SEP sesam

Backup in verteilten Umgebungen

Eine Einführung für Partner und Kunden, Backup-Architekten, -Designer und CIOs

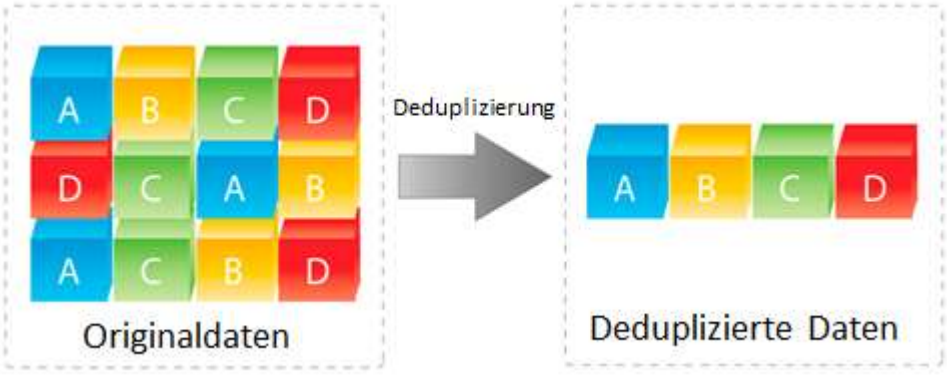
Inhalt

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Einleitung

Aufgrund der heutigen Möglichkeiten der Vernetzung und steigenden Bandbreite sind Firmenumgebungen fast zu 100% verteilte Umgebungen. D.h. ein meist leistungsfähiges Firmennetzwerk ist mit Netzwerkinseln in Außenstellen oder zu seinen Außendienstmitarbeitern nur über öffentliche Leitungen verbunden. Insbesondere die Globalisierung und Verteilung der Firmen weltweit vergrößern das Problem der Interkonnektivität immer wieder. Hierbei ist die Anbindung einer Netzwerkinsel zwar einerseits ein Kostenproblem, denn das Mieten großer Bandbreiten und Übertragen großer Datenmengen ist noch immer ein teures Unterfangen, andererseits aber auch oft technisch noch nicht oder nur mit sehr großem Aufwand möglich, je nachdem wie entlegen die Außenstelle ist. Daneben spielt für die Datenübertragung nicht nur die reine Datenmenge eine Rolle, sondern auch die Sicherheit und Stabilität. Je nach Wert und Klassifizierung der zu übertragenden Daten, die das sichere Firmennetz verlassen, muss dabei auch an Verschlüsselung und Integrität gedacht werden.

Um die zu übertragende Datenmenge über WAN zu reduzieren, war lange Zeit die Komprimierung die einzige Möglichkeit. Beginnend mit Single Instancing (doppelte Files z.B. Attachments an Mails werden nur einmal übertragen), hat sich mittlerweile die Deduplizierung am Markt etabliert. Hierbei wird der Datenstrom und damit auch Files in Blöcke zerlegt, wobei nach einem Vergleich die bereits am Backup-Server vorhandenen Blöcke erst gar nicht mehr übertragen werden, sondern nur noch ein Verweis auf den bereits vorhandenen Block eingetragen wird.



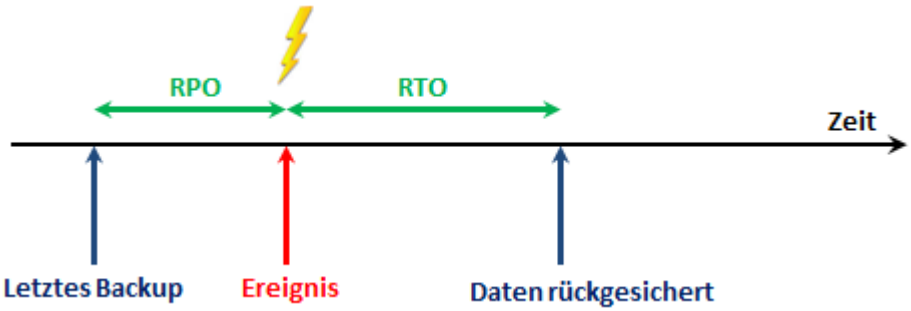
SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTHAMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Überblick

Zum Sichern von verteilten Umgebungen gibt es sehr unterschiedliche Ansätze je nach gegebener Infrastruktur, Datenmengen oder Zeitfenstern. Zwei wichtige Kriterien zum Festlegen der konkreten Anforderungen an Backup und Restore sind die beiden Werte RPO und RTO:

- **Recovery Time Objective (RTO)**
  Wie lange darf ein Geschäftsprozess/System ausfallen?
  Bei der Recovery Time Objective handelt es sich um die Zeit, die vom Zeitpunkt des Schadens bis zur vollständigen Wiederherstellung der Geschäftsprozesse (Wiederherstellung von: Infrastruktur - Daten - Nacharbeitung von Daten - Wiederaufnahme der Aktivitäten) vergehen darf. Der Zeitraum kann hier von 0 Minuten (Systeme müssen sofort verfügbar sein), bis mehreren Tage oder Wochen betragen.

- **Recovery Point Objective (RPO)**
  Wie viel Datenverlust kann in Kauf genommen werden?
  Bei der Recovery Point Objective handelt es sich um den Zeitraum, der zwischen zwei Datensicherungen liegen darf, das heißt, wie viele Daten/Transaktionen dürfen zwischen der letzten Sicherung und dem Systemausfall höchstens verloren gehen. Wenn kein Datenverlust hinnehmbar ist, beträgt die RPO 0 Sekunden.



