



Leitfaden zum Thema „Information Lifecycle Management“

(Stand: April 2004)

Inhaltsverzeichnis

1	Zusammenfassung	4
2	Definition und Abgrenzung.....	4
3	Treibende Kräfte	5
3.1	Datenwachstum	5
3.2	Gesetzliche und regulatorische Bestimmungen	5
3.3	Wert der Daten im Laufe des Lebenszyklus.....	6
3.4	Kostendruck.....	6
3.5	Technologische Innovationen	6
4	ILM-Konzept	7
4.1	Grundsätzliche Arbeitsweise.....	7
4.2	Externer Input	7
4.2.1	Wert der Informationen	7
4.2.2	Service Level Agreements.....	8
4.2.3	Zugriffsrechte / Datenintegrität.....	8
4.3	Speicherhierarchien / Speicherklassen	8
4.4	Strukturierung der Daten.....	8
4.5	Nutzen	9
5	Vorleistungen.....	9
5.1	Konsolidierung und SAN / NAS Infrastruktur.....	9
5.2	Virtualisierung.....	10
5.3	Storage Resource Management	10
6	ILM-Modell.....	10
7	Datensicherung.....	12
8	Praktische Beispiele zu ILM.....	13
8.1	E-Mail-Archivierung	13
8.1.1	Anforderungen an die E-Mail-Archivierung	13
8.1.2	Elemente einer ILM-Strategie bei der E-Mail-Archivierung.....	13
8.1.3	Implementierung einer ILM-Strategie bei der E-Mail-Archivierung	13
8.2	Lebensversicherung	14
9	ILM und System Managed Storage.....	14
10	Schlussbetrachtung	15

1 Zusammenfassung

Information Lifecycle Management (ILM) hat das Potential, Kostenpositionen (Total Cost of Ownership, Produktivität etc.) und den Umgang mit Service Level Agreements wesentlich zu verbessern. Dies verspricht Lösungen für drängende Probleme. Der Begriff Information Lifecycle Management (ILM) wird heute leider in vielen und zum Teil unscharfen Definitionen verwendet. Das heißt, jedes Mal, wenn für ein Informationsobjekt eine Entscheidung erforderlich ist oder auf ein Informationsobjekt eine Aktion angewandt wird, kann man auch das erforderliche Umfeld mit ins Spiel bringen und unter ILM positionieren. Unscharfe Definitionen ermöglichen es Herstellern, Beratern, Systemhäusern den größten Teil ihrer existierenden Produkte unter ILM zu positionieren. ILM ist jedoch kein Produkt, sondern eine Kombination aus Prozessen und Technologien. Ziel ist es, die richtige Information zur richtigen Zeit am richtigen Ort bei geringsten Kosten zu haben. Dies wird in einem permanenten Optimierungsprozess erreicht. Der Optimierungsprozess erhält seine Parameter zum einen durch externe Vorgaben (Wert der Informationen, Sicherheitsanforderungen, Service Level Agreements etc.) und zum anderen durch die vorhandene Speicherhierarchie mit den darunter liegenden Kostenstrukturen. Als Ergebnis des Optimierungsprozesses ergeben sich Entscheidungen, wo Informationsobjekte am besten zu speichern sind bzw. wie Backup-, Replikations-, Verdrängungs-, Verlagerungs- und Archivierungsfunktionen zu steuern sind. Für ein effizient arbeitendes ILM sind gewisse Vorleistungen erforderlich. Virtualisierung für den Online-, Nearline- und NAS-Bereich sind Beispiele. Aufgrund der Trennung der logischen Sicht von der physikalischen Sicht wird ILM in die Lage versetzt, Informationsobjekte aufgrund der Prozessentscheidungen optimal zu platzieren.

2 Definition und Abgrenzung

Information Lifecycle Management (ILM) ist ein Storage Management-Konzept, welches Informationsobjekte während der gesamten Lebenszeit aktiv verwaltet. Dabei bestimmt eine Regelmaschine unter Berücksichtigung von Vorgaben aus den Geschäftsprozessen und der Bewertung der Kostenstrukturen der Speicherhierarchie in einem Optimierungsprozess den best geeigneten Speicherplatz für die verwalteten Informationsobjekte. Vorgaben aus Geschäftsprozessen sind z.B. der Wert eines Informationsobjektes, gesetzliche oder regulatorische Vorgaben, Aufbewahrungszeiten, Zugriffsrechte, Service Level Vereinbarungen. Informationsobjekte sind z.B. Logische Units (LUN), Volumes, Dateien, Dateiverzeichnisse, Datenbank-Container, Sätze in einer Datenbank, E-mail-Nachrichten, Informations-Container für Web-Systeme. Im Optimierungsprozess werden Ereignis-gesteuert oder in geeigneten Intervallen sich im Laufe der Lebenszeit der Informationsobjekte verändernde Parameter (z.B. veränderter Wert der Informationsobjekte) neu bewertet und gegebenenfalls automatisch Aktionen angestoßen (z.B. Verlagerung von Informationsobjekten innerhalb der Speicherhierarchie, Archivierung, Löschen).

ILM ist kein Produkt sondern eine Kombination aus Prozessen und Technologien.

Hinweis: In Fachkreisen ist derzeit eine rege Diskussion zur Abgrenzung der Begriffe Information Lifecycle Management und Data Lifecycle Management (DLM). So wird z.B. die Auffassung vertreten, dass ausschließlich der Begriff DLM verwendet werden sollte, da der Begriff ILM für Verwirrung unter den potentiellen Anwendern sorgen könnte. Nach dieser Auffassung sind die unter dem Begriff ILM geführten Systeme nur in der Lage, ein reines Data Lifecycle Management mit der Übertragung und Archivierung von Daten zwischen verschiedenen hierarchischen Speichersystemen zu gewährleisten. Unter ILM jedoch wären auch erweiterte Funktionen wie Data Mining, Prozessverarbeitung und Entdeckung von Geschäftstrends zu verstehen. Diese Funktionen wären heute noch nicht integriert. Einige Analysten unterstützen diese Argumentation.

BITKOM möchte an dieser Stelle keine Bewertung abgeben und auch keine Position für die eine oder andere Meinung beziehen, sondern auf die unterschiedlichen Terminologievorstellungen

im Markt hinweisen. Im Rahmen dieses Leitfadens werden aus Vereinfachungsgründen deshalb die Begriffe DLM und ILM gleichwertig betrachtet und auch so beschrieben.

3 Treibende Kräfte

Die Notwendigkeit, ein effektives ILM umzusetzen, ist im Prinzip nicht neu. In der Vergangenheit wurde dieser Ansatz eher punktuell und weniger formal gelebt. Das IT- und Speicher-Umfeld hat sich aber in mehreren Bereichen gravierend geändert, so dass heute eine professionelle, umfassende Konzeption erforderlich ist. Wesentlich treibende Kräfte sind dabei das ungebremste Datenwachstum, die internen und externen Vorgaben in Form von gesetzlichen und regulatorischen Bestimmungen, der sich weiter verstärkende Kostendruck sowie die technologischen Innovationen in den letzten Jahren.

