerika bereits im Februar, April und Mai 2005 gegeben“ (Müller, 2008, S. 5).

Des Weiteren erwähnt Nelson auch, dass bei einer Backup-Strategie die einzelnen Applikationen und die Methode der Sicherung evaluiert werden müssen. Nur durch eine konkrete Backup Strategie, kann auch ein entsprechendes Recovery funktionieren (Nelson 2011, S. 169).

2.12 Monitoring

Auch Nelson sieht die Überwachung als zentraler Punkt einer Backupumgebung. Ohne Überwachung ist es nicht möglich festzustellen ob die Umgebung funktionsfähig ist und die Sicherung erfolgreich abgeschlossen worden ist. Nur durch das Testen eines Restore kann man davon ausgehen das ein Backup funktioniert hat. Dank einem Monitoring können auch ein Reporting geführt werden welches somit die Vereinbarungen eines Service Level Agreements sicherstellt. Mit Hilfe vom Reporting wird dem Management aufgezeigt das der Backupjob erfolgreich war, somit können im Notfall die Unternehmensdaten zurückgespielt werden. Auch die Soft- und Hardware sollten überwacht werden. Falls ein Tape-Laufwerk benützt wird müssen die Tapes überwacht werden ob diese noch einwandfrei funktionieren (Nelson, 2011, S. 247).

2.13 Verfügbarkeit

Reese erläutert, der SLA ist bei allen Cloud Anwendungen eines der wichtigsten Schlüsselelemente, welches die drei Themen beinhaltet: Verfügbarkeit, Ausfallsicherheit und die Leistung. Die Verfügbarkeit wird anhand eines definierten Zeitraums ausgerechnet. Wenn ein Monat über 720 Stunden verfügt und eine öffentliche Webseite für 710 Stunden erreichbar ist, ist dies eine Verfügbarkeit von 98.6%. Auf dem ersten Blick scheint dieser Wert hoch zusein, jedoch hat dies einen starken Zusammenhang, wie wichtig eine Applikation ist (Reese, 2009, S. 54). Reese erläutert hier das folgende Beispiel: „If, for example, Google's spider is down for 24 hours but you can still search and get results, would you consider Google to be down?“ (Reese, 2009, S.54) Die meisten Personen sagen somit das Hochverfügbarkeit 99.99% bis 99.999% ist, dies bedeutet bei 99.999%, dass ein System oder Applikation gerade mal für fünf Minuten und 15 Sekunden pro Jahr nicht erreichbar sein darf. G. Reese erläutert auch, dass Glück nicht Hochverfügbarkeit ist, bei den meisten Kunden hat der sogenannte Tag X noch nicht eingeschlagen, das heisst jedoch nicht dass das System hochverfügbar ist, weil noch nichts passiert ist. Die Hochverfügbarkeit wird von Jahr zu Jahr wichtiger für die Firmen als auch Endkonsumenten (Reese, 2009, S. 54). Um die Verfügbarkeit eines Systems auszurechnen, erläutert Reese (2009) folgende mathematische Formel die angewendet werden kann:

$$a = (p- (c \times d)) / p$$

Legende:

a = erwartete Verfügbarkeit
c = Wahrscheinlichkeit in %, dass ein Server Ausfall, in einem bestimmten Zeitraum auftreten wird

d = erwartete Ausfallzeit bei Server Ausfall

p = die Messperiode

Hier noch ein Beispiel: Bei einem Server der 40% Chancen hat, dass er einen Ausfall aufweist und danach 24 Stunden nicht erreichbar ist, wäre diese Betriebszeit wie folgt:

(8760 - (40% x 24)) / 8760, oder in Prozent ausgedruckt 99,9%. (S. 55)

Müller sagt, die Verfügbarkeit bestimmen die Kosten, die durch die Ausfallzeit entstehen können. Somit muss sich eine Firma die Frage stellen, wie lange ihr IT System maximal nicht verfügbar sein darf und wie viele Male es pro Jahr ausfallen darf. Ein weiterer Punkt ist noch, in welchem minimalen zeitlichen Abstand ein IT System bei einem Vorfall erneut ausfallen darf. Hierbei ist für jede Firma bei der Verfügbarkeit, die Kosten-Nutzen-Überlegung, die ausgerechnet werden muss. Bei deren Berechnung müssen als Erstes immer die Single Point of Failure analysiert werden, allenfalls können durch geringe Kosten diese redundant erschlossen werden (Müller, 2008, S. 196).

Verfügbarkeit in Prozent	Max. Ausfallzeit pro Jahr, gerundet in Abhängigkeit von der Betriebszeit			Bezeichnung
	7 Tage x 24h	7 Tage x 12 h	5 Tage x 12 h	
99,0	4 Tage	2 Tage	1,5 Tage	
99,5	2 Tage	1 Tag	16 h	
99,9	9 h	5 h	3h	Verfügbar
99,99	1 h	0,5 h	9,3 h	Hochverfügbar
99,999	6 min	3 Min	2 min	Höchstverfügbar
99,9999	32 sec	16 sec	11 sec	Unterbrechungsarm
...	...	...		...
100	0 sec	0 sec	0 sec	Kontinuierlich verfügbar, unterbrechungsfrei

Tabelle 4: Verfügbarkeit's Tabelle, Quelle: Darstellung entnommen aus Müller 2008, S. 197

Müller sagt auch, dass hier die Datensicherung das Kernelement der Verfügbarkeit ausmacht und bei allfälligen Fehlern die letzte Datensicherung zurück gespielt wird (Müller, 2008, S. 196).

2.14 Service Level Agreement

Auch in Bezug auf ein Backup to Cloud stehen SLA's im Mittelpunkt. Mahmood und Hill schreiben, das SLA (Service Level Agreement) ist der Kernpunkt einer Cloud Strategie. Das SLA ist eine Vereinbarung zwischen dem Dienstleister und dem Auftraggeber, welches die Schnittstellen der beiden Parteien bildet. Für den Auftraggeber sollten die ganzen Leistungen transparent gemacht und detailliert nieder geschrieben werden (Mahmood & Hill, 2011, S. 85). Müller nennt dies Service Level Management, dieser umfasst den ganzen Service Katalog. Neben dem SLA wird auch noch der OLA (Operational Level Agreement) berücksichtigt. Beim OLA werden die internen Dienstleister berücksichtigt, bei den externen Dienstleistern wird dies Underpinning Contracts UC genannt. Zum SLM gehört auch die Überwachung und Steuerung von den internen oder externen Lieferanten. Der sogenannte SLR (Service Level Requirements) gibt die jeweiligen Kundenanforderungen vor, der Service Manager verhandelt diese Verträge mit dem Endkunden (Müller, 2008, S. 172).