Jede Konfiguration hat ihre Vor- und Nachteile. Hier ein Überblick über die in den nächsten Kapiteln näher erläuterten Möglichkeiten zum Sichern von verteilten Umgebungen. Dabei werden in den ersten 4 Kapiteln die bisher üblichen Möglichkeiten aufgezeigt, bevor in den letzten beiden Kapiteln dann auf die Alternativen mit Deduplizierung eingegangen wird.

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

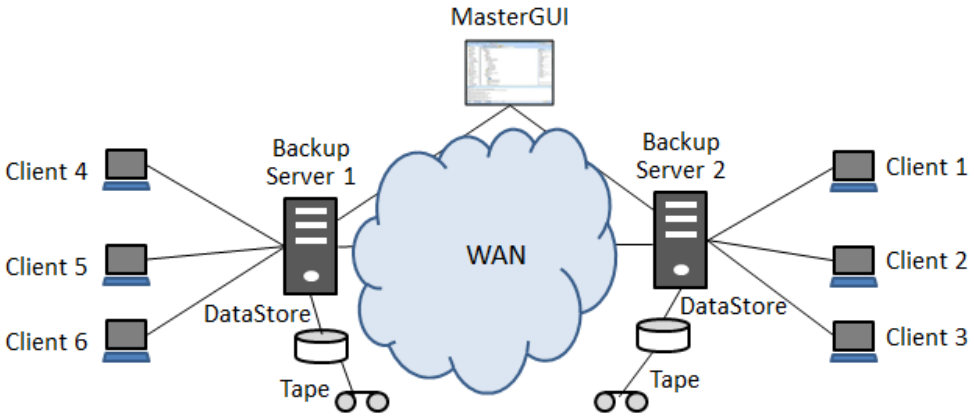
Methode	Bewertung
Separate Datenzonen	Flexibel, kostenintensiv, keine Synergieeffekte durch Zentralisierung, aber zentrale Kontrolle über MasterGUI möglich
Clients direkt über WAN sichern	Günstig, einfach, nur wenige Clients, geringe Datenmengen
Lokale Sicherung in der Außenstelle	Große Datenmengen, separate HW und lokale Datenhaltung bei zentraler Administration über AdminGUI
Replikation (rsync, GlusterFS, Ceph)	Mittlere Datenmengen, nicht von der Backup-SW gesteuert
Source Deduplication	Kleine Bandbreite möglich, initiales Backup und große Restores problematisch
Replikation mit Deduplizierung	Große Datenmengen möglich, volle Flexibilität

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTHAMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Separate Datenzonen

Dieser einfachste aller Fälle steht meist am Anfang. Entstanden entweder historisch mit der Zeit durch Wachstum, durch Übernahmen oder weil Außenstellen organisatorisch selbständig bleiben wollen. Hierbei können die komplett separierten Datenzonen vollkommen eigenständig und flexibel nach ihren Anforderungen agieren (z.B. unterschiedliche Lizenzmodelle, Backupstrategien, HW, etc.), im Extremfall sogar unterschiedliche Backup-Software-Produkte einsetzen.

Wenn in mehreren Datenzonen jeweils ein SEP sesam Backup Server steht, können alle Backup Server zentral über ein MasterGUI administriert werden. Logfiles und Backup Stati sind dadurch zentral einsehbar und Folgeaktionen wie eine Migration können angestoßen werden. Der Rechner mit dem MasterGUI muss kein Backup-Server sein, sondern nur ein beliebiger Rechner im Netzwerk mit einem installierten SEP sesam Client und Java.

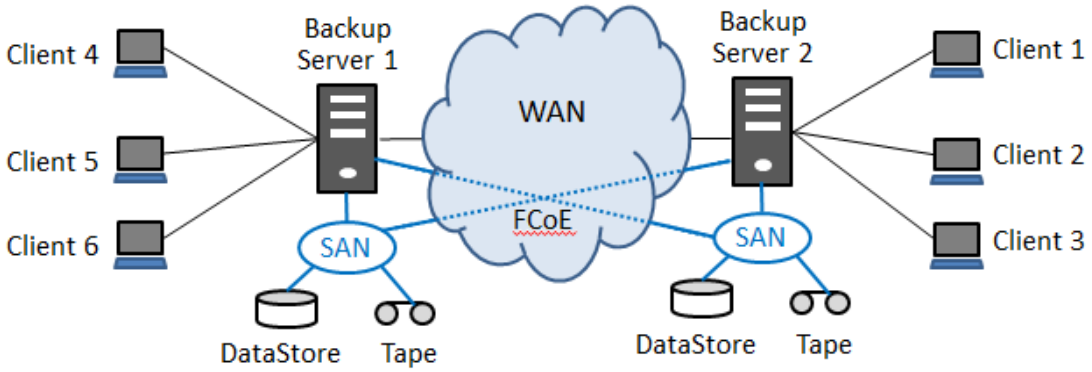


Vorteile	Nachteile
unabhängig und flexibel	hohe Kosten durch doppelte Verwaltung, Lizenzierung und Backup-Infrastruktur
keine Datenübertragungsprobleme	keine Synergieeffekte
lokale Datenhaltung, aber zentrale Kontrolle über MasterGUI über mehrere - auch viele - SEP Datenzonen	kein MasterGUI beim Einsatz von unterschiedlicher Backup-Software
Sinnvoll um Private Cloud Strukturen aufzubauen	

