

Archiv-, Backup- und Copy-Daten vereint

In diesem Beitrag:

- Das intellektuelle Vermögen langfristig erhalten und schützen
- Datenhaltung: Die Rolle des Archivs
- Das Design von «Data Pooling»



Data Pooling: Eine Alternative zur klassischen Datenhaltung

Unter dem Begriff «Data Pooling» soll eine neue, offene Architektur erstmals Archiv-, Backup- und Copy-Daten zusammenführen. Als Kern dient ein schnelles Online-Archiv. Es reduziert den primären Speicher auf die aktuell benötigten Daten. Das Backup von Files und E-Mails erübrigt sich. Ebenso deren Restore durch die IT. Compliance kommt zum Nulltarif dazu.

Alexander Rübensaal,
CEO ABC Systems AG

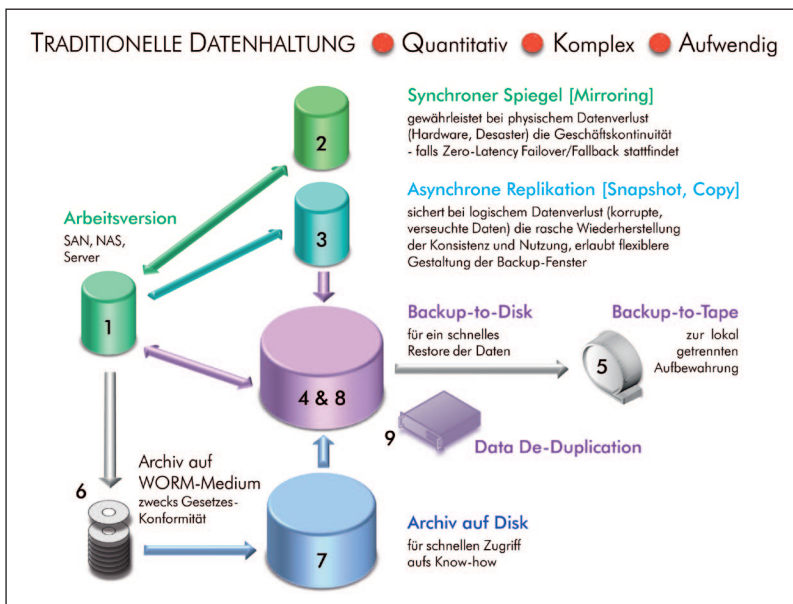
Identische Daten, wie beispielsweise eine Adresse, verwaltet die IT in der Regel «logisch» nur noch einmal. Die einst teuren Disk-Kapazitäten wurden im SAN/NAS physisch konsolidiert. Nach wie vor werden aber dieselben Daten immer wieder aufs Neue auf eigenen Medien gespeichert – bis zu acht Mal oder noch häufiger. Um sie dann anschließend wieder zu deduplizieren.

Die Folge ist ein Hardware-Zoo, in dem sich von allem etwas und immer mehr ansammelt. Das Skalie-

ren erfolgt konsequenterweise parallel in allen Facetten. Für die Speicherindustrie entsteht so nahezu ein Perpetuum mobile. Der Anwender kann nur nachbestellen.

Den Trends ins Auge sehen

93 Prozent aller Informationen werden heute bereits in elektronischem Format erstellt, so eine Studie der University of California in Berkeley. 70 Prozent ihrer wichtigsten Informationen speichern die Unterneh-



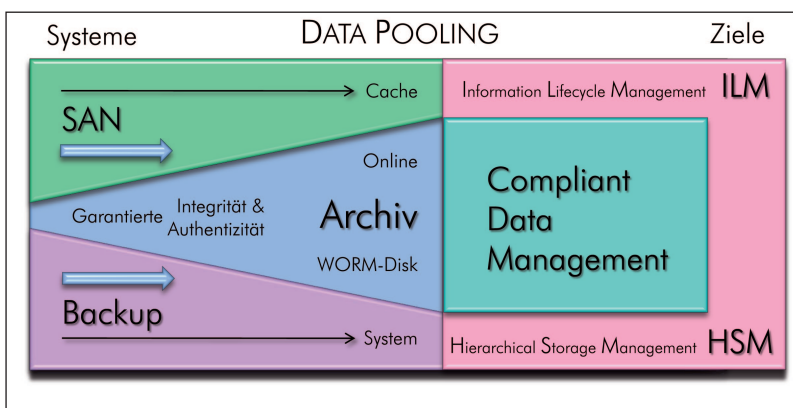
Das Design der gängigen Datenhaltung: Daraus resultiert ein rein quantitatives Konzept, verbunden mit Risiko-reichtum und hohen Kosten.

men auf ihren Messaging-Systemen, wie die Enterprise Storage Group erhoben hat. 80 Prozent der Daten auf den primären Speichersystemen werden nur selten genutzt. Das jährliche Wachstum nicht mehr geänderter Daten beträgt 65 Prozent (Hal Varian, UC Berkeley).