Hier werden die wichtigsten Punkte von Mahmood und Hill erläutert:

- Monitoring
- Prozesse
- Überwachung der Kapazität
- Datensicherheit
- Datenschutz
- Privatsphäre der Daten
- Betriebliche Integrität
- Schwachstellen Management
- Business Continuity
- Disaster Recovery
- Identitäts Management
- Eigentum
- Tätigkeitsbereich
- OLA
- Abgrenzungen
- Unterzeichnung und Gültigkeit

Diese Aufzählungen, beschreiben Mahmood und Hill, sind zentrale Punkte bei einem Cloud SLA (Mahmood & Hill, 2011, S. 85). Wieder, Butler, Theilmann und Yahyapour sehen dies jedoch anders und setzen sich mit den systematischen Geschäftsanforderungen von einem SLA auseinander. Bei den meisten Kunden ist die heutige Situation so, dass alle Risiken und Herausforderungen gelöst werden sollen. Die SLA müssen verhandelbar sein, dass sich dem Kunden die jeweiligen Risiken und Reaktionszeiten aufzeigen lassen. Die Kundenanforderungen müssen jedoch hierfür zuerst aufgenommen werden. Die Qualität des SLA lässt sich nur schwer beschreiben. Für die Autoren Wieder et al. gibt es zwei Bedürfnisse bezüglich eines funktionierenden Service Level Agreement. Das Erste ist, dass die gemeinsamen Bedürfnisse des Kunden und des Anbieters analysiert und niedergeschrieben werden. Nur so können beide einen gemeinsamen Weg Richtung Service Level Agreement gehen. Das zweite Bedürfnis hierfür ist, dass die SLAs systematisch verwaltet werden um einen transpa-

renten Weg für Kunde und Anbieter aufzuzeigen. Nur so kann eine lange Zusammenarbeit weitgehend garantiert werden (Wieder, Butler, Theilmann & Yahyapour, 2011, S.5).

In Bezug auf Backup und Recovery muss dies in einem SLA detailliert niedergeschrieben werden. Wenn ein Recovery Fall eintritt, muss definiert sein, wie lange es dauert, um das Backup zur Verfügung zu stellen und zurück zu spielen. Des Weiteren muss ein Reporting über den ganzen Status geben werden. Das Reporting zeigt die Informationen über den Bericht und die benötigte Zeit. Die benötigte Zeit wird anhand von einem Monitoring überwacht. SLA Verstösse können durch ein Monitoring aufgezeigt werden und die entsprechenden Lösungen gefunden werden. Jedes SLA muss die wirtschaftliche Situation des Unternehmens berücksichtigen und miteinbeziehen (Wieder et al., 2011, S.246). Furht und Escalante argumentieren, dass ein Service Level Agreement den entscheiden Faktor für die Gewinnung eines Projektes beeinflussen kann, wenn die technischen Anforderungen bezüglich den Kundenbedürfnissen stimmen. Jedes SLA verfügt über Informationen wie z.B. ein Profil der Last gegenüber der Reaktionszeit stellt. Nur so kann garantiert werden, dass auch bei einem Ausfall der Server unter hoher Last noch die entsprechend definierte Ansprechzeit wieder gibt (Furht & Escalante, 2010, S. 28). Nelson fügt noch hinzu, dass neben dem RPO und RTO auch beschrieben werden muss, welche Personen zwingend anwesend sein müssen falls ein Restore eintritt (Nelson 2011, S. 13).

2.15 Physische Sicherheit

Der Name Rechenzentrum (engl. Data Center) hat sich in den letzten Jahrzehnten massiv gewandelt. Früher haben Grossrechner viel Platz gebraucht und viel Lärm verursacht, diese Grossrechner wurden in dedizierten Räumen, sogenannten Rechenzentren, untergebracht. Durch den ganzen Informatik Hype in den 1990er Jahren haben sich die IT Grundlagen verändert, denn jetzt konnten ganze Geschäftsprozesse und Businessmodelle abgebildet werden. Das Internet hat die nötige Verbindung hergestellt. Durch die Abhängigkeit der IT wurden die Data Center immer wichtiger und eine Ausfallsicherheit musste immer mehr gewährleistet werden. Durch das ganze Datenwachstum wurde auch das ganze Datenbackup immer zentraler. Totalausfälle werden durch Datenspiegelung (Disaster Recovery) – Pläne auf ein Minimum reduziert (Hauri D., Mohler L, Deiniger S., 2012, S. 7).

Auch Wolfgang J. Fiedl befasst sich mit dem Totalausfall der IT und dessen Risiken. Bei der Sicherheit geht es um zwei Hauptthemen - die technischen und die organisatorischen. Diese zwei Hauptthemen müssen harmonieren, ansonsten entstehen hohe Folgekosten. Auch Fiedl argumentiert, dass viele Unternehmen bei einem Schaden oder dessen Folgekosten nach wenigen Tagen ruiniert sind. Je nach Branche ist die Länge des Überlebens verschieden, bei den meisten Unternehmen sind dies jedoch Tage oder maximal Wochen, die ein

Unternehmen ohne IT überleben kann. Fiedl schliesst auch den Schaden für Dritte nicht aus, wenn ein Gebäude z.B. brennt oder das eigene Gebäude kontaminiert wird, kann dies Auswirkungen auf dritte Unternehmen haben (Friedl, 1998, S.4).

> Hier setzt die Sicherheitswissenschaft an: Es gilt, ganzheitliche Sicherheitsanalysen zu erstellen und umzusetzen, die alle Firmen individuellen Abhängigkeiten und Gefährdungen in Qualität und Quantität erfassen; die daraus resultierenden Sicherheitskonzepte müssen in sich geschlossen und gleichwertig sein, wenn sie wirklich funktionieren und greifen sollen. (Friedl, 1998, S. 4)

Fiedl spricht auch die Problematik an, dass sich bereits realisierte Sicherheitsmassnahmen nur schwer darstellen lassen. Nur wenn ein Schaden eintritt, wird er finanziell dargestellt. Wenn ein Schaden nicht eintritt oder verhindert wurde, hat er nie stattgefunden oder wurde nicht wahrgenommen. Im Nachhinein ist es meist unmöglich nachzuweisen, um wie viel grösser der Schaden gewesen wäre, wenn die bestehenden Sicherheitseinrichtungen nicht gewählt worden wären. Die Schadensstatistiken der Industrieversicherer zeigt, dass die meisten Schäden durch Sturm, Feuer, Wasser, Vandalismus und andere Gefahren verursacht werden können. Die direkten Schäden machen oftmals einen kleineren Anteil aus als die Folgeschäden wie Betriebsausfälle. Viele Unternehmen haben bemerkt, dass professionelle Lösungen oftmals preiswerter sind als nichts unternehmen und dessen Folgen tragen zu müssen (Friedl, 1998, S.5). „Aber da auch extrem hohe Sicherheitsauflagen nicht absoluten Schutz und ständige Verfügbarkeit der EDV garantieren können, investieren manche Unternehmen in interne oder externe Backup-Konzepte“ (Friedl, 1998, S.5).