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Gekoppelte Datenzonen

Getrennte Datenzonen sind oft strategisch oder strukturtechnisch vorgegeben, dennoch besteht dann gerade im Enterprise-Bereich der Wunsch die Daten für Ausfallszenarien in die jeweils andere Datenzone auszulagern. Dies setzt natürlich ein leistungsfähiges WAN (am besten >1GBit/s) voraus. Eine Möglichkeit ist, die Backupziele d.h. Laufwerke der Gegenseite für Copy- oder Migrationsaufträge zu verwenden. Hierzu müssen die Remote-Laufwerke lokal konfiguriert werden. Konzeptionell relativ einfach zu planen, steckt hier die eigentliche Problematik im technischen Detail, insbesondere wenn mehrere technische Funktionalitäten kombiniert werden z.B. SAN/WAN-Kopplung/FCoE, Replikation, Deduplizierung etc. steigt mit der Komplexität automatisch das Risiko eines funktionierenden Disaster Recovery, was dem ursprünglichen Ziel des Ganzen entgegenspricht. Insbesondere der Aufwand, bis die jeweilige Konfiguration stabil in Betrieb ist, steigt exponentiell mit der Komplexität und muss unbedingt durch erhöhte zusätzliche Consultingleistung im Voraus mit eingeplant werden. Zum Beispiel hat die Erfahrung erst bei der Implementierung gezeigt, dass das Protokoll FCoE (Fibre Channel over Ethernet) sich je nach Switch-Einstellungen anders verhält als reines SAN über FC.



Das Konzept des Kopierens bzw. Migrierens der Daten auf ferne Laufwerke gestattet ein sicheres Auslagern der Daten. Bei Ausfall einer ganzen Datenzone muss diese aber erst wieder komplett aufgebaut werden, dann sind die Daten aus der entfernten Datenzone wieder zurück zu spielen. Ein Zugriff der anderen Datenzone auf diese Daten ist nicht möglich, womit diese auch nicht den Betrieb der ausgefallenen Datenzone übernehmen kann.

So ein Konzept ist mit SEP sesam erst bei Einsatz von Replikation (Si3R) möglich, wo die Daten durch die integrierte Importfunktion am entfernten Backup Server importiert werden

können und damit sofort wieder ein Backup Server für Backup und Restore zur Verfügung steht.

Je aufwändiger die Anforderungen, und je komplexer die Lösung, desto mehr verschiedene Lösungsansätze gibt es, die es zu vergleichen gilt in punkto Aufwand, Kosten und Nutzen. Z.B. tummeln sich in diesem Marktsegment auch viele Hardwarehersteller inkl. zugehörigem Softwarestack, die ihre Replikationsmöglichkeiten über große Entfernungen anpreisen. Oder es sind manuelle und zusätzliche Administrationsaufwände abzuwägen gegen Zusatzkosten für eine komplexere und damit auch oft risikoreichere Lösung. Letztendlich wird hier oftmals sogar die persönliche Vorliebe oder subjektive Meinung oder Erfahrung des Kunden als auch des technischen Beraters entscheidend sein.

Vorteile	Nachteile
beliebige Funktionalitäten in unterschiedlichen Lösungsansätzen implementierbar	steigende Komplexität und exponentiell steigender Consultinganteil
	schwer planbare Kosten

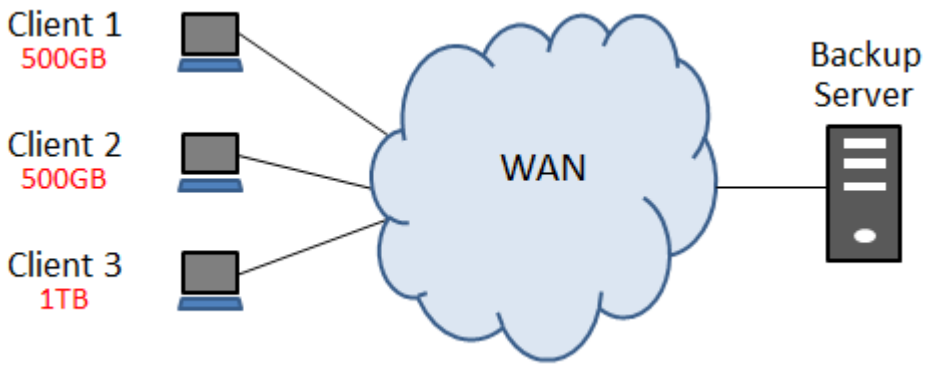
SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Clients direkt über WAN sichern

Aus Konfigurationssicht ist es kein Problem beliebige Clients im Netzwerk im SEP Backup Server zu konfigurieren, solange der Clientname im DNS auflösbar ist und sie über Netzwerk erreichbar sind. Völlig unabhängig davon wie nah oder fern (Firewall, anderes Netzsegment, WAN Strecke, etc.) dieser Client ist. Hierbei muss der Backup Server den Client auflösen und erreichen können als auch umgekehrt.