3.1 Datenwachstum

Jedes Unternehmen bzw. jede Institution kann das nach wie vor anhaltende extreme Datenwachstum in den unterschiedlichsten Bereichen mit eigenen Erfahrungen belegen.

Das prominenteste Beispiel ist sicherlich der E-Mail-Bereich; allgemein als Messaging-Systeme bezeichnet. Gartner sagt voraus, dass die Zahl, der täglich versandten E-mails, im Jahr 2005 ein Volumen von 36 Milliarden erreichen wird. Dies entspricht einer Verdreifachung, bezogen auf das Jahr 2001. Im Storage Magazine Artikel „Optimize your storage for fixed content“ (Ausgabe Mai 2003) wird die Aussage gemacht, dass ein Unternehmen mit ca. 3.000 Mitarbeitern ungefähr ein Terabyte an E-mail-Daten pro Jahr produziert.

Ein weiterer Bereich mit extrem hohem Wachstum ist im Gesundheitswesen zu finden. Im selben Storage Magazine Artikel wird aufgezeigt, dass ein Picture Archive and Communications System (PACS) in einem größeren Krankenhaus leicht mehr als fünf Terabytes pro Jahr an digitalen Röntgen- bzw. MRI-Informationen (Magnetic Resonance Imaging) erzeugen kann.

Ähnliche Wachstumsraten dürften auch in anderen Bereichen anzutreffen sein. Beispiele sind Forschung und Entwicklung, Bankensektor, Versicherungen, Video Editing, Video Streaming, Customer Relationship Management (CRM), Quick decision (parallele inhaltliche Auswertung mehrerer Quellen auf kritische Stichworte), Fraud detection und Überwachung.

3.2 Gesetzliche und regulatorische Bestimmungen

Gesetzliche und andere regulatorische Bestimmungen legen z.T. fest, wie und wie lange Informationen gespeichert und in welcher Form sie auf Anforderung zur Verfügung gestellt werden müssen. Die Basis sind Gesetzgebungen oder Vorgaben auf internationaler, nationaler, kommunaler oder Verbands- bzw. Organisations-Ebene. Ebenso sind Firmen-interne Vorgaben von Bedeutung. In internationalen Präsentationen bzw. Dokumentationen werden häufig amerikanische Gesetze und Vorgaben wie z.B. SEC Rule 17a-4, 21CFR Part 11, HIPPA, Sarbanes-Oxley, GoBs, DoD 5015.2 zitiert. Vergleichbare Bestimmungen können auch für andere Länder zitiert werden. In Deutschland greifen z.B. die Grundsätze der ordnungsgemäßen Buchführung (GoB), die Grundsätze zum Datenzugriff und zur Prüfbarkeit digitale Unterlagen (GDPdU), das Handelsgesetzbuch (§ 238 HGB) und die Abgabenordnung (§ 140 – 148 AO). Darüber hinaus gibt eine Vielzahl weiterer Regelungen. Ein Beispiel wären Vorgaben der Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin), die für Banken, Versicherungen und den Aktienhandel relevant sind. Generell ist in Europa ein Trend zu einer gewissen Harmonisierung zu erkennen, kurzfristig sollte die Erwartungshaltung aber nicht zu hoch gesteckt werden. Es ist Aufgabe der Verantwortlichen in Zusammenarbeit mit internen oder externen Revisionen, Wirtschaftsprüfern oder mit anderen Sachkundigen die Speicherart und Aufbewahrungsfristen festzulegen. Generell scheinen die meisten Bestimmungen größtenteils Technologie-neutral zu sein, so dass ein großer Handlungsspielraum in der technischen Umsetzung in einem ILM gegeben ist. Eines ist aber sicher, der Umfang der ordnungsgemäß zu speichernden Informationen und die Länge der Aufbewahrungsfristen werden erheblich zunehmen.

3.3 Wert der Daten im Laufe des Lebenszyklus

Der Wert der Daten ändert sich im Laufe der Zeit. Diese Tatsache kann am besten an einem Beispiel erläutert werden. Dazu werden einige Prozesse einer Banktransaktion simplifiziert betrachtet. Die wichtigste Information ist der Kontostand. Dieses Datum wird immer einen hohen Wert haben, da daraus Forderungen und Verbindlichkeiten abgeleitet werden. Eine konkrete Umsatzbewegung (Abhebung / Einzahlung) muss zuerst einmal bis dem Zeitpunkt online vorrätig gehalten werden, bis der Kontoauszug gedruckt ist. Ab diesem Zeitpunkt muss die einzelne Bewegung nicht mehr im Online-Bereich gehalten werden. Der Wert wird aber zur Erstellung von monatlichen Umsätzen bzw. zur Berechnung der Bilanzsumme benötigt. Nach Veröffentlichung der Bilanzsumme spielen die Einzelwerte eine untergeordnete Rolle, haben somit geringeren Wert. Der Wert dieser Daten ändert sich aber schlagartig, wenn entweder Reklamationen zu bearbeiten sind oder wenn Revisionen oder gar Steuerprüfungen anstehen. Unter Umständen muss auf einzelne Werte zugegriffen werden. Eine permanente Wertänderung tritt erst dann ein, wenn Reklamationsfristen bzw. vorgeschriebene Archivierungszeiten überschritten werden. Es liegt dann im Ermessen der Bank, die Daten zu löschen oder sie noch für einen längeren Zeitraum vorzuhalten. In manchen Fällen gibt es aber auch Situationen, wo Daten nach einer bestimmten Zeit gelöscht werden müssen. Derartige Fälle sind in anderen Bereichen bekannt, wo Überwachungsdaten (z.B. Videoüberwachung) nach einer bestimmten Anzahl von Tagen gelöscht werden müssen. Möglicherweise erfordert dies bestimmte Techniken, so dass die Daten auf keinen Fall rekonstruiert werden können.

3.4 Kostendruck

Reduzierung von Kosten ist ein permanentes Thema. Im Laufe der Zeit gab es nur kleinere Zeitspannen, wo der Kostendruck etwas in den Hintergrund getreten ist. Die E-Business-Phase ist ein derartiges Beispiel. Bei manchen Firmen hatte die Expansion die höchste Priorität. Im Ergebnis ging aber die Rechnung bei sehr vielen Firmen nicht auf. Heute hat das Thema Kostenreduktion wieder höchste Priorität. Hauptgründe mögen in der Globalisierung und der wirtschaftlich angespannten Lage in vielen Regionen liegen. Kostenreduktion muss an den richtigen Stellen ansetzen. Kosten können in Personal-, Technologie- und Prozesskosten aufgeteilt werden. Personalkosten tendieren permanent zu steigen, wenn man von einem konstanten Personalstand ausgeht. Technologiekosten sind in den vergangenen Jahren mit einem zweistelligen Prozentsatz gefallen, unter der Annahme gleiche Kapazität und dass kein Technologiewechsel vorgenommen wurde. Prozesskosten sind größtenteils individuell, können aber durch Automatisierung stark beeinflusst werden. Wenn man es auf einen kurzen Nenner bringt, dann erzwingen die Randbedingungen, dass mit weniger Input an Ressourcen immer mehr geleistet werden muss. ILM ist hier ein äußerst geeigneter Ansatz. ILM stellt automatisierte Prozesse bereit, um vorhandene Technologie optimal auszunutzen. Dadurch ergibt sich zusätzlich eine Reduzierung von Personalkosten, da die richtige Information zum richtigen Zeitpunkt am richtigen Ort automatisch bereitgestellt wird.