Wie die Abbildung 11 von Friedl aufzeigt, ist der Nutzen von sicherheitstechnischen Massnahmen in der Relation zu den Kosten am Grössten. Auch das Gegenteil ist möglich - dass eine sehr kleine sicherheitstechnische Verbesserung fast 100% abdeckt, jedoch ist es auch möglich, dass dies wiederum sehr teuer werden kann. Nach Friedl sind nur 1/3 aller Rechenzentren gegen Einbruch, Diebstahl, Sabotage oder Vandalismus geschützt. Daher sind Überwachungssysteme sehr wichtig für ein Rechenzentrum. (Friedl, 1998, S.14)

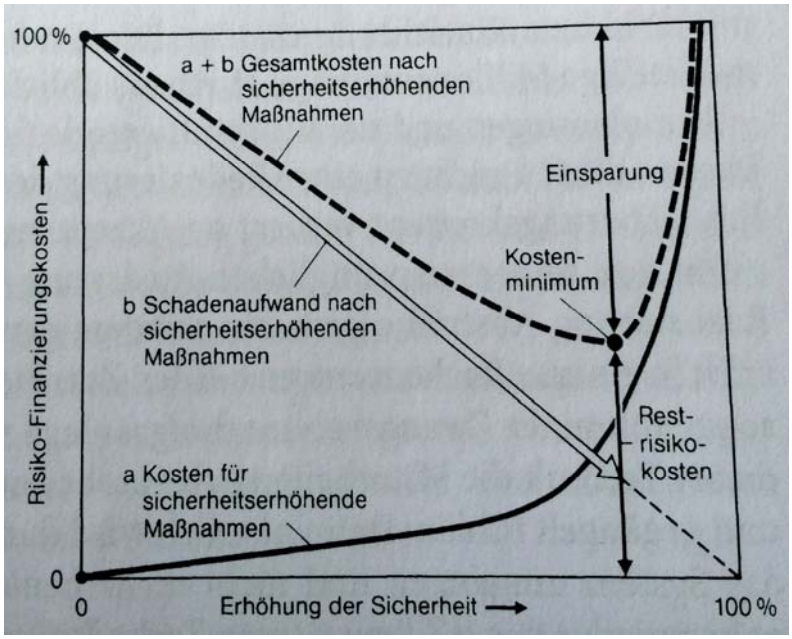


Abbildung 11: Finanzierungskosten eines Risikos, Quelle: Darstellung entnommen aus Friedl, 1998, S.15

2.16 Sicherheitsaspekte

Wie Skurk erwähnt, konnten Unternehmen früher mit einen mehrstündigen IT Ausfall überleben. Die kontinuierliche Verfügbarkeit der IT ist heutzutage unverzichtbar (BITKOM, 2013, S. 8). „Wie hoch sind die maximalen, tolerierbaren Ausfallzeiten der IT des Unternehmens?“ (BITKOM, 2013, S. 8) Bevor die technischen Komponenten ausgelagert werden, muss die Standortwahl getroffen werden, welche geographisch und politisch ausgewählt werden sollte. Wie Skurk erwähnt unterteilt das Uptime Institute in den USA die Verfügbarkeit in Tier Klassen. Die Unterteilung geht von Tier I bis Tier IV, wobei Tier I beispielsweise 99,671% Verfügbarkeit gewährt und Tier IV gar 99,995%Verfügbarkeit. Dies wird durch mehrfach redundante Komponenten erreicht. Das BSI unterteilt diese in 5 Klassen von VK 0 (95%) bis VK 5 100% kumulierte wahrscheinliche Ausfallzeit pro Jahr ist 0 Minuten. Bei der VK 4 ist es gerade einmal noch ca. 5 Minuten. Neben redundanten Kühlungssystemen, USV, Notstromdiesel Generatoren, sollen auch die Energieeffizienz stark berücksichtigt werden (BITKOM, 2013, S. 8).

Die berechnete Formal dafür ist wie folgt:

$$\text{Verfügbarkeit (in Prozent)} = \left(1 \cdot \frac{\text{Ausfallzeit}}{\text{Produktionszeit} + \text{Ausfallzeit}} \right) = 100$$

- 99,99 % * 52,6 Minuten/Jahr
- 99,999 % * 5,26 Minuten/Jahr
- 99,9999 % * 0,5265 Minuten/Jahr

Abbildung 12: Formel für die Verfügbarkeit, Quelle: Darstellung entnommen aus BITKOM, 2013, S. 8

Laut Müller lassen sich durch die Redundanz Single Points of Failure vermeiden. Die Redundanz unterteilt Müller in die fünf Kategorien wie Redundanzkategorien, Strukturelle Redundanz, Redundanzgrad, Redundanzgeschwindigkeit und die Redundanzqualität. In Fachkreisen wird meist über die N+1-Redundanz gesprochen. Bei der N+1 werden gleiche Komponenten parallel betrieben wie USC; Klimaanlage, Transformatoren, Dieselgeneratoren. Sobald eine dieser Komponenten ausfällt wird diese Aufgabe übernommen. Neben der Redundanz spricht man auch von der Latenzzeit, wie viel Zeit vergeht bis die Aufgabe übernommen werden kann (Müller, 2008, S. 144). Laut Friedl ist auch neben der Verfügbarkeit der physische Zugang zu den Daten ein zentraler Sicherheitspunkt. Schlussendlich kann man sagen, dass die Qualität der schwächsten Stelle über die Sicherheit der Daten entscheidet. Zu den Schutzmassnahmen eines modernen Rechenzentrums gehört Wachpersonal mit Zutrittskontrolle. Im Extremfall ist auch mit einem Einbruch zu rechnen, gar Sprengladungen könnten zum Einsatz kommen. Somit ist es wichtig, dass die Grundstücksumhüllung mit Zaun, Mauer und Tor gesichert ist. Es sollten auch nur autorisierte Personen das Areal und Gebäude betreten um das Risiko zu minimieren. Eine Aufzeichnung aller Aktivitäten auf dem Gelände ist zwingend (Friedl, 1998, S. 64).

3 Auswertungsteil

Um eine Auswertung vorzunehmen zu können, werden im folgenden Teil die Kriterien der Ausganslage nochmals überprüft. Zu den Kriterien zählen die Datensicherheit, der Unterschied von konventionellen Backups und eBackups bzw. Backup as a Service, die Auswirkungen auf das Datenschutzgesetz und die Sicherheit eines Rechenzentrums.

Müller und Wald haben mehrfach wiederlegt, welche enormen Kosten ein Datenverlust oder Diebstahl haben kann. Ein Imageschaden oder ein Verlust von einzelnen Dateien ist noch verkraftbar, wenn jedoch ganze ‚überlebenswichtige' Systeme nicht mehr wiederhergestellt werden können, kann dies im schlimmsten Fall zum Konkurs einer Firma führen (vgl. Müller, 2008; Wald, 2002). Autoren Brooks, et. Al. erläuterten, dass ein Backup eine Art Versicherungspolice ist. Erst wenn Probleme auftreten, ist sich eine Firma der Notwendigkeit eines Backups bewusst.

Die Studie von IDC zeigt auf, dass die Firmen in den nächsten Jahren mit einem exponentiellen Datenwachstum rechnen müssen, Big Data steuert dabei bereits heute einen erheblichen Teil dazu. Ohlhorst hat jedoch erwähnt, dass diese massiv steigende Datenmenge in einer Backup-Strategie berücksichtigt werden muss (vgl. Brooks et., al 2003; Gantz & Reinsel, 2012; Ohlhorst, 2013).