Diese Zahlen belegen klare Trends, denen sich Unternehmen stellen müssen. Schliesslich geht es darum, auf wachsendem Boden selbst zu wachsen und zu bestehen. Einerseits ist die unmittelbare Geschäftskontinuität zu sichern, d.h. die relevanten Daten müssen heute 7x 24h verfügbar sein. Andererseits gilt es, das intellektuelle Vermögen, das auf elektronischen Medien gespeichert ist, langfristig zu erhalten und zu schützen.

Der Erfüllung dieser Anforderungen stehen limitierte Ressourcen gegenüber. Im Sinne einer Engpassplanung ist bei diesem Optimierungsprozess von den unverzichtbaren Aspekten auszugehen. Dazu zählt das heute noch oft vernachlässigte Endlager der Informationen auf digitalen Medien: das Archiv. So wird die elektronische Archivierung gern in Bezug auf die Compliance-Thematik mit der Begründung

Der Weg zum automatisierten Information Lifecycle Management



«Wo kein Kläger, kein Richter» zurückgestellt. Dabei wird rasch erkannt, wie bedeutend der Selbstschutz des eigenen, konformen Handels sowie der Schutz des aufgebauten Know-hows zwecks seiner jederzeitigen Nutzung sind.

Nationale Compliances erstdefinierender Organisationen haben mittlerweile den Status internationaler Standards erreicht, wie z.B. DIN A4. Diesen kann sich ein exportorientiertes Land wie die Schweiz, bzw. ein Glied in einer stark arbeitsteiligen Logistikkette längst nicht mehr entziehen. Daher drängt es sich geradezu auf, dem unvermeidbaren Endlager der Daten – dem Archiv – eine primäre Rolle der Datenhaltung zu übertragen. Nur so lassen sich von vorneherein weitere, parallel zu den Bestehenden geführte und auszubauende Speichertöpfe verhindern. Dies bedingt ein unmittelbares Auslagern der Daten ins Archiv, womit der Rucksack fürs Tagesgeschäft sofort im Sinne von HSM (Hierarchical Storage Management) von Ballast befreit und effizienter wird. Voraussetzung für automatisierbares Handeln ist zudem, im Moment der Datenentstehung gleichzeitig auch ihren «Verfall» festzulegen. Nur so lässt sich ein automatisches Information Lifecycle Management (ILM) verwirklichen.

Alle diese Ansprüche kann nur ein Disk-basiertes, schnelles Online-Archiv erfüllen. Es muss aber noch mehr bieten wie etwa auch die inhaltliche Integrität der Daten langfristig zu garantieren. Dazu bedarf es kryptologischer Algorithmen (SHA-1, MDA-5), welche die Files in eindeutige «Fingerabdrücke» umwandeln und sie zwecks periodischem Abgleich redundant ablegen.

Diese Anforderungen werden von folgenden Lösungen nur auf halbem Weg und daher also nicht erfüllt:

