

Ordnung mit digitaler Archivierung

Im digitalen Zeitalter werden Systemverantwortliche und CIOs nahezu täglich mit dem Zuwachs an elektronischen Daten und deren Verwaltung konfrontiert.

VON NIKOS BACHZETSIS

IN KÜRZE

- Eine Archivierungslösung muss rechtliche und regulatorische Vorgaben befolgen.
- Um den Überblick zu behalten, ist eine ausgeklügelte Archivierungsstrategie wichtig.
- Daten werden archiviert, wenn sie sich nicht mehr verändern und nur noch selten verwendet werden.

INFORMATION WORKER

Daten können nicht einfach irgendwie gespeichert werden. Unternehmen müssen dabei neben rechtlichen und regulatorischen Vorgaben auch Anforderungen an Recovery Time Objective (RTO) berücksichtigen. Zentral ist aber vor allem die Frage, wie Informationen während ihres gesamten Lebenszyklus schnell aufgefunden werden können. Viele Unternehmen greifen daher auf digitale Archivierungslösungen zurück. Sie erhoffen sich dadurch im Infrastruktur-Bereich eine Entlastung der Primärspeicher sowie eine Verkürzung des Backups. Ausserdem ermöglichen die Lösungen eine Revisions-sichere Aufbewahrung von Belegen sowie eine Kostenreduktion und eine Prozessoptimierung. Eine Studie aus dem Jahre 2007 von Accenture und Avanade bestätigt, dass das Top-Thema für die nächsten 24 Monate bei Schweizer CIOs die digitale Archivierung ist.

Strategie ist entscheidend

Den wichtigsten Teil beim Design einer Archivierungslösung bilden die Anforderungsanalyse und die Erarbeitung der Strategie. Ein wichtiger Aspekt ist, dass die Entwicklung einer Strategie nichts mit der Produktfindung

zu tun haben darf, sondern sich als Gerüst versteht, welches beschreibt, was wie, wo und wann archiviert zu werden hat. Eine Archivierungsstrategie kann mehrere Projekte beinhalten. Es ist daher ratsam, ein Archivierungsprojekt in kleinen Schritten zu starten und dieses dann kontinuierlich zu erweitern. In einem ersten Schritt sollten die archivierungsabhängigen Prozesse analysiert werden, um daraus den Business Case abzuleiten, welcher die wichtigste Entscheidungsgrundlage für die Firmenführung bildet.

Archivierung versus Backup

Digitale Archive sind Aufbewahrungsorte für betriebliche Informationen, welche nicht mehr bearbeitet werden, auf die aber in Zukunft zugegriffen werden muss. Es handelt sich um Daten mit festem Inhalt, welcher nicht mehr verändert werden darf. Das primäre Ziel digitaler Archive ist es, diese Aufzeichnungen vor Änderungen zu schützen und die notwendigen Mittel zur Verfügung zu stellen, um die Aufzeichnungen zu finden.

Backups hingegen sind Kopien aktiver Produktionsdaten, welche verwendet werden, wenn es Probleme gibt und eine Kopie zur Rückgewinnung der Daten erforderlich ist.

Backups sind generell kurzlebig und werden oft überschrieben.

Demgegenüber konzentrieren sich Archive nicht auf die Wiederherstellung, sondern erlauben den Abruf von Informationen auf granularer Ebene wie zum Beispiel einer Datei, eines E-Mails oder anderer einzelner Unterlagen. Archive sind keine Kopien von

Produktionsinformationen, sondern vielmehr das Original. Wenn Daten sich nicht mehr ändern oder nicht mehr häufig verwendet werden, ist es sinnvoll, sie in ein Archiv zu verlagern, wo sie ausserhalb des Backup-Bereiches sind, aber noch auf sie zugegriffen werden kann.

Verschiedene Archivlösungen

Bei einer Archivierungslösung handelt es sich nicht um eine Hard- oder Software, welche autonom in einer bestehenden Umgebung installiert und betrieben werden kann. Vielmehr ist es eine Kombination von verschiedenen Komponenten und Technologien. Dies ist einer der Hauptgründe, wieso die Einführung einer Archivierungslösung komplex sein kann. Sehr gute und detaillierte Kenntnisse über die bestehende Infrastruktur, die Applikationen und Prozesse sind ein Muss und spielen in bezug auf die Einführung, Integration und den Betrieb eine zentrale Rolle.