Bei der traditionellen Datensicherung gibt es verschiedene Methoden der Datensicherung. Bei der differentiellen Datensicherung sehen die Autoren Müller und Nelson dieselben Vorteile, das kurze Backupzeitfenster gegenüber einer kompletten Datensicherung. Bei einem Restore hingegen sieht Nelson den Vorteil beim Differentialbackup, denn es braucht nur ein Vollbackup und die Differenz dazu. Müller hingegen sieht diesen Schritt bezüglich Restore als Nachteil, denn diese Daten müssen wieder von der differenziellen Datensicherung überschrieben werden. Nelson erwähnt jedoch im Gegensatz zu Müller noch folgendes Problem, durch die Veränderung kann ein Differentialbackup schnell grösser sein, als die komplette Datensicherung. Bezüglich inkrementeller Datensicherung sind sich Müller und Nelson einig, Nelson sieht die inkrementelle Datensicherung als die meist verwendete Datensicherung. Nelson fügt noch das Level Based Backup hinzu, wobei Müller das Vater-Sohn-Prinizipt erwähnt (vgl. Müller, 2008; Nelson, 2011).

Nelson erwähnt, dass ein Tape-Laufwerk das günstigste Medium im $/TB ist, da die Einstiegskosten gering sind und die Daten einfach an einen anderen Ort transportiert werden können. Nelson sieht den Nachteil jedoch im Medium selbst, da dieses sehr komplex ist und noch mechanisch funktioniert und die Zuverlässigkeit relativ gering ist (Nelson, 2011, S.48).

Laut Nelson ist die Deduplizierung eine der neusten Technologien, welche die Backup-Welt revolutionieren wird. Aufgrund des enormen Datenwachstums, auf das auch die Studie von Ohlhorst verweist, wird die Deduplizierungs-Technologie einen zentralen Einfluss in der Backupwelt einnehmen. Osuana et al. und Nelson erwähnen das die Statistik bzw. das Verhältnis je nach Daten unterschiedlich ausfallen kann und je nach Produkt ein anderes Verhältnis hervorruft. Osuana A. et al. sehen bei der Deduplizierung auch einen Vorteil beim TCO, durch die diversen Einsparungen beim Storage, Strom und einem besseren RTO. Bei der Deduplizierung auf dem Hypervisor hat Nelson klar von der Block Level Deduplizierung gesprochen und diese als primäre Backup Technologie für die Zukunft genannt (vgl. Nelson, 2011; Ohlhorst 2013; Osuna et al., 2011).

Gadatsch und Meyer haben erwähnt, dass ohne ein Datensicherungskonzept Verluste von Daten entstehen können. Diese Kosten sind indirekte Kosten und können bis zu 55% der IT Kosten ausmachen. Wie Müller, Wald, Hoppe & Priess erwähnten, können die Kosten eines Datenverlustes sehr hoch ausfallen oder gar das Überleben der Unternehmung gefährden (vgl. Gadatsch & Meyer, 2010; Hoppe & Priess, 2003; Müller, 2008; Wald, 2002). Die konkreten Themen eines Datensicherungskonzeptes des BSI sind im Anhang aufgeführt (BSI, 2013).

Eine der wichtigsten Erkenntnisse von Nelson ist, dass nur durch die neuesten Technologien wie Snapshot, Cloud Storage, Deduplication und Deduplication auf Hypervisor Ebene ein BaaS realisiert werden kann. Nelson sieht den Vorteil bei BaaS, da es weniger Initialkosten braucht im Gegensatz zu einem herkömmlichen Backup. Ab einer Grösse von 200GB oder 30 Servern ergibt laut Nelson ein konventionelles Backup aus Kostensicht mehr Sinn, da bei einer BaaS Lösung die RTO und RPO Werte allenfalls nicht mehr eingehalten werden können. Somit muss der Fokus bei BaaS sicherlich auf die Erfüllung der RTO und RPO Werte liegen. Dies könnte wie Nelson erwähnt bei einer BaaS Lösung je nach Datenleitung ein Problem werden. In Bezug auf die Internetleitung könnte eine externe Disk in einem Notfallszenario eine Lösungsoption sein. Bei einem Baas wird immer mehr nach dem „pay as you go“ Ansatz bezahlt. Mit Hilfe von Deduplication können die Kosten gesenkt werden, falls die vom Cloud-Anbieter auch unterstützt wird. In Bezug auf die Sicherheit empfiehlt auch der EDÖB eine Verschlüsslung der Daten zu verwenden, bei einer BaaS Lösung wäre eine optimale Lösung, wenn die Verschlüsselung vor der Übertragung schon geschehen würde (vgl. EDÖB, 2011; Nelson, 2011).

Der EDÖB gibt relativ klare Richtlinien bezüglich des Datenschutzes in Verbindung mit den Cloud-Services vor. Wenn ein Backup as a Service bezogen wird, sollte sicherlich auf die Verschlüsslung geachtet werden. Wenn der Service von der Schweiz bezogen wird, ist die Rechtssicherheit gegeben, im Gegensatz zum Ausland. Wie Thür erwähnt hat, sollten ameri-

kanische Tochtergesellschaften besonders genau analysiert werden, denn der Patriot Act kommt auch dort zum tragen (Tage-Anzeiger, 2013a). Wie Reese und Mahommd & Hill erläutern, ist neben der Verfügbarkeit des Services der SLA einer Cloud Anwendung ein zentraler Bestandteil. Nur durch den SLA und OLA können genau die Grenzen zwischen Anbieter und Nutzer gezogen werden und schliessen offene Fragen somit aus. Eine Cloud Anwendung sollte auch in einem Tier III oder in der BSI Klasse 4 oder 5 gehostet werden, denn die Daten müssen von den allfälligen Gefahren geschützt werden. Durch die N+1 Redundanz bei den Rechenzentren sind diese durch allfällige Ausfälle sehr gut abgedeckt (vgl. BITKOM, 2013; Mahmood & Hill, 2011; Reese, 2009).

3.1 Prüfung der Hypothese

Die aufgestellte Hypothese wird aus folgenden Gründen falsifiziert:

Nelson behauptet, dass sich ein eBackup bzw. Backup as a Service aus Kostensicht nur bei einer Datenmenge unter 200GB oder einer Infrastruktur unter 30 Servern lohnt. Eine lokale Backup-Lösung kostet selbst unter der Berücksichtigung der Umlaufkosten einer Firma weniger.

Gemäss Winkler und Meine muss ausserdem beim Entscheid für ein eBackup zwingend mindestens ein zweites, lokales Backup erstellt werden, was die Sicherheit erhöht, jedoch auch die Kosten für die Gesamtlösung in die Höhe treibt.

Zur Berücksichtigung aller Sicherheitsfragen, ist ein Datensicherungskonzept laut Gadatsch & Meyer für eine Firma notwendig, um ein Datenverlust zu vermeiden und um die indirekten Kosten, welche einen Datenverlust mit sich bringt, so gering wie möglich zu halten.

Fügt man die Aussagen von Nelson, Winkler und Meine sowie Gadatsch & Meyer zusammen, so reicht es für eine Firma nicht aus, sich nur aufgrund von Kostenfragen für ein eBackup zu entscheiden. Für eine Entscheidung ist es sehr wichtig, auch Faktoren wie die Anforderungen an die RTO und RPO Werte, der SLA‘s und OLA‘s des Service Anbieters sowie der Datensicherheit zu berücksichtigen (vgl. Gadatsch & Meyer, 2010; Nelson, 2011; Winkler & Meine, 2011).