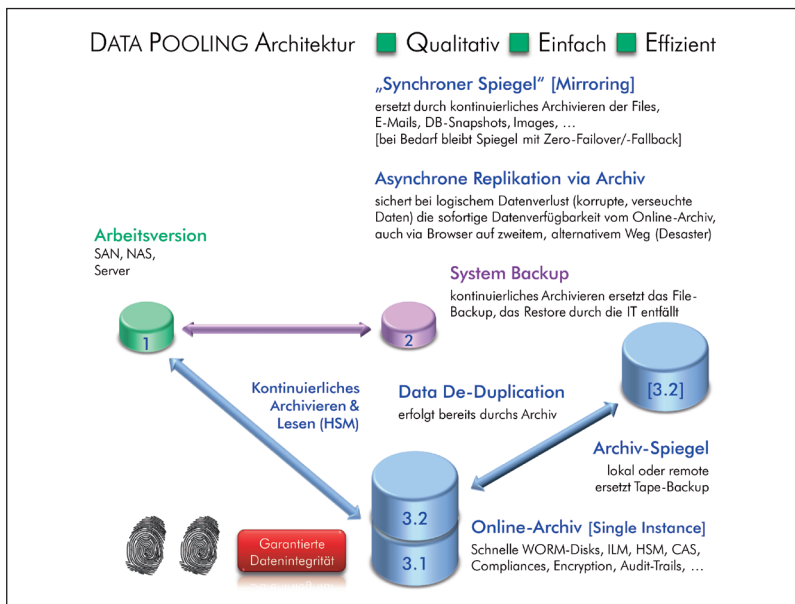
- Archivierungs-Software und -Systeme, die sich dem Anwender gegenüber zwar als Read Only verhalten, aber auf veränderbare oder unzuverlässige Medien aufzeichnen
- Archivierungs-Systeme, die auf WORM-Disks ohne kryptologische Hash-Funktionen schreiben. Denn das Read Only auf dem physischen Layer allein kann die Authentizität der Daten nicht garantieren. Derartig basierte Systeme dann noch zu spiegeln bringt nichts ausser weiteren Kosten.

Der angelegte Massstab mag hart erscheinen, doch er bewahrt den Anwender vor grossem und nur sehr teuer korrigierbarem Schaden.

Da hilft es auch nichts, bei der Produktevaluation auf Zertifizierungen abzustellen. Ohne kritisches Hinterfragen sind sie riskant und rasch fehlleitend. Denn je nach Alter bestätigen sie nur, was dem Zertifizierenden (oft ein Technologiefremder, wie etwa eine Revisionsgesellschaft) nach bestem Wissen mit den dazumal verfügbaren Technologien als erreichbar erschien. Best Effort war bereits genügend. So werden



Archiv-, Backup- und Copy-Daten vereint



Das Design des Data Poolings

Archivsysteme (Software und/oder Hardware) mit dem Siegel «revisionssicher» angeboten, obwohl sie infolge des nur halb zurückgelegten Weges immanent untauglich sind. Solche Verunsicherungen erschweren dem Anwender zusätzlich den Umgang mit der digitalen Archivierung.

Data Pooling

Das Design des «Data Poolings» erscheint daher naheliegend, da jeder Anwender früher oder später von selbst danach verlangen wird. Revolutionär mag es nur dann anmuten, wenn es ein Anbieter propagiert – schaltet er sich damit doch kurzfristig einen verlässlichen Profitgenerator aus. Auch wenn das Design der offenen Architektur des Data Pooling völlig neu ist, so ist es doch wegen seiner geringen Modulanzahl äusserst risikoarm. Dazu tragen auch die Kern-Bau-

Durch die qualitative Ausrichtung des Lösungsdesigns verschiebt sich der Hauptnutzen vom Lieferanten zum Anwender



steine bei, die sich längst über Jahre in der Praxis international bewährt haben (Release 6). Die Architektur überzeugt vor allem durch ihre Einfachheit bei umfassender Funktionalität.

Die bekannten, synchronen Spiegel sind praktisch ausschliesslich so implementiert, dass sie vom primären Speicher auf den sekundären replizieren. Fällt nun der primäre aus, so bleibt allen Servern nur noch der konsolidierte Crash. Wieso also solches noch mit sehr viel Geld «vorbereiten»? Solche Nullwertlösungen lassen sich zwar leicht übertreffen, doch der kontinuierlich erstellte Spiegel des Archivs bringt mehr als 100 Prozent im Vergleich. Er garantiert nämlich auch die inhaltliche Integrität der Daten bis zu deren Verfallsdatum. Denn das erreicht selbst ein synchroner Spiegel mit Zero-Latency Failover/Fallback nicht (es gibt dazu übrigens attraktive Appliance-Lösungen).

Mit der Verlagerung von quantitativer hin zu qualitativer Ausrichtung des Lösungsdesigns verschiebt sich auch der Hauptnutzen vom Lieferanten zum Anwender.

Die von der Schlierener ABC Systems entwickelte ABC (Archiv, Backup, Copy)-Data-Pooling-Architektur konnte sich bereits in einer öffentlichen Ausschreibung gegen 31 interessierte Mitbewerber durchsetzen. Obwohl alles rein Disk-basiert über 100e von TB über 5 Jahre konzipiert, erwies sich die Lösung als die Wirtschaftlichste – auch im Vergleich mit Tape-Libraries. Dabei waren noch nicht einmal die Cartridges, das Wegfallen des Tapehandlings sowie die Tape-Migration nach 5 Jahren (für die Langzeitaufbewahrung) in die Waagschale geworfen worden.