3.5 Technologische Innovationen

Mit dem Aufkommen neuer Speichertechnologien, z.B. ATA-Platten (Advanced Technology Attachment) mit unterschiedlichen Preis-/Performance Positionierung im Vergleich zu Fibre Channel / SCSI Platten, wird das klassische Speichermodell erweitert. Heute haben wir mindestens drei Haupthierarchien, wenn man die unterschiedlichen Technologien im Online-Bereich (High-end, Mid-tier bzw. Entry RAID-Systemen) und die unterschiedlichen Technologien im Nearline-Bereich (Bänder oder auch Cartridges, CD, DVD, etc.) in Betracht zieht. Die Hierarchie beginnt mit dem hoch performanten Online-Bereich. Die wesentliche Charakteristik im Online-Bereich ist, dass die Daten unverzüglich verfügbar sind. Im Anschluss an den Online-Bereich sind Systeme mit ATA-Platten positioniert. Diese Hierarchie hat ebenfalls Direkt-Zugriffseigenschaften ist aber im Gegensatz zum hoch performanten Online-Bereich mit hoch kapazitiven Platten mit moderateren Durchsatzwerten und geringeren Kosten pro Megabyte aufgebaut. Wesentliche Einsatzszenarien sind in der Archivierung und in der Platte-zu-Platte Datensicherung zu sehen. Im Online-Bereich liegen die Zugriffszeiten für einzelne Ein/Ausgaben im Millisekunden-

den-Bereich. Am anderen Ende der Hierarchie steht der Nearline-Bereich mit Bandtechnologien etc. bei wesentlich geringeren Kosten pro Megabyte. In dieser Hierarchie liegen die Zugriffszeiten im Bereich von einigen 10 Sekunden bis in den Minutenbereich, unter der Voraussetzung, dass die Bänder in automatischen Libraries verwaltet werden. In Ergänzung an die Nearline-Ebene kann man noch eine Auslagerungs-Ebene mit den geringsten Kosten pro Megabyte betrachten. Aufgrund der manuellen Bereitstellung der Datenträger liegen die Zugriffszeiten im Auslagerungsarchiv erheblich über der Nearline-Ebene mit Library-Unterstützung. Die Herausforderung liegt nun darin, die Hierarchien mit Hilfe des ILM optimal auszunutzen.

4 ILM-Konzept

4.1 Grundsätzliche Arbeitsweise

ILM ist kein Produkt, sondern eine Kombination von Prozessen und Technologien. Sie bestimmen, wie Informationen durch eine Speicherinfrastruktur bestehend aus unterschiedlichen Speicherhierarchien/-klassen fließen. Betrachtet werden dabei nicht einzelne Ein-/Ausgabe-Aufträge (I/Os) sondern Informationsobjekte. Informationsobjekte sind z.B. Volumes (LUNs), Dateisysteme, Dateien, E-Mail-Nachrichten inklusive Anhänge, digitalisierte Verträge/Policen, Patientenunterlagen in Krankenhäuser. Die Prozesse bestimmen, auf welcher Speicherhierarchie/-klasse ein Informationsobjekt zu einem bestimmten Zeitpunkt liegen soll. Die Speicherhierarchie/-klasse wird dabei in einem Optimierungsprozess bestimmt. Variablen für den Optimierungsprozess sind auf der einen Seite der Wert der Information im Informationsobjekt, die Zugriffshäufigkeit, das Alter des Informationsobjekts, gesetzliche/regulatorische Anforderungen etc. und auf der anderen Seite die Kosten pro Speichereinheit, die geforderte Zugriffsgeschwindigkeit, die verfügbare Kapazität etc. in der jeweiligen Speicherhierarchie/-klasse. Der Optimierungsprozess ist ein wiederkehrender Prozess. Er wird zum ersten Mal bei der Erzeugung eines Informationsobjektes durchlaufen. Anschließend werden die Informationsobjekte in regelmäßigen Abständen neu bewertet. Falls nach der Bewertung festgestellt wird, dass ein Informationsobjekt sinnvoller in einer anderen Speicherhierarchie/-klasse gespeichert werden soll, veranlasst der Prozess die notwendigen Schritte (z.B. Verdrängung oder Verlagerung in eine andere Speicherhierarchie/-klasse, Archivierung auf Langzeitspeicher, Löschen des Informationsobjekts).

4.2 Externer Input

Der oben beschriebene Optimierungsprozess benötigt zusätzliche Entscheidungskriterien, die als externe Vorgaben (nicht aus der IT-Abteilung) eingebracht werden müssen. Die wichtigsten sind:

- Der Wert der Informationen,
- Die vereinbarten Service Level Agreements (SLA) und
- Definition der Zugriffsrechte / Datenintegrität.

4.2.1 Wert der Informationen

Die Festlegung des Werts der Informationen muss als unternehmerische Vorgabe erfolgen. Die Basis liegt in den betriebswirtschaftlichen Geschäftsprozessen mit der Abbildung der Geschäftslogik. Oft wird man feststellen, dass die Geschäftsprozesse nicht hinreichend klar definiert sind. In derartigen Fällen sollte man externe Berater einschalten. Wenn einmal der Wert der Informationen definiert ist, kann auch der Schutz der Informationen konkretisiert werden. Die wichtigsten Überlegungen sind, welche Anforderungen z.B. an RPO (Recovery Point Objectives) und RTO (Recovery Time Objectives) gestellt werden. RPO definiert den maximalen Informationsverlust, der eintreten darf, wenn ein fataler Fehler eintritt. In einer etwas unpräziseren Formulierung würde man fragen: Wie viele Transaktionen dürfen verloren gehen, wenn entweder das System abstürzt oder im schlimmsten Fall das gesamte Rechenzentrum inklusive der Daten nicht mehr zur Verfügung steht. RTO hingegen definiert die Zeit zwischen einem Schadensereignis und dem Punkt, an dem das IT-System wieder funktionsfähig sein muss. In bestimmten Branchen ist dies nicht alleine eine interne Entscheidung, sondern wird durch Verbandsvorgaben oder sogar durch Gesetze beeinflusst.