3.2 Praxisbezug auf eBackup

Die Preise eines eBackups können je nach Firma variieren. Bei der Bestimmung des Preises muss auch immer die Ausganslage des jeweiligen Unternehmens berücksichtigt werden. Wie die folgende Tabelle aufzeigt, wurde eine nicht repräsentative Umfrage bezüglich eBackup bzw. Backup as a Service erstellt, um die Angebote von Schweizer Anbieter aufzuzeigen. Welche Firmen angefragt wurden, ist im Anhang ersichtlich. Die Kosten bezüglich einem

lokalen Backup variieren nach Unternehmen, zwei Firmen werden jedoch auch noch im Anhang ausgewiesen:

Firma	Cloud Backup	DR	Virtuell & Physikalische Server	Deduplizierung	Internetleitung	Preis pro TB (Monat)	Software
Green	Nur als Projekt möglich	-	-	-	-	-	-
nexellent	Nur als Projekt möglich	-	-	-	-	-	-
swissbackup24.ch	Ja	Nein	Ja	Ja	Standard DSL	CHF 442.00	Onbackup.de
ProCloud	Ja	Ja	Ja	Ja	Standard DSL	CHF 457.80	-
Nugolo	Ja	Ja	Ja	Ja	Standard DSL	CHF 499.00	-
Backupswiss	Ja	Nein	Ja	Nein	Standard DSL	CHF 800.00	Duplicati (open source) oder Langmeier Backup
DataTrust AG	Ja	Nein (in planung)	Ja	Ja	Standard DSL	Preis Pro Server CHF 120.- pro Monat Bei 5 Servern sind es CHF 183.-	Telebackup
Arcplace	Ja	Ja	Ja	Ja	Standard DSL	CHF 1081.00	asigra.com
Informatio	Ja	Ja	Ja	Ja	Standard DSL	Preis Pro Server CHF 20.- + 1TB CHF 409.-	Cetra
Netkom	Ja	Ja	Ja	Ja	Standard DSL	CHF 266.66 (CHF+ 3500.-) Ein-	Cetra

						malig	
Wolfsync	Ja	(mög-lich) mit Appli-ance	Ja	Ja Block-basierend	Standard DSL	Einfache CHF 250.- Business CHF 350.00	-

Tabelle 5: BaaS Provider aus der Schweiz, Quelle: Eigene Darstellung, 2013

3.3 Inhaltliche Abgrenzung

Bei dieser Arbeit geht es nicht darum die verschiedenen Produkte, welche auf dem Markt bestehen, zu untersuchen oder eine Software zu evaluieren. Vielmehr sollen die Gefahren wie Datenhaltung, Datenschutz, Datenübertragung des eBackups aufgezeigt werden. Dabei soll lediglich die Gesetzgebung der Schweiz im Fokus stehen. Der Hauptfokus bezüglich Gesetzgebung liegt hier bei den KMU's. Die Gesetze für Banken/Versicherungen werden in dieser Arbeit nicht berücksichtig und somit Banken/Versicherungen als Kunden ausgeschlossen. Unterschiede in den Kosten der einzelnen Produkte werden ebenfalls nicht berücksichtigt oder untersucht. Es werden auch keine Datenablagen wie (Dropbox, Wuala, Google Drive usw.) mit eBackup verglichen, denn es geht um Server und Storage Daten die gesichert werden müssen. Es wird jedoch der Unterschied zwischen einem traditionellen und einem neuzeitigen eBackup verglichen. Clients wie Notebook oder Workstations werden nicht als zentraler Fokus behandelt. Virtual Tape Library (VTL) sowie Disaster Recovery werden in dieser Arbeit nicht erwähnt oder verglichen.

4 Konklusion und Ausblick / Schlussfolgerung

Die Sicherheit der Daten hat einen immer höheren Stellenwert bei jeder Unternehmung. Ein Datenverlust oder einen kompletten Datenverlust kann sich fast keine Firma mehr leisten. Bei der konventionellen Datensicherung können diverse Sicherungs-Technologien verwendet werden. Wichtig ist dabei, dass ein Datensicherungskonzept vorhanden ist. Auch wenn eine BaaS verwendet wird, sollte auch auf die redundante und performante Internetleitung geachtet werden.

Ziel dieser Arbeit war die Unterschiede zwischen einem traditionellen Backup und einem e-Backup aufzuzeigen. Am Anfang der Arbeit wurde die Grundlage der Datensicherung erarbeitet, warum eine Datensicherung überhaupt erstellt werden muss. Nachfolgend wurden die Technologien bzw. Methoden der lokalen bzw. traditionellen Datensicherung verglichen. Im Weiteren wurden auch mehrere Varianten der Daten-Deduplizierung verglichen, da diese als wichtige Grundlage für die heutigen BaaS Angebote dient. Dies wiederspielgelt sich auch bei der nicht repräsentative Umfrage. Bei einem BaaS Angebot müssen jedoch auch zwingend Themen wie SLA, OLA und der Standort des Rechenzentrums berücksichtig werden.

Je nachdem, wo sich die Daten befinden, gelten auch die entsprechenden Datenschutzgesetze. In der Schweiz müssen die entsprechenden Kriterien überprüft werden und auch der Patriot Act darf hierbei nicht ausser Acht gelassen werden.

Das Ergebnis dieser Arbeit belegt, dass je nach Ausgangslage eines KMU's ein eBackup Lösung Sicherheitsvorteile mit sich bringt, jedoch muss auch bei dieser Variante zwingend ein Datensicherungskonzept erstellt werden. Winkler und Meine erwähnten, dass auch bei einem BaaS ein zweites Backup vorhanden sein sollte. Ob sich ein KMU diese Sicherheit leisten wird, liegt in der eigenen Verantwortung. Die Erweiterung eines lokalen Backups mit einem eBackup würde die Sicherheit zusätzlich erhöhen, jedoch auch zusätzliche Kosten mit sich bringen. Für gewisse Firmen bietet ein BaaS eine sichere Lösung, da es ausserhalb der lokalen Infrastruktur gelagert wird. Die Kosten sind bei einem BaaS Service dadurch höher, zusätzlich gewinnt man aber auch bei der Datensicherheit, sofern das Datacenter einen entsprechenden Tier-Level besitzt. Die Anforderungen sind immer kundenspezifisch und können nicht als allgemeingültig betrachtet werden.

Wichtig sind bei einem Backup as Service auch die RPO und RTO Anforderungen. Laut Nelson kann ein RTO jedoch durch eine lokale Appliance beschleunigt werden. Zusätzlich sinken die Internet Anbindungspreise von Jahr zu Jahr, deshalb ist die kritische Betrachtung Nelsons in Bezug auf die Internetleitung nur teils gerechtfertigt. Mit Hilfe der Deduplizierung-Technologie werden auch nicht mehr hohe Bandbreiten benötigt. Dies ist auch aus der nicht

repräsentativen Umfrage ersichtlich. Bezüglich den Angeboten aus der Schweiz variieren die Preise sehr stark, in naher Zukunft wird sich zeigen, welche Backup as a Service Anbieter sich durchsetzen werden (vgl. Nelson 2011, Winkler & Meine, 2011).

Bechtle Dübendorf wird Backup as a Service im 2014 im Portfolio aufnehmen, mit Hilfe dieser Arbeit stehen wichtige Grundlagen für die Konzeption des Services zur Verfügung. Dabei ist auch die Erkenntnis sehr wertvoll, dass bei einer eBackup Lösung auch der Kunde wichtige Themen erledigen muss, wie z.B. Datensicherungs- und Sicherheitskonzepte.