Bei vielen Archivierungsprojekten stellt sich die Frage, ob dedizierte oder integrierte Archivierungsapplikationen zum Einsatz kommen sollen. Beide bringen unterschiedliche Vor- und Nachteile (vgl. Grafik auf S. 39).

Neben der Archivierungsapplikation stellt der Archivspeicher einen wichtigen Teil jeder Lösung dar. Der Begriff HSM (Hierarchical Storage Management) beschreibt das Konzept des effizienten Einsatzes von verfügbaren Speichertechnologien und -ressourcen. Moderne Archivierungslösungen verfügen über ein integriertes Lifecycle-Management, welches Archivdaten aufgrund ihres gegenwärtigen Lebensabschnittes in das vorgesehene Speichersystem verschieben, damit verfügbare Ressourcen optimal verwendet und ausgelastet werden.

Mit der rasanten Entwicklung und tendentiell sinkenden Preisen im Bereich der Disk-basierten Speichertechnologie zeichnet sich klar ein Trend ab, welcher weg von den klassischen

VOR- UND NACHTEILE VON ARCHIVLÖSUNGEN

DEDIZIERTE ARCHIVAPPLIKATIONEN

Vorteile

- ▶ Schnelle Implementierung und Resultate
- ▶ Erlaubt «Best-of-breed»-Vorgehen

Nachteile

- ▶ Mehrere Such- und Retrieve-Schnittstellen
- ▶ Managementkosten (mehrere Anbieter)

INTEGRIERTE ARCHIVAPPLIKATIONEN

Vorteile

- ▶ Ein Archiv mit einer Such- und Retrieve-Schnittstelle
- ▶ Managementkosten (nur ein Anbieter)

Nachteile

- ▶ Lange und relativ komplexe Implementierung
- ▶ Lösungen sind normalerweise nicht überall «Best-of-breed»

optischen Speichern zu Harddisk-basierten Soft-Write-once-read-many-Lösungen (WORM) geht. Die Zugriffszeiten werden dadurch verbessert und Datenmigrationen klar vereinfacht.

E-Mail-Archivierung

Die Einführung von E-Mail-Archivierung ist klar eine der am häufigsten eingesetzten Anwendungen. Die Gründe dafür sind sicherlich das enorme Datenaufkommen auf den E-Mail-Servern und die Tatsache, dass sich laut einer Statistik der Enterprise Strategy Group (ESG) aus dem Jahre 2005 bis zu 75 Prozent des geistigen Eigentums von Unternehmen in E-Mails befinden.

Grundsätzlich unterscheidet man zwei Typen von E-Mail-Archivierung: Die Mailbox-Archivierung und die Journal-Archivierung. Bei der Mailbox-Archivierung ist das Ziel die Archivierung der Inhalte von elektronischen Postfä-

chern. Die Journal-Archivierung zeichnet den gesamten internen und externen E-Mail-Verkehr auf.

Mailbox-Archivierung

Eine Mailbox-Archivierung entlastet den Primärspeicher, verkürzt Backup- und Recovery-Zeiten, steigert die Performance der E-Mail-Server, bietet eine Volltextsuche und schafft persönliche Archive ab.

In der Praxis findet man vor allem drei Methoden der Mailbox-Archivierung: Im Conditional Mode werden E-Mails beispielsweise nach Alter oder Grösse archiviert, während im Threshold Mode Nachrichten beim Erreichen eines Speichergrenzwertes archiviert werden. Im Manual Mode bestimmt der Anwender derweil selber, nach welchen Kriterien die E-Mails archiviert werden.

Bei allen beschriebenen Methoden werden die E-Mails ins Archiv verschoben, und auf dem E-Mail-Server verbleiben sogenannte Shortcuts, welche auf das archivierte Element verweisen.