Das eigentliche Problem beginnt erst mit der Datenübertragung. Je geringer die Bandbreite der Netzwerkverbindung insbesondere bei WAN-Verbindungen, desto geringer die sinnvoll zu übertragende Datenmenge. Sinnvoll bedeutet hier das Einhalten von Backupzeitfenstern oder Restore SLAs. Aber auch das längere Belegen von Bandbreiten durch das Backup, so dass z.B. mit anderen Applikationen nicht mehr richtig zu arbeiten ist, kann die Sinnhaftigkeit in Frage stellen. Oftmals werden für das Backup eigene Standleitungen mit fester Bandbreite reserviert.

Werden nun einzelne Clients direkt remote über den Backup Server gesichert, funktioniert das Konzept aufgrund der Limitierungen in der Datenübertragung meist nur mit wenigen Clients und geringen Datenmengen.



Beispiel:

- 1 Außenstelle mit 2 Clients in Summe 1TB Daten
- Änderungsrate 10%
- Eine WAN-Leitung mit 155Mbit/s (ATM)

⇨ Dauer für eine Vollsicherung = 1.000.000MB * 8bit / 155Mbit/s = ~14 Stunden

⇨ Dauer für eine inkrementelle Sicherung = ~1,4 Stunden

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTHAMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Wie man der einfachen Rechnung entnehmen kann, macht dies wenn überhaupt dann nur bei sehr schnellen WAN-Verbindungen Sinn.

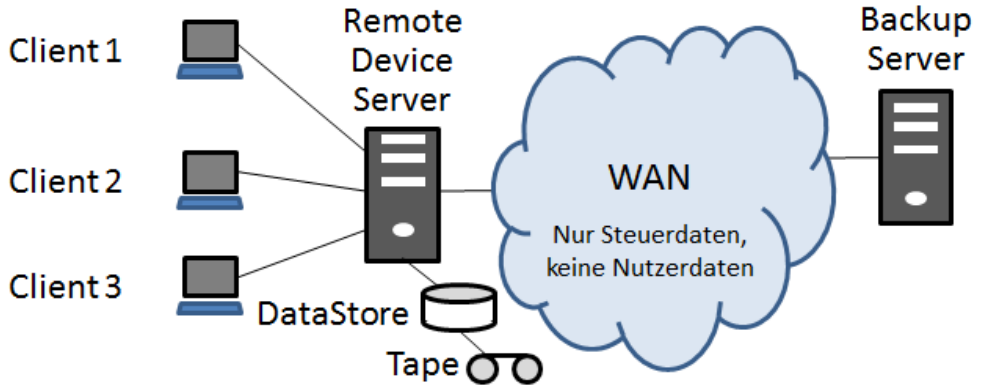
Vorteile	Nachteile
zentrales Backup	kleine Datenmengen
einfachste Konfiguration	wenige Clients
keine Extrakosten	schnelles WAN empfehlenswert
direkte Sicherung auf Tape möglich	

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Lokale Sicherung in der Außenstelle

Wenn die Datenmenge zu groß wird für eine Übertragung über WAN müssen die Backupdaten lokal gespeichert werden. Hierzu muss man in der Außenstelle einen Remote Device Server (RDS) aufsetzen, der stellvertretend für den Backup Server die Daten auf eigene Backupmedien (Disk und/oder Tape) schreibt. Bei Einsatz von einem oder mehreren RDS (die Konkurrenz nennt es Media Server oder Storage Node) bleibt die gesamte Kontrolle beim Backup Server, wohingegen der Datenfluss bei Backup, Restore oder Migration im RDS erfolgt.

Dieses Konzept erfordert zwar zusätzlich HW und Administration und widerspricht bzgl. der Datenlagerung einem zentralen Backupkonzept, hat aber erhebliche Performance-Vorteile insbesondere beim Restore, da die Backupdaten lokal verfügbar sind.



Vorteile	Nachteile
Große Performance durch lokale Datenhaltung	keine zentrale Datenhaltung
zentrale Steuerung des Backups	zusätzliche Kosten für HW und Administration
große Datenmengen, viele Clients	

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTHAMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Replikation

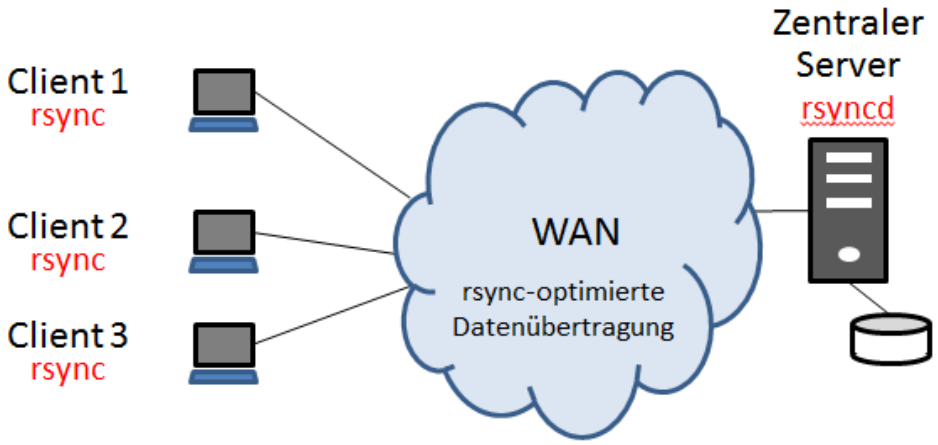
Schon immer bestand das Problem, Daten insbesondere aus Disaster Recovery Gründen an andere Standorte zu verlagern. Sollte mit einem Rechner oder einem ganzen Standort ein Katastrophenfall eintreten, sollen sämtliche Daten an einem anderen Ort verfügbar sein, so dass sehr einfach und schnell mit Ersatz-Hardware eine funktionierende Umgebung aufgesetzt werden kann. Dies kann entweder ein anderes Rechenzentrum sein oder eine Auslagerung der Daten an einen katastrophensicheren Ort. Früher hat man hierzu z.B. Bandkopien in Bunker oder Bergwerke transportiert. Mit Zunahme der Bandbreiten sind aber auch Datenreplikation oder –spiegelung disk-to-disk, synchron oder asynchron, in begrenztem Umfang übers Netzwerk möglich geworden.