5 Quellenverzeichnis

[BITKOM, 2013]
Bundesverband Informationswirtschaft, Telekommunikation und neue Medien BITKOM (2013). Betriebssicheres Rechenzentrum. Zugriff am 26.12.2013. Verfügbar unter http://www.bitkom.org/files/documents/131213_Leitfaden_BRZ_web.pdf

[Borges & Schwenk, 2012]
Borges, G. & Schwenk, J. (2012). Daten- und Identitätsschutz in Cloud Computing, E-Government und E-Commerce. Berlin: Springer.

[Brooks, McFarlance, Pott, Trcka & Tomaz, 2003]
Brooks, Ch., McFarlane, P., Pott, N., Trcka, M., & Tomaz, E. (2006). IBM Tivoli Storage Management Concepts. (5. Aufl.). San Jose, California: IBM Corporation, International Technical Support Organization

[BSI 2004]
Bundesamt für Sicherheit in der Informationstechnik BSI. Datensicherungskonzept. Zugriff am 25.11.2013. Verfügbar unter http://www.reinhard-wolf.de/upload/pdf/003_Datensicherungskonzept.pdf

[BSI, 2013]
Bundesamt für Sicherheit in der Informationstechnik BSI. (2013). M 6.33 Entwicklung eines Datensicherungskonzepts. Zugriff am 18.11.2013. Verfügbar unter https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/m06/m06033.html

[CSC, 2011]
CSC (2011). Big Data Infographic. Zugriff am 07.12.2013. Verfügbar unter http://www.csc.com/insights/flxwd/78931-big_data_universe_beginning_to_explode

[Davis & Patterson, 2012]
Davis, K. & Patterson, D. (2012). Ethics of Big Data: Balancing Risk and Innovation. Sebastopol: O'Reilly Media

[Drakos & Paquet, 2009]
Drakos, N. & Paquet R., (2009). Technology Trends You Can't Afford to Ignore. Zugriff am 01.12.2013. Verfügbar unter http://www.gartner.com/it/content/1258400/1258425/january_6_techtrends_rpaquet.pdf

[DSG, 2011]
Bundesgesetz über den Datenschutz DSG. Zugriff am 13.11.2013. Verfügbar unter http://www.admin.ch/opc/de/classified-compilation/19920153/201401010000/235.1.pdf

[EDÖB, 2011]
Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter EDÖB (2011). Erläuterungen zu Cloud Computing. Zugriff am 15.11.2013. Verfügbar unter http://www.edoeb.admin.ch/datenschutz/00683/00877/index.html

[EDÖB, 2012]
Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter EDÖB (2012). Urteil des Bundesverwaltungsgerichts vom 10. April 2012. Zugriff am 18.11.2013. Verfügbar unter http://www.edoeb.admin.ch/datenschutz/00628/00664/index.html

[Finn, Lownds, Luescher & Flynn, 2013]
Finn, A., Lownds, P., Luescher, M., & Flynn, D. (2013). Windows Server 2012 Hyper-V Installation and Configu-ration Guide. Indianapolis: John Wiley & Sons

[Friedl, 1998]
Friedl, W. (1998). Rechenzentrums-Sicherheit: sicherheitstechnische Beurteilung, Massnahmen gegen Gefährdungen. Berlin: Springer

[Furht & Escalante, 2010]
Furht, B. & Escalante A. (2010). Handbook of cloud computing. New York: Springer

[Gadatsch & Mayer, 2010]
Gadatsch, A. & Mayer, E. (2010). Masterkurs IT-Controlling : Grundlagen und Praxis für IT-Controller und CIOs - Balanced Scorecard - Portfoliomanagement - Wertbeitrag der IT - Projektcontrolling - Kennzahlen - IT-Sourcing - IT-Kosten- und Leistungsrechnung (4. Aufl.). Wiesbaden: Vieweg.

[Gantz & Reinsel, 2012]
Gantz, J. & Reinsel D., (2012). THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digi tal Shadows, and Biggest Growth in the Far East. Zugriff am 01.12.2013. Verfügbar unter http://idcdocserv.com/1414

[Hauri, Mohler, Deiniger, 2012]
Hauri, D., Mohler L., Deiniger S. (1998). Datentresor Schweiz. Basel: IWSB - Institut für Wirtschaftsstudien Basel AG

[Hoppe & Priess, 2003]
Hoppe, G. & Priess, A. (2003). Sicherheit von Informationssystemen: Gefahren, Massnahmen und Management im IT-Bereich. Herne: NWB, Neue Wirtschafts-Briefe.

[Hurwitz, Nugent, Halper & Kaufman 2013]
Hurwitz, J., Nugent, A., Halper, F. & Kaufman M. (2013). Big data for dummies. Hoboken: Wiley

[Tages-Anzeiger, 2012]
Haupt, K., (2012). Wenn der Virenschutz zu wenig hilft. Tages-Anzeiger, Zugriff am 01.12.2013. Verfügbar unter http://www.tagesanzeiger.ch/digital/internet/Wenn-der-Virenschutz-zu-wenig-hilft/story/25608988

[Tages-Anzeiger, 2013a]
Barandun, A., (2013). Wirtschaft: Geheimnummern und heikle Daten. Tages-Anzeiger, 21. Dezember 2013, Seite 43.

[Tages-Anzeiger, 2013b]
Knellwolf, T., (2013). Ich würde nie über einen US-Service mailen. Tages-Anzeiger, Zugriff am 13.12.2013. Verfügbar unter http://www.tagesanzeiger.ch/schweiz/standard/Ich-wuerde-nie-ueber-einen-USService-mailen/story/23317779

[Krutz & Vines, 2010]
Krutz, R. & Vinces, R. (2010). Cloud security : a comprehensive guide to secure cloud computing. Hoboken: Wiley.

[LTO, 2013]
Linear Tape-Open Technology LTO. Zugriff am 16.10.2013. Verfügbar unter http://www.lto.org/technology/generations.html

[Mahmood & Hill, 2011]
Mahmood, Z. & Hill, R. (2011). Cloud computing for enterprise architectures. London: Springer.

[Mahmood, 2013]
Mahmood, Z. (2013). Cloud Computing for Enterprise Architectures (Computer Communications and Networks). London: Springer.

[Melani, 2013]
Bundesbehörden der Schweizerischen Eidgenossenschaft. Melde- und Analysestelle Informationssicherung MELANI. Zugriff am 20.11.2013. Verfügbar unter http://www.melani.admin.ch/themen/00166/00171/

[Müller, 2008]
Müller, K. (2008). IT-Sicherheit mit System: Sicherheitspyramide - Sicherheits-, Kontinuitäts- und Risikomanagement - Normen und Practices - SOA und Softwareentwicklung. (3. Aufl.). Wiesebaden: Vieweg

[Nelson, 2011]
Nelson, S. (2011). Pro Data Backup and Recovery. Berkeley: Apress

[Neue Zürcher Zeitung, 2012]
Steier, H., (2012). Anti-Terror-Gesetz als Geschäftsrisiko. Neue Zürcher Zeitung, Zugriff am 11.11.2013. Verfügbar unter http://www.nzz.ch/aktuell/digital/patriot-act-cloud-computing-1.17410220

[Ohlhorst, 2013]
Ohlhorst, F. (2013). Big data analytics: Turning big data into big money. Hoboken: Wiley

[Winkler & Meine, 2011]
Winkler, V. & Meine, B. (2011). Securing the cloud: cloud computer security techniques and tactics. Amsterdam: Elsevier Syngress.