4.2.2 Service Level Agreements

Service Level Agreements (SLA) legen fest, welche Leistung z.B. ein Endanwender erwarten kann, wenn er auf Informationsobjekte zugreift. SLAs sollten grundsätzlich für jede Leistungsbereitstellung definiert werden. In einem ILM gewinnt dieser Aspekt zusätzliche Bedeutung, da dynamische Prozesse aufgrund des sich verändernden Wertes der Informationen ablaufen. Z.B. können Informationsobjekte aus dem Online-Bereich auf langsamere, kostengünstigere Speichermedien verdrängt werden, wenn darauf längere Zeit nicht mehr zugegriffen wurde. Sollte auf die verdrängten Informationsobjekte wieder zugegriffen werden müssen, dann muss klar sein, dass der Zugriff länger dauern wird, aber eine bestimmte maximale Zeit nicht überschritten werden darf.

Aufgrund der SLA-Festlegungen ergibt sich die Technologie, auf die Informationsobjekte nach den jeweiligen Zeiten verdrängt werden können.

4.2.3 Zugriffsrechte / Datenintegrität

Analog zum Wert der Daten müssen Zugriffsrechte prinzipiell aus unternehmerischer Sicht definiert werden. Beispielsweise ist zu entscheiden, ob Daten nur intern oder auch extern zugänglich sein sollen. In der Praxis existieren natürlich zuerst einmal intern unterschiedliche Benutzergruppen mit unterschiedlichen Zugriffsrechten. Diesen Vorgaben muss ILM Rechnung tragen, so dass diese Vorgaben durch die ILM-Aktionen nicht verletzt werden. Nach einer Verdrängung müssen z.B. zuerst einmal dieselben Zugriffsrechte weiter bestehen, wie vor der Verdrängung. Das gleiche gilt für Backup, Replikation, Archivierung etc. Schutzanforderungen können sich im Lebenszyklus eines Informationsobjektes verändern.

Unternehmerische Vorgaben müssen auch festlegen, für welche Datenbereiche absolute Datenintegrität gefordert wird. Datenintegrität bedeutet, dass Veränderungen in jeglicher Form ausgeschlossen werden.

4.3 Speicherhierarchien und Speicherklassen

ILM kann umso wirksamer arbeiten, je besser der gesamte Speicherbereich strukturiert ist. Dies erfolgt üblicherweise über Speicherhierarchien und Speicherklassen. Speicherhierarchien ergeben sich in der Klassifizierung von Speichertechnologien nach technischen oder nach Kostengesichtspunkten in einer absteigenden bzw. aufsteigenden Ordnung. Technische Aspekte und Kostengesichtspunkte sind oft voneinander abhängig. Z.B. sind Speichersysteme mit schnellem Zugriff in einer höheren Preisklasse angesiedelt als Speichersysteme mit langsamerem Zugriff.

Speicherklassen ergeben sich aus der logischen Strukturierung des Speicherbereichs. Somit können Speicherhierarchien noch feiner strukturiert werden. Darüber hinaus kann z.B. nach Anzahl der automatisch zu erzeugenden Kopien von Informationsobjekten differenziert werden, auch wenn die Speicherobjekte in derselben Speicherhierarchie liegen.

4.4 Strukturierung der Daten

Bisher wurden die Daten pauschal bewertet und die Container (Informationsobjekte) bestimmt. Dieses Raster muss verfeinert werden, da sonst die höchste Anforderung auch auf Daten angewendet werden müsste, die nicht diese Anforderungen haben. Mit der Strukturierung der Daten teilen wir den gesamten Datenbestand in logisch zusammenhängende Teilbereiche auf. Die klassische und wohl auch einfachste Art der Strukturierung liegt in der Strukturierung nach Aufgabengebieten (oder auch Abteilungen). Beispiele sind Forschung/Entwicklung, Einkauf, Verkauf, Marketing oder Personalabteilung. Eine Datenstrukturierung kann auch nach Benutzergruppen erfolgen. Eine orthogonale Einteilung ergibt sich durch die Trennung von „strukturierten“ und „unstrukturierten“ Daten (siehe Absatz 3 „Treibende Kräfte“). Interessanter wird das Modell, wenn die vertikalen Verfahren einzelne Datenobjekte attributieren können. Ideal wäre es, wenn z.B. ein Prozess im SMTP-Server bereits die Steuerrelevanz von

eingehenden E-mails bestimmen könnte. Damit müssten nicht alle E-Mail-Vorgänge auf Verdacht archiviert werden.

4.5 Nutzen

Ziel ist es, zu jeder Zeit ein unternehmensweites Optimum zu gewährleisten, indem Daten in unterschiedlichen Speicherklassen entsprechend ihrer Wichtigkeit bei geringsten Kosten gespeichert werden, ohne Anwender beim Zugriff auf die Daten übergebühlich zu behindern. Wesentliche Variablen für den Optimierungsprozess sind somit die Service Level Agreements (seien sie externer oder interner Art) und das vorhandene Budget bzw. vorhandene Konfigurationen.

Als Grundgedanke liegt dem Konzept zugrunde, dass sich der Wert der meisten Informationen und damit das Zugriffsmuster auf deren Informationsobjekte im Laufe der Zeit verändern. Über automatisierte Prozesse wird sichergestellt, dass die Informationsobjekte zu jeder Zeit am richtigen Ort mit den geringst möglichen Kosten gespeichert werden. Damit reduzieren sich die Investitionskosten.

Hardwarekosten sind jedoch nur ein kleiner Teil der Gesamtkosten. Unterschiedliche Total Cost of Ownership Betrachtungen gehen davon aus, dass dieser Teil nur ein Drittel bis ein Fünftel ausmacht. Die restlichen Kosten sind Infrastruktur-, Administrations- und Betriebskosten. Eine rechtzeitige Archivierung von Informationsobjekten entlastet z.B. den Online-Bereich, für den aufwändige Datensicherungsprozesse ablaufen. Archivierte- bzw. verdrängte Informationsobjekte brauchen im Prinzip nur einmal gesichert werden. Durch Archivierung kann somit das tägliche, wöchentliche oder monatliche Datensicherungsvolumen reduziert werden. Noch gravierender ist dies im Bereich der Wiederherstellung. Falls der Online-Bereich nicht von alten Informationsobjekten entlastet ist, müssen alle Objekte erst zurückgeladen werden, bevor die Anwendung wieder online gehen kann.

ILM ist die Basis zur Gewährleistung von Service Level Agreements, da im dynamischen Prozess immer für eine optimale Belegung der einzelnen Speicherklassen gesorgt wird. Service-Rechenzentren können diese Service Level Agreements direkt an ihre Kunden weitergeben. Eingehaltene Service Level Agreements wirken sich auch auf die Kundenbindung aus. Kundenanfragen, seien sie direkt über das Internet gestellt oder indirekt über den Mitarbeiter, können innerhalb kürzester Zeit beantwortet werden. Die Kundenzufriedenheit steigt entsprechend.