Im einfachsten Fall kann z.B. unter UNIX/Linux mit einem in eine crontab eingetragenen rsync-Kommando Daten in regelmäßigem Rhythmus remote repliziert werden. Rsync ist ebenso über cygwin auch auf Windows verfügbar, es gibt aber auch eine Vielzahl von rsync-basierten GUI-Tools. Rsync ist insofern für die Übertragung von Dateien über WAN optimiert, dass es Dateien in Blöcke zerlegt, Checksums bildet und nur geänderte Blöcke überträgt. Ebenso ist verschlüsselte (ssh) und komprimierte Datenübertragung möglich. Da rsync aber vorwiegend dateienbasiert arbeitet, ist für die Replikation ganzer Verzeichnisse manuelles Skripting oder der Einsatz weiterer rsync-basierter Tools ratsam. Mit rsync könnten entweder die Originaldaten der Clients oder aber lokale Backupdaten (z.B. DataStore eines RDS) repliziert werden. Damit könnten auf diesem Weg sogar deduplizierte Daten repliziert werden.

Eine weitere Möglichkeit der Replikation von Daten ist auf Basis von verteilten Dateisystemen wie GlusterFS, Ceph oder DFS, die die Replikation als für den Anwender konfigurierbare integrierte Funktionalität bieten.

Bei allen dieser Alternativen ist in jedem Fall das Thema Skalierbarkeit genauer zu betrachten und im konkreten Einsatzfall zu testen.

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com



Vorteile	Nachteile
zentrale Datenkopie für Desaster	große manuelle Aufwände
sehr kostengünstig	wenig Support
mittlere Datenmengen, mehrere Clients	Erhöhter Speicherplatzbedarf
Lesbare da Original- Datenformate	

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTHAMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Source Deduplication

Um das einfache Konfigurationskonzept der direkten Sicherung der Clients am zentralen Backup-Server beizubehalten, dennoch aber auch größere Datenmengen und eine größere Anzahl von Clients sichern zu können, empfiehlt sich Source Deduplication. Im Gegensatz zur Target Deduplication bedeutet dies, dass die Datenquelle d.h. der Client die zu sichernden Daten in Blöcke zerlegt, pro Block einen Hashwert bildet, diesen mit dem zentralen Index am Backup Server vergleicht und dann nur die nicht bereits am Backup Server vorhandenen Blöcke überträgt. Damit kann die zu übertragende Datenmenge erheblich reduziert werden, da nicht nur geänderte Dateien, sogar nur geänderte und nicht vorhandene Blöcke übertragen werden. Zwar erhöht diese Technologie etwas die Rechenlast auf dem Client, aber je nach aktueller Deduplizierungsrate kann die Datenmenge bis zu einem Faktor 10 oder mehr reduziert werden. Und damit auch die Dauer eines Backups. Zudem reduziert wird die Datenmenge durch Komprimierung der zu übertragenden Blöcke.

Aufgrund des Restoreverfahrens, wobei zusammengehörige Blöcke beliebig im Backup Medium verteilt sind, ist Deduplizierung eine reine diskbasierte Technologie. Desweiteren entspricht eine deduplizierte Sicherung logisch immer einer Vollsicherung, so dass es keine separaten Vollsicherungen mehr gibt. Es werden dennoch mit jeder Sicherung nur wenige Daten übertragen.

Eine Deduprate wird meist angeben in Relationen z.B. 10:1, was einer Reduktion der Daten um 90% entspricht. Eine gute Deduprate sollte mindestens 3:1 sein, damit sich die Reduktion signifikant von einer einfachen Komprimierung unterscheidet. Außerdem sollten die zusätzlichen Kosten der Lizenzierung für die Deduplizierung einhergehen mit einer signifikanten Kostenersparnis beim Speichermedium. Zur Bestimmung der ungefähr zu erwartenden Deduprate, ist eine Testinstallation im individuellen Datenmix empfehlenswert.