Für ein frisches Lösungsdesign ist es unerlässlich, sich von den Paradigmen und den Vorgaben der Speicherindustrie zunächst einmal zu lösen. Zudem begünstigen knappe Ressourcen stets die Entwicklung neuer Ideen. Reicht das Budget nicht, sind vermehrt Alternativen gefragt. So auch in wirtschaftlich härteren Zeiten, die zudem zu mehr Offenheit gegenüber Neuem führen.

Das Design einer neuen Architektur ist umso besser, je geringer die Anzahl der zur Lösung der Aufgaben notwendigen Layer/Module bleibt. So lässt sich nach Gartner aufgrund der Total Cost of Ownership (TCO), von denen die Anschaffung ja gerade nur 20 Prozent ausmacht, ein Komplexitätslevel bestimmen. Ein zusätzlich benötigtes Modul erhöht daher die Komplexität der Lösung nicht nur um 1, sondern gleich um Faktor 5. Bei einem Vergleich des noch üblichen Designs der Datenhaltung mit dem des Data Pooling resultiert aus dem Wegfall von 4–5 Layern eine Vereinfachung der Komplexität um 20–25 Zähler. Die Wahl der Bausteine ist dabei Match-entscheidend. Nach mehrjähriger Suche und Evaluation erwies sich ein Produkt von Nexsan als praktisch exklusive



Leistungsmerkmale von Assureon ABC

CAS / Fingerprint	Erstellung einer eindeutigen, alphanumerischen Identifikation, um mittels 'Fingerabdruck' die Authentizität jeder archivierten Datei zu garantieren
Versioning	Verwaltung verschiedener Versionen derselben Datei
Encryption	Verschlüsselung der Daten mit AES 256 Bit
Security	Überwachte Kontrolle der Zugriffe aufs Archiv inklusive Audit Trails
WEB Interface	Passwort geschützter des Zugriff aufs Archiv über Browser
ILM	ILM Information Lifecycle Management (Aufbewahrungsdauer der Daten)
HSM / Backup	Automatisches Verlagern von Daten vom Server und Client ins Archiv nach vordefinierten Regelwerken
Reacquisition	Schneller Online-Aufruf archivierter Dateien im Originalformat
Reporting	Nachweis über den zeitlichen Verlauf des Datenzugriffs, Vollständigkeit des Archivs, rechtsverbindliche Löschung nach der Aufbewahrungsfrist
Classification	Datenklassifikation in Übereinstimmung mit den Unternehmensvorgaben
Time Stamp	Unabhängige, externe Datums-/Zeitstempelung zwecks rechtlicher Beweiskraft der Dateierstellung
Database	Speicherung und Indexierung der Informationen in einer internen Datenbank
Disaster Prevention	Aktiv-Aktiv Replikation der Daten zwischen entfernten Standorten

Checkliste zur Evaluation einer Lösung fürs Data Pooling

sich das Universalarchiv bereits durch die Entlastungen und Einsparungen beim primären Speicher und Backup/Restore finanziert hat.

Das Zusammenlegen verschiedener Bereiche der Datenhaltung verlangt aber noch mehr. Denn je weniger Layer/Module zur Lösung eines Problems erforderlich sind, umso besser und gleichzeitig einfacher ist das Lösungsdesign. Dazu zählt auch, vielfältige Aufgaben gleich auf einen Schlag gemeinsam erledigen zu können.

Bei der standardmässigen Multisite-Replikation kann der Anwender wählen, die Daten je Standort einfach oder redundant zu archivieren. Das Auto-failover für Reads zwischen den Archiven/Standorten ist als Standard implementiert. Der Vordereingang zum Universal-Archiv befindet sich völlig transparent bei den bedienten Servern. Sollten diese einmal ausfallen oder nicht zur Verfügung stehen,

Was Assureon als universelles Online-Archiv gleichzeitig in den Bereichen des primären Speichers und Backup/Restore leisten kann

Grundlage fürs Data Pooling, und zwar aufgrund seines Funktionsumfangs. Assureon ABC – Archiv & Backup Center – entstand nach den Vorgaben internationaler Normen wie FDA, HIPAA, SEC oder SOX. So kann mit all den bei Assureon ABC bereits vorhandenen «Schaltern» auf einen Schlag auch den einschlägigen Schweizer Gesetzen wie BGA, DBG, GeBüV, MWSTG, OR und StGB entsprochen werden. Quasi zum Nulltarif, weil ja der gesamte Funktionsumfang zählt und

kann der «Hintereingang» zu den Daten benutzt werden. Dieser entspricht dem Datenzugriff via Browser – entsprechend den gewährten Rechten.