Journal-Archivierung

Die Journal-Archivierung unterstützt Unternehmen bei der Erfüllung der Rechtskonformität, der Beweisführung und dem Records Management. Vom gesamten E-Mail-Verkehr wird jeweils eine Kopie jeder Nachricht archiviert. Je nach Standort gelten unterschiedliche Gesetzgebungen und Anforderungen bezüglich der Aufzeichnung von E-Mails. Generell kann man feststellen, dass bei Erfüllung der schweizerischen Auflagen diese für sehr viele Standorte weltweit auch als genügend gelten. Im konkreten Fall müssen jedoch immer länderspezifische Abklärungen getroffen werden. Damit zu einem späteren Zeitpunkt ein E-Mail in den entsprechenden Kontext gesetzt werden kann, genügt die Einführung einer E-Mail-Archivierungslösung nicht. Für diese Funktionalität ist eine Koppelung zwischen dem Nachrichtenarchiv und einer ECM- oder DMS-Lösung notwendig.

Papier-Belege archivieren

Das Digitalisieren und Archivieren von Papierdokumenten findet in den verschiedensten Bereichen Anwendung. Typische Beispiele sind das Einscannen von Kundenbestellungen, Lieferscheinen oder Kreditoren-Rechnungen. So werden Informationen schnell durchsuch- und auffindbar, und Papierarchive werden eliminiert. Zudem können Inhalte dank der OCR/ICR-Technologie automatisch extrahiert werden. Des Weiteren wird eine Prozessoptimierung und -automatisierung möglich, sobald die Dokumente und deren Inhalt elektronisch vorhanden sind.

Nach dem Einscannen von Dokumenten erhält man deren Abbild (Image) und Inhalte (Indexwerte) in elektronischer Form. Es gilt nun, diese Daten in strukturierter Form zu speichern. Typischerweise übernimmt diese Aufgabe ein DMS (Dokumenten-Management-System). Die Hauptaufgabe des DMS ist die Verwaltung der Images und der dazugehörigen Indexwerte und die Konvertierung sämtlicher Inhalte in ein Langzeitdatenformat wie z.B. TIFF, PDF, PDF/A. Im weiteren verwaltet das DMS den Lebenszyklus aller Dokumente. In bezug auf die Archivierung ist das ILM (Information Lifecycle Management) eine wichtige Funktion, da dort das System aufgrund definierter Kriterien entscheidet, auf welchen Speichertyp und in welcher Art und Weise die Dokumente abgelegt werden. Falls Dokumente rechtskonform gespeichert oder die Originale (Papier) vernichtet werden sollen, müssen die archivierten Daten auf einen WORM-Speicher geschrieben oder mittels digitaler Signatur signiert werden.

Die Erfahrungen zeigen, dass kein Archivierungsprojekt gleich ist wie ein anderes. Dies liegt nicht nur an den unterschiedlichen Anforderungen, sondern auch daran, dass ein digitales Archiv eine zentrale Rolle in der bestehenden Infrastruktur einnimmt.

NIKOS BACHZETSI IS SENIOR CONSULTANT
BEI ARCPLACE AG.

PDF/A ALS LANGZEITFORMAT

Um Zugänglichkeit und Reproduzierbarkeit der archivierten Geschäftsbelege zu garantieren, ist es ausschlaggebend, dass die Aufzeichnungen in Datenformate umgewandelt werden, welche langfristig kompatibel sind. Die am häufigsten verwendeten Formate für die elektronische Archivierung sind TIFF und PDF.

TIFF garantiert die Reproduzierbarkeit langfristig und besitzt eine bewährte Struktur. TIFF ist leicht zu erzeugen, kann aber nicht so einfach durchsucht werden. Für das PDF-Format hat die ISO im Jahr 2005 ein neues Dateiformat definiert, das als PDF/A (ISO 19005-1) bekannt ist. Dieses neue Format bietet einen Mechanismus für die Darstellung elektronischer Dokumente in einer Art, welche ihr Erscheinungsbild über die Zeit bewahrt und dabei unabhängig ist von den Werkzeugen und Systemen, die für deren Erstellung, Speicherung oder Aufteilung verwendet wurden. Ausserdem können die PDF/A-Dokumente leicht durchsucht werden.

Für weitere Informationen kontaktieren sie bitte:



Arcplace AG, Seehofstrasse 6, CH-8008 Zürich
T: +41 44 260 80 00 / www.arcplace.ch / info@arcplace.ch

Digital Archiving Solutions