5 Vorleistungen

5.1 Konsolidierung und SAN/NAS Infrastruktur

ILM ist zuerst einmal konzeptionell unabhängig, ob die IT-Konfiguration in einem dezentralisierten oder zentralisierten Modell betrieben wird, solange Kommunikationspfade zu den dezentralen Standorten bestehen. Im dezentralisierten Modell ist aber der Kommunikations- und Administrationsaufwand um ein vielfaches höher. Vor der Einführung eines ILM sollte daher alle Konsolidierungspotentiale ausgeschöpft werden. Konsolidierung dürfte die Schlüsseltechnologie sein, um zum einen vernünftig zusammenhängende Speicherhierarchien zu schaffen und zum anderen die Anzahl der Speicherklassen in einem vernünftigen Umfang zu halten. Praxisbeispiele für effektive Konsolidierungen sind aus dem E-Mail-Bereich bekannt. So lassen sich mit entsprechenden Produkten mehrere hundert E-Mail-Server auf wenige Server konzentrieren. Der Anschluss einer Verdrängungs- bzw. Archivierungsebene ist bei wenigen Servern wesentlich einfacher zu realisieren als bei einigen hundert möglicherweise weit verteilten Servern.

Intelligente Speichernetzwerke, wie Storage Area Networks (SAN), Network Attached Storage (NAS) und IP Storage, erleichtern die operative Umsetzung der dynamischen ILM-Entscheidungen erheblich. In intelligenten Speichernetzwerken hat im Prinzip jeder Server auf jeden Speicherbereich Zugriff. Dies schafft z.B. die Voraussetzung, um bei einer Erstzuweisung die bestgeeignete Speicherklasse auszuwählen.

5.2 Virtualisierung

Ein effektives ILM geht von einem problemlosen Zugriff auf Ressourcen aus. Bei der Entscheidung, auf welchen Technologien Informationsobjekte abgelegt werden, sollten technische Attribute eine möglichst geringe Rolle spielen. Dies wird heute über Virtualisierung erreicht. Zu unterscheiden ist zwischen Virtualisierung im Online-, Nearline- und NAS-Bereich.

Im Online-Bereich ist zwischen Server-, Netzwerk- und Speichersystem-basierten Lösungen zu unterscheiden. Bei den Netzwerk-basierten Lösungen haben sich im wesentlichen die In-band Lösungen durchgesetzt. Diese Lösungen basieren auf Appliances, die zwischen Server und Speichersystem platziert werden. In-band Lösungen fassen auch heterogene Speichersysteme zu einem Pool zusammen und präsentieren Teilbereiche als Virtuelle LUNs. Vergleichbare Lösungen gibt es auch für den Nearline-Bereich. Auch hier wird die Trennung der logischen Sicht von der physikalischen Sicht realisiert. Im Nearline-Bereich ist die Einführung einer logischen Sicht noch wichtiger, da traditionelle Anwendungen mit den Nearline-Gerätecharakteristika konfrontiert sind. Ein hohes Maß an Virtualisierung wird z.B. auch durch NAS-Systeme erreicht. NAS-Systeme exportieren Datei-orientierte Informationsstrukturen. Die Virtualisierung erlaubt es sogar, dass unterschiedliche Plattformen auf dieselbe Datei zugreifen (beispielsweise Windows über das CIFS-Protokoll, UNIX über NFS). NAS-Systeme verbergen physikalische Speichereigenschaften in Richtung Anwendungsserver komplett.

5.3 Storage Resource Management

Im Storage Resource Management (SRM) werden die Kenndaten der vorhandenen Speichersysteme, physikalischen Platten, logischen und virtuellen Volumes, Dateisysteme, Dateien, Behälter von Datenbanken, Datenbankinhalten etc. erfasst und in einer zentralen, strukturierten Datenbank zur langfristigen Verwaltung abgelegt. Dies ist die Grundlage für eine erfolgreiche Server- und Plattform-übergreifende Datenkonsolidierung und Auswertung der Speichernutzung für Trendanalysen, Schwellwertüberwachung oder Verbrauchsabrechnungen. Aus den Daten können die Speicherklassen und deren notwendigen Größen abgeleitet werden.

6 ILM-Modell

Das ILM-Modell besteht zum einen aus den Prozessinstanzen und zum anderen aus der Übergangsmatrix. Standardprozessinstanzen sind: Erstplatzierung, Datensicherung inklusive Recovery, Replikation, Verdrängung (HSM), Archivierung, Verlagerung und Löschung von Informationsobjekten

In der **Erstplatzierung** legt ILM fest, in welcher Speicherklasse ein Informationsobjekt gespeichert werden soll. Heute erfolgt dies noch überwiegend durch eine starre Zuordnung von Dateisystem zu Virtuellen Volumes oder LUNs. Es gibt aber auch Dateisysteme, die diesen Prozess bereits automatisieren.

Die klassische **Datensicherung** basiert in der Regel auf einem Dateiversionsschema. Das heißt, die Informationsobjekte werden in regelmäßigen Abständen auf ein anderes physikalisches Medium kopiert, ohne die vorhergehende Kopie zu zerstören. Dies ermöglicht nicht nur den Zugriff auf die letzte Version eines Informationsobjekts sondern auch auf ältere Versionen. Datenbanken und Anwendungen realisieren die Wiederherstellung bis zu einem bestimmten Transaktionsstand über die Logfiles.

Replikation gibt es in der Ausprägung „kontinuierliche Spiegelung“ und „Point-in-Time“ Kopien. Beim kontinuierlichen Spiegeln wird jeder Schreibvorgang entweder synchron oder asynchron ebenfalls im gespiegelten Datenbestand ausgeführt. Daten, die in einem Datenbestand geändert werden, werden so schnell wie möglich auch im anderen Datenbestand geändert. Originaldatenbestand und Datenbestand sollten immer miteinander verbunden sein. Point-in-Time Kopien erzeugen eine komplette Kopie eines Datenbestandes in genau einem bestimmten Zeitpunkt. Auch wenn im Originaldatenbestand Änderungen durchgeführt werden, verbleibt die Kopie in dem logischen Zustand, wie sie in dem genau bestimmten Zeitpunkt erstellt wurde.

Verdrängung verschiebt ein Informationsobjekt von einer Speicherklasse in eine andere ohne dass die Zugriffssyntax geändert werden muss. Eine systemintegrierte Verdrängung wird zuerst das Informationsobjekt in die neue Speicherklasse kopieren und dann das Originalobjekt in einen Zeiger umwandeln, der auf das kopierte Objekt in der anderen Speicherklasse verweist. Dies hat den Vorteil, dass die Anwendung oder der Anwender mit derselben Zugriffssyntax zugreifen kann, als wäre die Datei noch im Online-Bereich. Der einzige Unterschied aus Endbenutzersicht ergibt sich in der Zugriffszeit, da das System bei einem Zugriff das Informationsobjekt zuerst in die originäre Speicherklasse zurückholen muss.