Der sinnvolle Einsatz dieser Technologie - d.h. eine gute Deduprate - ist von mehreren Faktoren abhängig:

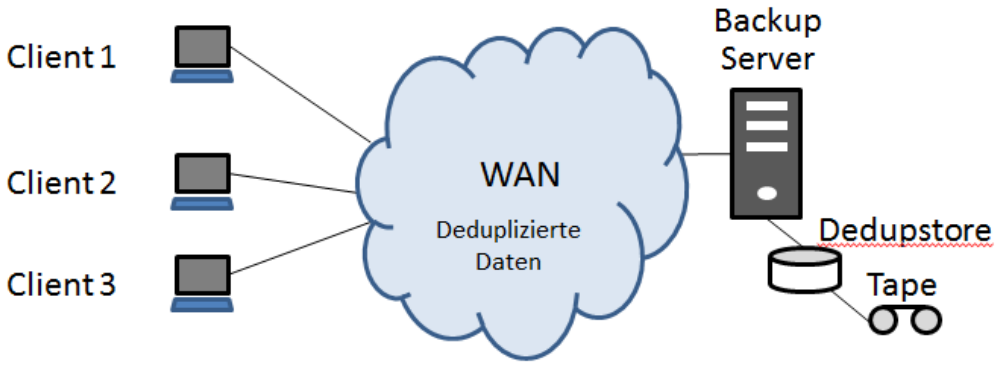
- Änderungsrate (sollte <5% sein)
- Aufbewahrungszeit (längere Aufbewahrung erhöht die Deduprate, aber auch die Datenmenge am Sicherungsziel)
- Globale Deduplizierung (je mehr Daten von Backups in einen zentralen Index einfließen, desto besser wird die Deduprate)
- Datenformate (schlecht sind z.B. Bild- oder Videodaten)

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

- Verschlüsselung (die Daten sollten erst nach erfolgter Deduplizierung komprimiert oder verschlüsselt werden)
- Der Deduplizierungsalgorithmus (SEP sesam verwendet einen eigenentwickelten und 4-fach patentierten, hocheffizienten Algorithmus mit variabler Blocklänge)

Aspekte, die unbedingt bei der Source Deduplication noch zu beachten sind:

1. **Die initiale Sicherung**
   Der Nachteil einer WAN-Verbindung macht sich insbesondere beim ersten Backup bemerkbar, wo nahezu die volle Datenmenge einmal übertragen werden muss. Dies ist natürlich abhängig davon, was bereits im Dedupstore vorhanden ist. So ist z.B. die zu übertragende Datenmenge einer zweiten Windows VM schon erheblich geringer. Für die initiale Sicherung sollte speziell Zeit eingeplant werden (z.B. Wochenende).
2. **Der Restore**
   Das Rücksichern einzelner Files oder kleinere Datenmengen ist aufgrund der WAN-Verbindung noch möglich, muss aber mehr rückgesichert werden, sind gesonderte Maßnahmen erforderlich. Insbesondere der Disaster Recovery Fall, wo komplette Server wieder aufgesetzt werden müssen, muss bedacht werden. Eine in der Praxis oft angewendete Methode ist, einen Server in der Zentrale zu betanken und dann per Express physikalisch in die Außenstelle zu liefern. Hierfür muss man natürlich die Erreichbarkeit der Außenstelle betrachten.
3. **Backup auf Tape**
   Sollte man die Anforderung eines Backups auf Tape haben (z.B. Archivierung aus Compliance-Gründen), dann muss dies in einem nachgelagerten Migrationsschritt erfolgen. Hierbei sollten sinnvollerweise die dedupizierten Daten vor Ablage auf den Bändern wieder „rehydriert“ d.h. die Deduplizierung rückgängig gemacht werden.



Beispiel:

- 1 Außenstelle mit 2 Clients in Summe 1TB Daten
- Änderungsrate für Blöcke 2%
- Eine WAN-Leitung mit 155Mbit/s (ATM)
- Deduprate von 10:1

⇨ Dauer für eine Sicherung =
1.000.000MB * 8bits *0,02 *0,1 / 155Mbits/s = ~1,7 Minuten

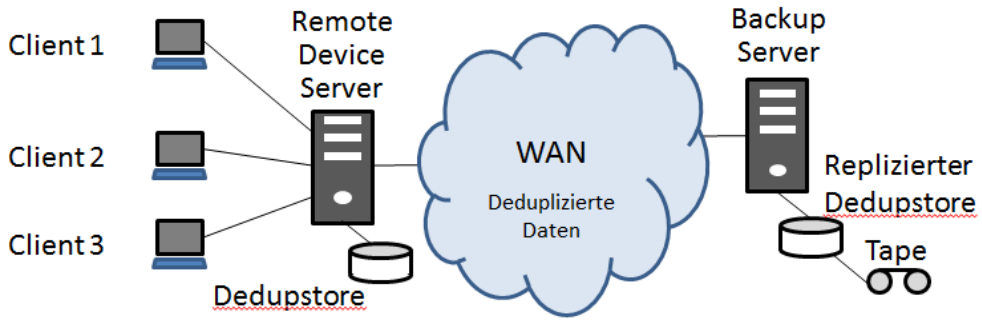
Vorteile	Nachteile
zentrales Backup	initiales Backup, Datenmenge wie bisher
einfache Konfiguration	große Restores, Disaster Recovery
permanente logische Vollsicherung	muss auf Disk gehen, nachgelagerte Migration auf Band
mittlere Datenmengen, mehrere Clients	globale Deduplizierung empfehlenswert
	mehr Rechenleistung auf den Clients erforderlich
	erhöhtes Verlustrisiko der Daten bei Inkonsistenzen der Datenspeicherung

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Replikation mit Deduplizierung