[Osuna, Balogh, Galante de Carvalho, Javier & Mann, 2011]
Osuna, A., Balgoh, E., Galante de Carvalho, A., Javier, R., & Mann, Z. (2011). Implementing IBM Storage Data Deduplication Solutions. San Jose, California: IBM Corporation, International Technical Support Organization

[Osuna, Cecchetti, Franz & Mencarelli, 2010]
Osuna, A., Cecchetti, L., Franz, E., & Mencarelli, R. (2010). TS7680 Deduplication ProtecTIER Gateway for System z. San Jose: IBM Corporation, International Technical Support Organization

[Patak 2010]
Patak, S., (2010). Sicherheit MOUNT10 COMBO ECO PRO aus rechtlicher Sicht . Zugriff am 27.10.2013. Verfügbar unter https://www.mount10.ch/pdf/Externes-Gutachten-MOUNT10.pdf

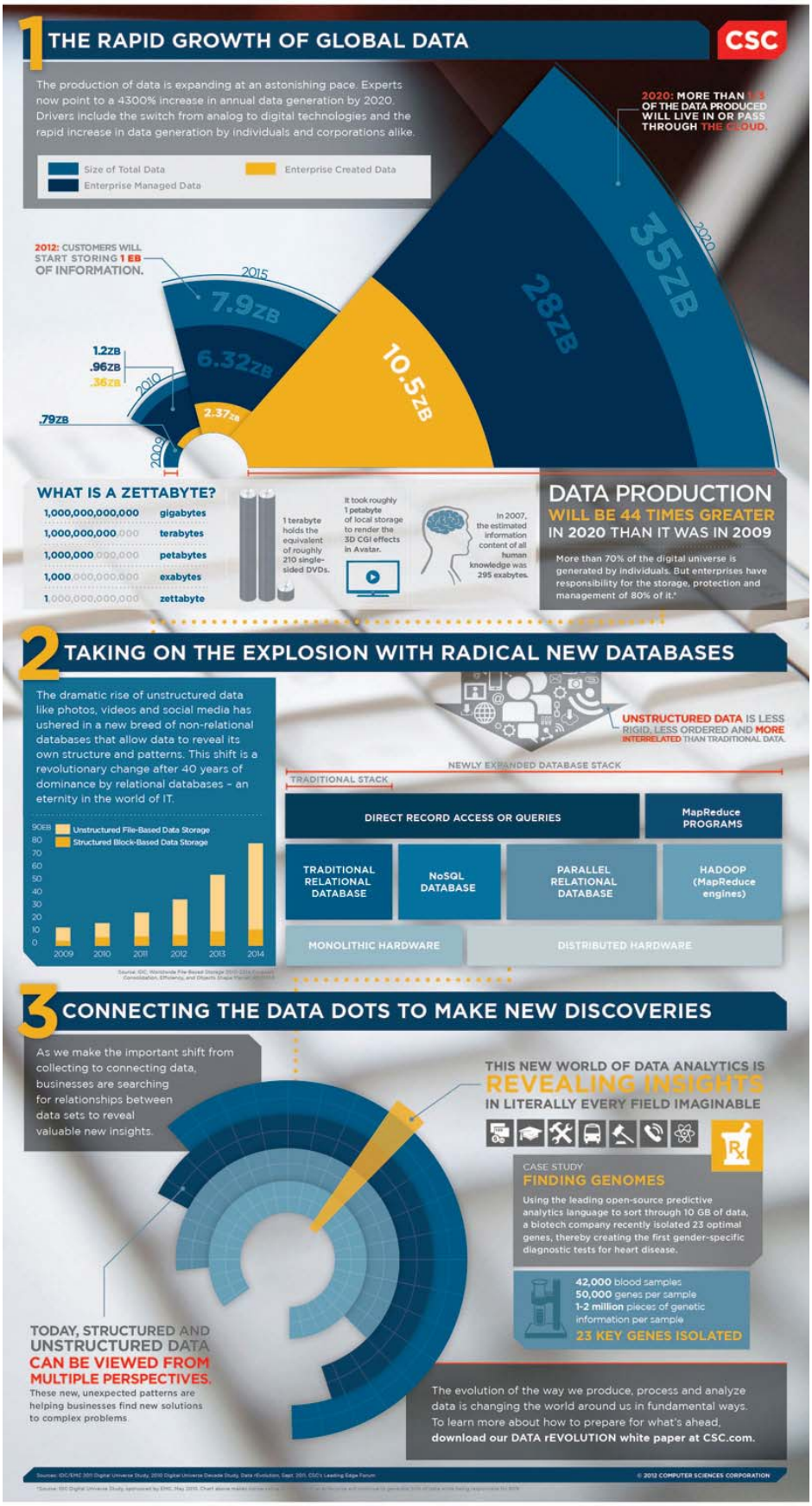
[Reese, 2009]
Reese, G. (2009 Cloud application architectures : [building applications and infrastructure in thecloud. Sebastopol: O'Reilly

[Wald, 2002]
Wald, E. (2002). Backup & Disaster Recovery. Bonn: MITP-Verlag

[Wieder, Butler, Theilmann & Yahyapour, 2011]
Wieder, Ph., Butler, J., Theilmann, & W., Yahyapour, R. (2011). Service Level Agreements for Cloud Computing. New York: Springer.

6 Anhang

Die Studie vom IDC 2011 Big Data Infographic.



Folgende Firmen haben keine Antwort auf Meine Umfrage gegeben:

Swissonlinegroup, Swisscom, Data-Safe, Uscp, Smartbee, Hosttech Swissbackupbank, Aa-fe-backup, ebckup.me, Mit-Group, MTF, Aconel, Viag, Vocielan, Net-Point und Triasys

Vorlage für die Backup to Cloud Umfrage:

Guten Tag

Ich möchte sie um einen unverbindliche Offerte für Ihr Backup to Cloud Angebot von 1 Terrabyte anfrage?

Ich schreibe derzeit eine Semesterarbeit bei der Kalidos FH (Bachelor für Wirtschaftsinformatik).
Meine Hypothese lautet wie folgt:
Gegenüberstellung von einem eBackup versus einem konventionellen Backup in Bezug auf das Datenwachstum von einem 1 Terrabyte bei einer Kostenreduktion von 50%

Nach dem ich die Theorie aus der Literatur verglichen habe, möchte ich marktgerechte Zahlen als vergleich nehmen. Des Weiteren ist eine Kollege (CIO einer Vermögensverwaltung in Zürich) an diesem Service sehr interessiert. Also richtiger Kunde steckt hinter der ganzen Anfrage.