Verdrängung ist gegen Systeme abzugrenzen, wo ein lineares Dateisystem realisiert wird, das sich wie ein Online-Dateisystem verhält, die Informationsobjekte aber zügig auf Offline-Speicherklassen geschrieben werden. Diese Systeme kennen im Prinzip keine Policies, sondern nutzen den Online-Speicher als Cache. Eine deterministische Vorhersage, ob ein Informationsobjekt noch im Cache enthalten ist, ist nicht möglich. Aufgrund der Cache-Größe und des Zugriffsverhalten können allenfalls Wahrscheinlichkeiten angegeben werden.

Archivierung ist die gezielte Speicherung eines einzelnen Informationsobjekts oder einer zusammengehörenden Menge von Informationsobjekten in einer anderen Speicherklasse und die Definitionen, wie lange die Informationsobjekte nicht gelöscht werden dürfen. Wesentliches Merkmal von Archivierung ist die zusätzliche Speicherung von logischen Metadaten zu jedem Archivierungsauftrag. Die Metadaten können entweder manuell erstellt oder es kann eine automatische Indexierung durchgeführt werden. Die Vorgabe dieser Zeiten erfolgt basierend auf den Geschäftsprozessen. Beispielsweise gibt es im deutschen Steuerrecht neben vielen anderen Fristen eine typische Aufbewahrungsfrist von sechs oder zehn Jahren. Abhängig von der Branche gibt es auch wesentlich längere Aufbewahrungsfristen (z.B. Konstruktionszeichnungen bzw. Genehmigungsunterlagen für ein Kernkraftwerk). Die ILM-Prozesse müssen die Einhaltung dieser Fristen gewährleisten.

Eine **Verlagerung** eines Informationsobjektes ist im Prinzip ein irreversibler Vorgang. Verlagerung ist eine wichtige Prozessinstanz, wenn alte Hardware gegen neue ausgetauscht werden soll. Informationsobjekte sind schwerpunktmäßig LUNs oder Volumes. In der Verlagerung liegt die Herausforderung in der transparenten Abwicklung des Prozesses, der in der Regel längere Zeit in Anspruch nimmt und im Anschluss in der transparenten Änderung der Adressierungsmechanismen. Im Gegensatz zur Verdrängung werden im Anschluss daran an eine Verlagerung alle Bereiche auf dem Original freigegeben.

Am Ende des Lebenszyklus eines Informationsobjekts stehen die Prozessinstanzen **Löschen bzw. Zerstören**. Beim Löschen wird der Platz nur logisch freigegeben. Zerstören bedeutet, dass die Datenbereiche mehrmals mit bestimmten alternierenden Bit-Mustern überschrieben werden, so dass eine Datenrekonstruktion (auch mit ausgefeilten physikalischen Analysetools) ausgeschlossen ist.

Der dynamische Prozess wird mit einer **Übergangsmatrix** (Abbildung 1) beschrieben. Nach der Erstzuweisung sind alle Übergänge in alle anderen Zustände möglich. Von einem Informationsobjekt kann ein Backup erzeugt, ein Replikat erstellt, das Informationsobjekt kann verdrängt, verlagert, archiviert oder auch gelöscht werden. Nach dem Löschen, speziell nach dem Zerstören ist kein Übergang in einen anderen Zustand mehr möglich, wohingegen aus gesicherten oder replizierten Informationsobjekten wieder eine Erstzuweisung abgeleitet werden kann. Verdrängte oder archivierte Informationsobjekte können wieder einer Erstzuweisung zugeführt werden. Vom Backup-Zustand wird oft in den Replikatzustand übergegangen, indem nach dem Schreiben der Datensicherung auf ein Sicherungsmedium dieses anschließend automatisch repliziert wird. Darüber hinaus gibt es weitere vorstellbare Übergänge. Z.B. können Informationsobjekte aus dem Backup-Bereich heraus verdrängt oder archiviert werden. Für Informationsobjekte mit festem Inhalt ist auch eine Erstzuweisung im Archivbereich möglich. Spezielle Systeme liefern die Basis zur Erfüllung von gesetzlichen und regulatorischen Anforderungen im Archivierungsbereich.

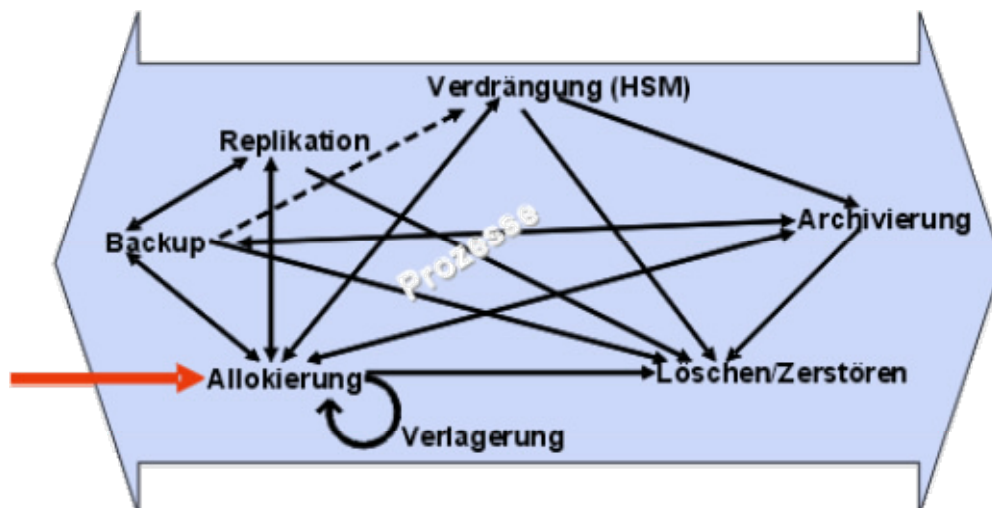


Abbildung 1: Übergangsmatrix

7 Datensicherung

Unter Anderem werden sich die Verfahren im Bereich Datensicherung bei Nutzung von ILM-Konzepten stark verändern.

Häufig werden heute undifferenzierte Policies verwendet. So findet man oft festgelegte Vorgehensweisen für alle oder nur wenige Gruppen von Daten. Beispielsweise ist es gängige Praxis, tägliche inkrementelle und wöchentliche Vollsicherungen zu erstellen. Dazu kommen Monats- und Jahres-Sicherungen mit den dazugehörigen Auslagerungsregeln.

Eine derartige Vorgehensweise führt dazu, dass einerseits erhebliche Datenmengen unnötig oft mitgesichert werden. Auf der anderen Seite werden, eben weil keine saubere Differenzierung nach Datenklassen vorgenommen wird, aktuelle Backup-Techniken wie Snapshots auf Platte, die sehr schnelle Daten-Wiederherstellungen ermöglichen, nicht eingesetzt.