Um eine Replikation effizienter zu gestalten, ist hierfür Deduplizierung das ideale Mittel zum Zweck. Die zu übertragende Datenmenge wird soweit reduziert, dass bei heute üblicher WAN-Technologie (ATM) sogar eine synchrone Replikation auch nahezu beliebig großer Datenmengen möglich ist. War diese Technologie in der Vergangenheit vorwiegend eine Domäne der Storage-Hardwarehersteller, so hat es heute zunehmend Einzug in die Software gehalten. Replikation durch die Software hat den Vorteil, dass die Backup-Software die Kontrolle über alle Backups behält und im Disaster Fall sofort Restores aus dem Replikat erledigen kann. Hardware-Hersteller stellen die für die Backup-Software meist transparente Replikation als Vorteil dar, werden aber mit der Unterstützung von Schnittstellen wie Symantec OST (Open Storage Technology) oder EMC DDBoost für NetWorker mit Data Domain der Anforderung zur Kooperation mit der Software gerecht. Allerdings sind diese Schnittstellen immer proprietär, so dass man damit an einen bestimmten Software-Hersteller gebunden ist.

Replikation mit SEP sesam erfolgt über einen Remote Device Server, der zum Beispiel in einer Außenstelle einen DedupStore für die Backups der dortigen Clients konfiguriert hat. Die Replikation wird zentral am Backup Server konfiguriert und angestoßen. Hierbei können beliebige auch günstige Disk Arrays verwendet werden.



Je nach Infrastruktur gibt es verschiedene Replikationsarten so z.B. n:1, d.h. viele DedupStores werden in eine Zentrale repliziert oder 1:m, d.h. ein DedupStore wird an mehrere Standorte repliziert. Aber auch beliebig komplexe Mischkonzepte n:m sind möglich. Alle Arten werden Schritt für Schritt in die SEP sesam Si3R Replikation eingeführt. Stand heute erfolgt im SEP sesam die Replikation nur asynchron und ist max. n:1 im GUI konfigurierbar.

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTHAMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Vorteile	Nachteile
eine oder mehrere Datenkopien für Desaster Fälle	Kosten durch eigene Lizenz im Classic Lizenzmodell
nahezu beliebige Datenmengen	Kosten für zusätzlichen Speicherplatz
Kostengünstige Disk Arrays ausreichend	Leistungsfähigere HW für den RDS
Unterschiedliche Replikationsarten	

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Sichern der Laptops von Außendienstmitarbeitern

Eine besondere Art von Außenstellen sind Mitarbeiter im Außendienst. Durch den mobilen Einsatz von Laptops ist eine zeitplangesteuerte Sicherung schwierig. Auch in Standorten, wo Mitarbeiter ihre Laptops am Abend immer mit nach Hause nehmen, ist kein Zeitfenster für eine Vollsicherung.

Die erste Maßnahme ist ein separater Zeitplan für Laptops. Hierbei sollte die Sicherung nicht über Nacht erfolgen, sondern über Mittag. 2 Maßnahmen können hier bei der Planung helfen:

- **Wake on LAN (WoL)**

  Mit dieser im BIOS einzuschaltende und im SEP sesam GUI zu konfigurierende Funktionalität kann der Backup Server einen Client bei angeschlossenem LAN-Kabel durch Ansprechen über die MAC-Adresse fürs Backup eigens hochfahren. Nach erfolgtem Backup fährt der Rechner im Normalfall über die definierten Systemeinstellungen wieder in den Ruhezustand.

- **Skript zum Test der Verbindung**

  Sinnvoll kann sein ein PRE-Skript zu konfigurieren, dass vor der Sicherung das Vorhandensein eines Rechner durch einen Verbindungstest feststellt. Über einen Zähler kann man den Administrator oder Rechnerbesitzer über Email warnen, dass eine bestimmte Zeit keine Sicherung mehr erfolgte und Abhilfe geschaffen werde sollte.

Desweiteren sollte es für Laptops unbedingt die Möglichkeit geben, selbst eine Sofortsicherung zu starten, sobald man eine entsprechende Verbindung und genügend Zeit hat. Das dafür benötigte Konzept zur Rollen- und Rechteverteilung ist zur Zeit für SEP sesam in der Entwicklung. Die Funktionalität der Übernahme von Userprofilen aus Active Directory oder via LDAP wird über das WebUI in einem zweiten Schritt eingeführt. In SEP sesam kann man sich heute damit behelfen, dass Außendienstmitarbeiter das AdminGUI installieren und in der Userrolle darüber eine Sofortsicherung durchführen.

Besonders effektiv beim Sichern von Laptops mit Vollsicherungen über Tage ist der Einsatz von Source Deduplication, was die Sicherungszeit einer Vollsicherung erheblich minimiert und auch für Außendienstmitarbeiter erträglich macht.

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTHAMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Netzwerkstrukturen

Wenn man das Backup über verteilte Umgebungen betrachtet, darf man auch die Netzwerkstruktur nicht außer Acht lassen. Hierbei gibt es eine Vielzahl von Unterschieden und Fällen, mit denen der SEP sesam jeweils umgehen können muss.

Die SEP sesam Ports

Die SEP sesam Kommunikation arbeitet immer nach dem Client-Server-Prinzip, d.h. es verbindet sich immer ein Client zu einem Daemon. Für die verschiedenen SEP sesam Daemon sind Standard TCP Ports festgelegt. Die Daemon laufen je nach installierten Modulen auf dem SEP sesam Server, dem SEP sesam RDS oder dem Clienten.