Meine Fragen noch zu Ihrer Lösung:
- Lassen sich Physikalische und Virtuelle Server Sichern (Hypervisor VMware ev. in Zukunft Hyper-V) ?
- Wird eine Deduplizierung verwendet?
- Benötigte Datenleitung?
- Ist auch ein Desaster Recovery mit ihrer Lösung in einem späteren Schritt möglich? (Phase 2)

Die Betriebssysteme vom Kunden sind:
- Betriebssystem Windows 2008R2 oder Windows2012
- Linux Redhat (optional falls möglich)

Derzeit wird Backup Exec von Symantec verwendet und auf ein LTO4 Bandlaufwerk gesichert.
1 mal in der Woche ein Fullbackup und 6 Tage ein Inkrementelles Backup.

Bei Fragen stehe ich Ihnen jederzeit zu Verfügung

Freundliche Grüsse

Sandro Eggenberger

Lokale Kosten des Kunden A:

Backup Lösung Kunde A. (Tape Laufwerk) (15 Mitarbeiter und 500 Gigabyte Daten und 8 Virtuelle Server)		
Kostentreiber	Monatlich	Einmalig
Investition in Hardware		CHF 3’290.-
Investition in Software		CHF 1’890.-
Aufwand für Installation und Inbetriebnahme		CHF 1’600.-
Wartungskosten für Hard- und Software für 3 Jahre	CHF 40.-	
Energiekosten (inkl. Klimatisierung)	CHF 180.-	
Bänder (Initial & Ersatz von 2 Bänder/Monat)	CHF 35.-	
Betrieb (Bandwechsel und Auslagerung, Restores Prüfung)	CHF 210.-	
Total:	CHF 465.-	CHF 6’780

Die Backupkosten auf drei Jahre gerechnet ergibt folgende Kalkulation. Die Investitionskosten und die Monatlichen Betriebskosten werden addiert und mit durch 36 Monate dividiert ergibt, dies auf drei Jahre folgende Kosten:

36 x 465.- + 6'780 = 16’740 / 36 = **CHF 465.00 pro Monat**

Lokale Kosten des Kunden B:

Backup Lösung Kunde B. (Tape Laufwerk) (30 Mitarbeiter und 2 Terrabyte Daten und 15 Virtuelle Server)		
Kostentreiber	Monatlich	Einmalig
Investition in Hardware		CHF 5’290.-
Investition in Software		CHF 8’290.-
Aufwand für Installation und Invetriebnahme		CHF 2’400.-
Wartungskosten für Hard- und Software	CHF 50.-	
Energiekosten (inkl. Klimatisierung)	CHF 210.-	
Bänder (Initial & Ersatz von 2 Bänder/Monat)	CHF 55.-	
Betrieb (Bandwechsel und Auslagerung, Restores Prüfung)	CHF 360.-	
Total:	CHF 657.-	CHF 15’980

Die Backupkosten auf drei Jahre gerechnet ergibt folgende Kalkulation. Die Investitionskosten und die Monatlichen Betriebskosten werden addiert und mit durch 36 Monate dividiert ergibt, dies auf drei Jahre folgende Kosten:

36 x 657.- + 15'980 = 39’632 / 36 = **CHF 1’100.90 pro Monat**

Quelle des BSI, Zugriff am 27.10.2013. Verfügbar unter: https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/m06/m06033.html

M 6.33 Entwicklung eines Datensicherungskonzepts

Verantwortlich für Initiierung: IT-Sicherheitsbeauftragter

Verantwortlich für Umsetzung: IT-Sicherheitsbeauftragter, Leiter IT, Verantwortliche der einzelnen Anwendungen

Die Verfahrensweise der Datensicherung wird von einer grossen Zahl von Einflussfaktoren bestimmt. Das IT-System, das Datenvolumen, die Änderungsfrequenz der Daten und die Verfügbarkeitsanforderungen sind einige dieser Faktoren. Im Datensicherungskonzept gilt es, eine Lösung zu finden, die diese Faktoren berücksichtigt und gleichzeitig unter Kostengesichtspunkten wirtschaftlich vertretbar ist.

Die technischen Möglichkeiten, Datensicherungen durchzuführen, sind vielfältig. Jedoch wird die Auswahl immer von den genannten Faktoren bestimmt. Daher gilt es zunächst, die Einflussgrössen der IT-Systeme und der damit realisierten IT-Anwendungen zu bestimmen und nachvollziehbar zu dokumentieren. Anschliessend muss die geeignete Verfahrensweise entwickelt und dokumentiert werden. Zum Abschluss muss durch die Behörden-/Unternehmensleitung die Durchführung angeordnet werden.

Das Datensicherungskonzept muss für die Gewährleistung einer funktionierenden Datensicherung die Datenrestaurierbarkeit mittels praktischer Übungen als Verpflichtung vorsehen.

Die Ergebnisse sollten aktualisierbar und erweiterbar in einem Datensicherungskonzept niedergelegt werden. Ein möglicher Aufbau eines Datensicherungskonzepts ist im nachfolgenden Inhaltsverzeichnis beispielhaft aufgezeigt:

Inhaltsverzeichnis Datensicherungskonzept

1. Definitionen

- Anwendungsdaten, Systemdaten, Software, Protokolldaten
- Vollsicherung, inkrementelle Datensicherung

2. Gefährdungslage zur Motivation

- Abhängigkeit der Institution vom Datenbestand
- Typische Gefährdungen wie ungeschulte Benutzer, gemeinsam genutzte Datenbestände, Computer-Viren, Hacker, Stromausfall, Festplattenfehler
- Institutionsrelevante Schadensursachen
- Schadensfälle im eigenen Haus

3. Einflussfaktoren je IT-System

- Spezifikation der zu sichernden Daten
- Verfügbarkeitsanforderungen der IT-Anwendungen an die Daten
- Rekonstruktionsaufwand der Daten ohne Datensicherung
- Datenvolumen
- Änderungsvolumen
- Änderungszeitpunkte der Daten
- Fristen
- Vertraulichkeitsbedarf der Daten

- Integritätsbedarf der Daten
- Kenntnisse und datenverarbeitungsspezifische Fähigkeiten der IT-Benutzer

4. Datensicherungsplan je IT-System

4.1 Festlegungen je Datenart

- Art der Datensicherung
- Häufigkeit und Zeitpunkt der Datensicherung
- Anzahl der Generationen
- Datensicherungsmedium
- Verantwortlichkeit für die Datensicherung
- Aufbewahrungsort der Backup-Datenträger
- Anforderungen an das Datensicherungsarchiv
- Transportmodalitäten
- Rekonstruktionszeiten bei vorhandener Datensicherung

4.2 Festlegung der Vorgehensweise bei der Datenrestaurierung

- Randbedingungen für das Datensicherungsarchiv
- Vertragsgestaltung (bei externen Archiven)
- Refresh-Zyklen der Datensicherung
- Bestandsverzeichnis
- Löschen von Datensicherungen
- Vernichtung von unbrauchbaren Datenträgern
- Vorhalten von arbeitsfähigen Lesegeräten

5. Minimaldatensicherungskonzept

6. Verpflichtung der Mitarbeiter zur Datensicherung

Prüffragen:

Existiert ein aktuelles Datensicherungskonzept?

Sind sämtliche betroffenen IT-Systeme im Datensicherungskonzept aufgeführt?

Sind die Mitarbeiter über den sie betreffenden Teil des Datensicherungskonzepts unterrichtet?

Wird die Umsetzung des Datensicherungskonzepts regelmässig kontrolliert?

Stand: 13. EL Stand 2013