Konsequenter Einsatz von ILM-Verfahren wird für Daten, die nicht mehr verändert werden, nur noch in großen Abständen Sicherungen vorsehen. Im Extremfall wird eine Sicherung gezogen und diese sicher aufbewahrt. Nach Ablauf der definierten Vorhaltefrist (Retention) sind die Daten sowieso obsolet und können gelöscht werden.

Daten, bei denen die Wahrscheinlichkeit von Updates gering ist, müssen selbstverständlich weiter gesichert werden, aber nur dann, wenn tatsächlich Veränderungen stattgefunden haben.

Für Daten, die durch Verdrängung (HSM) ausgelagert wurden, existieren bereits heute integrierte Verfahren, bei denen nur die Daten-Platzhalter, die im Vergleich zu den Originaldaten nur einen Bruchteil der Größe darstellen, gesichert werden.

Für das andere Extrem, aktuelle hochfrequentierte Datenbanken, Netzwerk-Shares und andere, stehen, wie oben bereits angedeutet, zum Beispiel Snapshot-Verfahren zur Verfügung, bei denen praktisch ohne relevante Backup-Fenster sogenannte Point-in-Time – Kopien auf Platte angefertigt werden. Diese ermöglichen, wenn erforderlich, sehr schnelle Daten-Wiederherstellungen.

8 Praktische Beispiele zu ILM

8.1 E-Mail-Archivierung

8.1.1 Anforderungen an die E-Mail-Archivierung

Ein gutes Beispiel für die Anwendung einer ILM-Strategie stellt die E-Mail-Archivierung dar. E-Mails ist seit einigen Jahren nicht mehr aus Unternehmenskommunikation wegzudenken. In vielen Fällen hat die E-Mail-Kommunikation bereits klassische Kommunikationsformen wie z.B. Brief oder Telefax ganz abgelöst, und ist heute integraler Bestandteil der internen und externen Unternehmenskommunikation.

Allgemein zählen E-Mails zum unternehmerischen Schriftverkehr. Unternehmen, die z.B. nach ISO 9000 zertifiziert sind, sind daher verpflichtet, die komplette Unternehmenskommunikation lückenlos zu dokumentieren.

Weitere Anforderung an die Archivierung von E-Mails ergeben sich aus Handels- und Steuerrecht. Kommt z.B. ein Kaufvertrag durch E-Mail zustande, müssen die entsprechenden E-Mail-Dokumente 6 Jahre lang aufbewahrt werden.

Im unternehmerischen Alltag stellt sich zumeist noch eine weitere andere Anforderung: Nicht alle gesendeten oder empfangenen E-Mails haben rein geschäftlichen Hintergrund. Die teilweise private Nutzung eines E-Mail-Systems ist in vielen Unternehmen gängige Praxis. Für den privaten Mailverkehr greifen grundsätzlich die Vorschriften des Fernmeldegeheimnisses.

8.1.2 Elemente einer ILM-Strategie bei der E-Mail-Archivierung

Wird eine ILM-Strategie bei E-Mails angewandt, so bedeutet dies deutlich mehr als die reine Archivierung von E-Mails. Bei einem klassischen Archivierungsvorgang werden alle E-Mails als ein großer Pool an Daten betrachtet, der Inhalt wird lediglich für die Indizierung betrachtet. Es erfolgt keine zeitliche oder inhaltliche Bewertung.

Im Rahmen einer ILM-Strategie müssen alle E-Mails inhaltlich betrachtet und bewertet werden. Aufgrund dieser Bewertung wird dann definiert ob diese E-Mail z.B. archiviert werden muss, bzw. für wie lange diese E-Mail im Archiv sein muss. Besondere Bedeutung entfällt hier die Trennung zwischen geschäftlichen und privaten E-Mails.

8.1.3 Implementierung einer ILM-Strategie bei der E-Mail-Archivierung

Im E-Mail-Bereich sind vor allem folgende Punkte bei der Implementierung einer ILM-Strategie zu beachten:

- Ist eine klare Richtlinie für den Umgang mit privaten E-Mails getroffen?
- Welche gesetzlichen Vorschriften und sonstige Normen treffen auf das Unternehmen zu?
- Mit welchen Maßnahmen erfolgt die Inhaltliche Bewertung von E-Mails?
- Mit welchen Maßnahmen werden die Inhalte den ILM Richtlinien zugeordnet?
- Durch welche Maßnahmen wird die Unveränderbarkeit der Inhalte gewährleistet?
- Durch welche Maßnahmen wird die Lesbarkeit der Inhalte für den Aufbewahrungszeitraum gewährleistet?
- Durch welche Maßnahmen wird die zeitnahe Bereitstellung der Inhalte gewährleistet?
- Durch welche Maßnahmen ist ein Lesezugriff für externe Audits möglich?
- Wie erfolgt die Löschung von Informationen?
- Wie wird die Einhaltung von Unternehmensrichtlinien gewährleistet?
- Wie werden die Hard- und Software-Komponenten, z.B. Speichermedien, gegen Bedrohungen, Bedienungsfehler und technische Defekte geschützt?

Die Implementierung einer ILM Strategie umfasst neben den technischen Maßnahmen vor allem Eingriffe in die Organisationsrichtlinien. Der Erfolg des ILM-Prozesses wird maßgeblich durch die Umsetzung dieser Richtlinien getragen.

So kann z.B. durch gezielte organisatorische Maßnahmen der E-Mail-Verkehr kanalisiert werden. Dadurch werden sehr aufwändige technische Maßnahmen zur Inhaltserkennung stark vereinfacht oder ganz überflüssig. Für eingehende Bestellungen per E-Mail kann z.B. eine dedizierte Mailbox eingerichtet werden. Ein anderes Beispiel ist die Einrichtung einer „Gruppen-Mailbox“ für eine Abteilung. Durch diese Maßnahmen dient dann der Mailboxname als Indikator für nachfolgende ILM-Prozesse. Ist dies nicht möglich, müssen z.B. alle Mail indiziert werden und über Key-Wörter etc. zugeordnet werden.

Am Beispiel E-Mail wird sehr gut deutlich, dass ILM keine einzelne Lösung darstellt, sondern eine Kombination von Prozessen und Technologien.

Bereits bestehende Lösungen wie z.B. eine Backup/Recovery-Lösung für den Mailserver stellen Teilkomponenten einer ILM-Strategie dar, und können lückenlos in eine einheitliche ILM-Strategie integriert werden.