Dienst	Port	Beschreibung
sm_ctrld	11301	unverschlüsselte Steuerkommunikation (**Ctrl**)
sm_sshd	11322	Steuerkommunikation über **SSH** Tunnel
sm_stpd	11001	Port auf Device Server für **ftp** Datentransfer
sm_stpd	11000	Port auf Device Server für **http** Datentransfer
sm_stpd	11443	Port auf Device Server für **https** Datentransfer
rmi GUI	11401	Port auf SEP sesam Server für **GUI** Verbindungen
rmi Web	11403	Webserver auf SEP sesam Server für RestAPI Verbindungen
sm_db	11201	Datenbank Port auf SEP sesam Server (PostgreSQL)
Si3-T	117xx	Port des Si3-T Dedupe Prozesses auf SEP sesam Server und RDS

Die meisten der Ports sind über die Konfigurationsdateien des SEP sesam Servers, des RDS bzw. des Klienten anpassbar.

Um die Sicherheit weiter zu erhöhen ist die Kommunikation zum Client auch verschlüsselt über SSH oder HTTPS möglich.

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com

Firewall-Umgebung

Wachsende Sicherheitsanforderungen in Unternehmen sorgen dafür, dass heute Firewalls nicht nur Unternehmen vom Internet abschotten, sondern auch innerhalb von Unternehmen zur Trennung von Niederlassungen und Bereichen eingesetzt werden.

Um einen Sicherungs-Client über eine solche Firewall zu betreiben sind zwei Schritte notwendig:

- Konfiguration der Firewall-Optionen im Sesam Client
- Freigabe der am Client konfigurierten Ports in der zu überwindenden Firewall

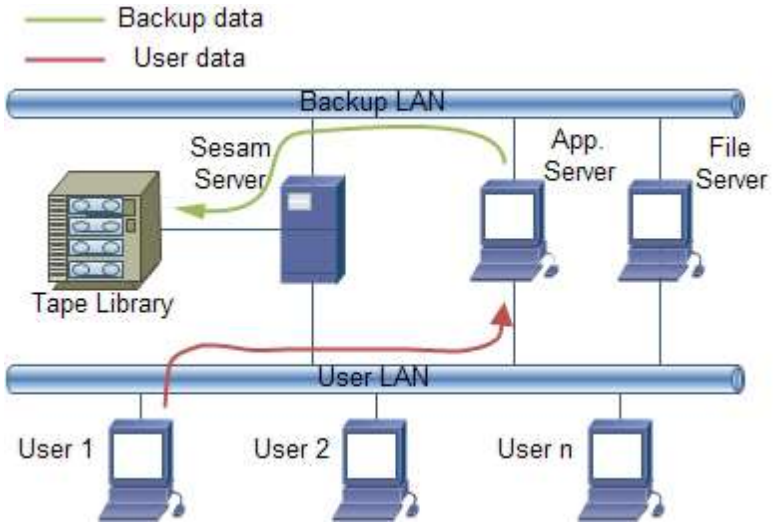
Zu beachten ist, dass es zwei verschiedene Möglichkeiten der Steuerkommunikation und drei verschiedene Möglichkeiten des Datentransfers gibt. Die Steuerkommunikation und der Datentransfer arbeiten völlig unabhängig voneinander, d.h. sie können beliebig kombiniert werden.

Im Standardfall benötigt der SEP sesam somit mehrere Ports zur Kommunikation. Sollten firmeninternen Sicherheitsvorschriften entsprechend streng sein, ist es unter Verwendung des http/https-Protokolls möglich, dass nur noch EIN definierter Port für die Verbindung zum Client benötigt wird.

2.LAN Segment

Zur Sicherung der Daten über ein separates Netzwerksegment sind sowohl der Sesam Server als auch die so zu sichernden Clients mit einer 2.Netzwerkschnittstelle auszustatten. Dann ist jedem 2. Netzwerkinterface eine eigene IP-Adresse und einen Hostnamen zu geben. Die Abbildung zeigt das Netzwerk-Prinzip bei einem solchen Vorgehen. Im ursprünglichen Sinn heißt LAN-free-Backup, dass der Backup-Datenstrom nicht über das User-LAN geht, obwohl bei der Nutzung eines 2.Netzsegments ebenfalls ein LAN zum Einsatz kommt. Man sollte daher ein separates LAN-Segment zum Backup nicht mit einer SAN-Struktur verwechseln. Bei einem SAN können die Sicherungslaufwerke (Shared Drives) nur von einem Host zur selben Zeit exklusiv genutzt werden, während sich bei einem 2.LAN-Segment auch die Parallelisierung von Backups durch den Sesam nutzen lassen.

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTHAMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com



IPv6

Der zur Neige gehende Zahlenraum der heutigen IP-Adressen bedingt für die Zukunft einen Umstieg auf den stark erweiterten Adressraum von IPv6. Die Software muss mit den längeren und neustrukturierten IP-Adressen umgehen können. Dies betrifft jede Software wie Betriebssystem, Middleware und Applikationen. Und damit auch die Backup-Software.

Für SEP sesam sind dafür zur Zeit (Stand Q1/2015) ausgiebige Tests im Gange.

SEP AG, Ziegelstrasse 1, 83629 Weyarn, Germany, +49 8020 180-0
NORTH AMERICA | SEP Software Corp., 4900 Pearl East Circle, West Tower Suite 260, Boulder, CO 80303, USA, +1.303.449.0100
FOR FURTHER INFORMATION, VISIT www.sep.de / www.sepsoftware.com