Optional steht auch eine Volltextsuche zur Verfügung. Diese kann sich alternativ zur Klassifikation der Daten über ein ECM, DMS, etc. anbieten und nach Bedarf auch ein Business Re-Engineering übers Universal-Archiv erlauben.

Zumindest folgende Fragen sind dabei zu prüfen:

- Lohnt sich der tägliche Klassifikationsaufwand im Verhältnis zur eventuellen nur seltenen Wiederverwendung der archivierten Daten?
- Wie lässt sich die Vollständigkeit des Archivs bei manueller Klassifikation der Daten garantieren?

Je mehr die primäre Speicherung der Daten auf Assureon übertragen wird, desto mehr gewinnt der schnelle Zugriff an Bedeutung. Dazu können optional High Speed Links (20 Gbps Infiniband) zwischen den Servern und dem Archiv implementiert werden. Ein Online-Archiv mit 2x 4TB oder 1x 8TB Netto-Kapazität und ohne Limitation der Anzahl der archivierten Objekte beläuft sich kostenmässig auf rund 35 000 Franken und kann als Blackbox (n-way Cluster) auf etliche Petabytes im Feld skalieren. In Anbetracht des Funktionsumfangs kann dies für KMU bereits dazu führen, auf ein separates SAN/NAS und ECM, DMS, etc. zu verzichten. Das Archiv kann bei Bedarf jederzeit vollumfänglich oder partiell auf Disk exportiert werden. Das Vorgehen ist allerdings Compliance-entsprechend zu verbiefen.

Der Bund stellt seit kurzem gratis ein Softwaretool zur Verfügung, mit dem sich auch relationale Datenbanken (Oracle, MS SQL) in eine Langzeitarchivierung einbinden lassen. □

Beispiel eines Lösungslayouts aus der Praxis

Vereinfachung von SAN/NAS und Backup/Restore

Vielältige Aufgaben ...	Assureon ABC löst alle in Einem ...
San Entlastung & Redimensionierung	
SAN stösst an Kapazitätsgrenze	mit HSM werden selten, kaum mehr genutzte Daten ins Online-Archiv verschoben
SAN ist zu ersetzen	Daten werden im Archiv auf File-Ebene automatisch dedupliziert
SAN ist nicht gespiegelt	Ersatz kann nur noch als Cache dimensioniert werden, weil das Online-Archiv so schnell wie's SAN ist zudem sind Daten stets im Archiv redundant verfügbar und erst noch mit garantierter Integrität
SAN-Unterhalt ist zu hoch	Daten werden im Archiv redundant nachgeführt, kontinuierlich, quasi synchron
SAN (noch) nicht vorhanden	SAN redimensionieren auf Disk-Cache, da Daten im Online-Archiv schnell abrufbar
SAN Ausfall bei Disaster	Verzicht zugunsten von Servern, da kontinuierliche Verschiebung ins Archiv mit redundanter Speicherung
Zugriff auf Daten im Archiv kann bei SAN-Ausfall direkt über Browser erfolgen, was einem zweiten, unabhängigen Zugang zu den Daten entspricht	
Ersatz von File-Backup & Restore durch IT	
Backup stösst an Kapazitätsgrenze	Ausbau erübrigt sich fürs File-Backup, nur noch System- und DB-Backup
Backup ist zu ersetzen	Ersatz erübrigt sich fürs File-Backup, nur noch System- und DB-Backup
Backup-Zuverlässigkeit	kontinuierliches Speichern im Archiv, das selbstheilend korrupte Daten wieder herstellt
Remote-Backup der Aussenstellen	einfache Zentralisierung des File-Backups durch Automatisierung
Backup Operating ist aufwendig	entfällt fürs File-Backup
Backup-to-Disk/-Tape	Disk- und Cartridge-Kapazitäten werden fürs File-Backup nicht mehr benötigt
Daten-Deduplizierung	erfolgt automatisch im Archiv auf File-Ebene
Restore ist aufwendig	entfällt fürs File-Restore, Anwender kann Daten nicht verändern

DATA POOLING Anwendungslayout