8.2 Lebensversicherung

Am weiteren Beispiel der Daten zu einer Lebensversicherung lässt sich die Wirkungsweise einer ILM Strategie ebenfalls gut verdeutlichen. Bei Abschluss einer Lebensversicherung werden die vom Kunden gemachten Angaben zu Gesundheitszustand, Krankheiten sowie der gesamte Schriftverkehr zwischen Kunden und Versicherungsunternehmen in elektronischer Form gespeichert und in einem Archivsystem elektronisch abgelegt. Vor und kurz nach Vertragsabschluss gibt es üblicherweise noch Rückfragen zwischen Antragsteller und Versicherungsunternehmen, auf die entsprechenden Dokumente wird häufiger zugegriffen. Daher werden die Dokumente auf einem Speichersystem mit kurzen Zugriffszeiten abgelegt. Nach dieser Phase wird sehr lange – bis zum Eintreten des Leistungsfalls – nicht mehr auf diese Dokumente zugegriffen, sie werden daher automatisch auf günstigere Datenträger (z.B. Magnetband) migriert. Im Leistungsfall wird nun bis zur erfolgten Abwicklung wieder häufiger auf diese Dokumente zugegriffen, sie sollten daher nun automatisiert wieder auf Datenspeicher mit kürzeren Zugriffszeiten verlagert werden. Nach erfolgter Abwicklung sollen die entsprechenden Unterlagen noch eine bestimmte Frist (z.B. 10 Jahre) aufbewahrt werden. Den Dokumenten wird daher jetzt eine andere Klasse – nämlich genau diejenige, die diese Anforderungen reflektiert – zugeordnet.

Die automatische Berücksichtigung der jeweils besten Speichertechnologie, die automatische Berücksichtigung der Aufbewahrungsvorschriften, sowie letztendlich die Vernichtung, d.h. das Löschen der Dokumente nach Ablauf der Aufbewahrungsfrist sind wesentliche Bestandteile des Information Lifecycle Managements.

9 ILM und System Managed Storage

Die oben angeführten Beispiele zeigen deutlich, dass es wichtig ist, die abgespeicherten Dokumente zu klassifizieren und zu kategorisieren, d.h. in unterschiedliche Gruppen einzuordnen. Die Menge der heutzutage anfallenden Informationen erlaubt es einfach nicht, jedem Objekt einzelne, individuelle Attribute zuzuordnen. Stattdessen werden diese Attribute nur für die jeweiligen Klassen vergeben. Sie sind damit automatisch für alle Daten in der jeweiligen Klasse gültig. Die einzelnen Daten/Objekte werden über entsprechende Auswahlroutinen den Klassen zugeordnet. Diese Vorgehensweise ist nicht neu, sondern wurde bereits vor vielen Jahren in der Großrechnerumgebung unter dem Begriff SMS (System Managed Storage) eingeführt.

SMS basiert auf unterschiedlichen Klassen, beispielsweise:

1. in der Data-Class werden Standardattribute für die Erstellung von Dateien zugewiesen (Platzbedarf, Dateioorganisation etc.)
2. in der Storage-Class werden die Anforderungen bzgl. Zugriffszeit (Performance) und Verfügbarkeit festgelegt.
3. in der Management-Class werden die Aspekte der Datensicherung, der Migration und der Aufbewahrungsfristen definiert.

Jeder Datei wird auf der Basis von definierten Auswahlregeln die entsprechende Klasse zugewiesen. Die Klassennamen und Attribute werden auch in dem Katalog (d.h. dem Inhaltsverzeichnis aller Dateien) mit abgelegt, stehen also während der gesamten Lebensdauer der Datei zur Verfügung. Die Auswahlregeln werden zentral vom Speicheradministrator definiert. Für die Zuweisung der „passenden“ Klassen stehen dabei vielfältige Parameter zur Verfügung (Dateiname; Name des Benutzers, der die Datei anlegt; etc.). Die gewünschten Klassen können auch beim Anlegen der Datei mit angegeben werden; in den Auswahlregeln wird dann entschieden, inwieweit diese berücksichtigt oder verworfen werden sollen. Des Weiteren sorgt die Integration in das Sicherheits-System des Großrechners dafür, dass die jeweiligen Klassen nur von den dafür autorisierten Benutzern benutzt werden können.

Die im Rechenzentrum installierte Speicherperipherie wird in einer SMS-Umgebung vom Systembetreiber in verschiedene Gruppen (Storage-Groups) eingeteilt. In den Auswahlroutinen wird aus den zugewiesenen Klassen die passende Storage-Group ermittelt und zugewiesen.

Die Einführung und Nutzung von SMS hat in der Großrechnerumgebung dazu geführt, dass trotz steigenden Datenaufkommens die Zahl der Speicher-Administratoren nicht gestiegen ist. Statt wie früher Attribute einzelner Dateien zu administrieren, werden heute lediglich die entsprechenden Klasseninformationen angepasst.

Soll zum Beispiel die Anzahl der aufzubewahrenden Datensicherungen für alle Vertriebsanwendungen geändert werden, so ist lediglich der Parameter in der entsprechenden Management-Klasse anzupassen.

Anwender, die Großrechnersysteme betreiben, wünschen sich diese massiven Vorteile auch für die Umgebung der „offenen Systeme“ (UNIX, Windows etc.).

Das oben beschriebene „policy-based Data-Management“ stellt damit einen wesentlichen Bauteil des ILM dar.

10 Schlussbetrachtung

ILM ist noch eine relative junge Disziplin. Es gibt zwar viele Teil-Lösungen, aber eine integrierte Gesamtlösung wird sicherlich noch einige Jahre auf sich warten lassen. Eine Kernkomponente kann ein zentraler Informationsobjekt-Katalog sein. Dieser Katalog gibt zum einen Auskunft über die originalen Informationsobjekte und verknüpft zum anderen alle davon abgeleiteten Informationsobjekte wie z.B. Sicherungsobjekte, Replikate oder auch Archivierungsobjekte. Eine wesentlich effizientere Arbeitsweise wird ermöglicht, wenn das System aus der Interpretation von Zugriffsmustern und Benutzerverhalten standardisierte oder initial definierte Policies dynamisch anpasst. Möglicherweise bedarf es zur Umsetzung dieser Vision völlig neuer Systemarchitekturen, die zum einen die Schnittstelle zu den Geschäftsprozessen darstellen, indem sie entsprechende logische Attribute verwalten können und zum anderen Plattform-übergreifend und völlig Hardware-neutral arbeiten.

Einzelne Lösungen zeigen, dass die Industrie den Weg in Richtung Gesamtlösung bereits eingeschlagen hat. Mit Teillösungen wird man vermutlich einen hohen Prozentsatz der heutigen Schmerzpunkte lösen können. Es gibt somit wenig Argumente, mit ILM nicht sofort zu beginnen. Dabei ist die Umsetzung von Vorgaben aus den Geschäftsprozessen der eigentliche Kernpunkt. Hier ist noch erhebliche Integrationsleistung erforderlich. Es ist oft schwierig, zum einen alle gesetzlichen und regulatorischen Anforderungen zu kennen und auf der anderen Seite den Wert der Informationsobjekte korrekt einzuschätzen. Dies ist umso schwieriger, wenn Gruppen- oder Abteilungsinteressen innerhalb eines Unternehmens divergieren. Hier hilft nur eine neutrale Beratung von externen Sachverständigen.