

Datenmigration

Dateivirtualisierung und Datenklassifikation

Wiederbelebung des ILM-Konzepts

30.08.2007 | Autor: Rainer Graefen

Die Antwort der Speicherindustrie auf die wachsenden Datenmengen, hieß vor etwa vier Jahren ILM, Information Lifecycle Management. Um ILM ist es scheinbar ruhig geworden. Kombiniert man aber die neuen „Hypes“ Dateivirtualisierung und Datenklassifikation miteinander, so finden sich hier alle guten Ideen des ILM-Konzeptes in neuem und auch praktikablem Gewande wieder.



Immer mehr Informationen werden nur noch in digitaler Form um die Welt geschickt. Jedes Unternehmen ist an diesem Kreislauf gleichzeitig als Produzent und als Verbraucher beteiligt. Speichertechnisch betrachtet kein größeres Problem. Das Unternehmen Hitachi Global Storage Technologies produziert schon heute Festplatten mit einem Terabyte Kapazität, Samsung und Seagate werden in wenigen Monaten folgen und Hitachi Data Systems kann allein mit seinem neuesten Speichersystem, der Universal Storage Platform V, insgesamt einen Speicherraum mit bis zu 250 Petabyte Daten adressieren.

Selbst tief gestaffelte Ordnerstrukturen wären vollständig unzureichend, um diese Datenmassen in eine beherrschbare Struktur zu zwingen. Das merkt jeder Anwender schon an seiner lokalen Festplatte. Innerhalb weniger Wochen, kann sich fast niemand mehr erinnern, wo er selten genutzte Dateien abgelegt hat. Im Internet hat man das Wiederfinden von Informationen mit leistungsstarken Suchmaschinen à la Google gelöst. Nicht vollständig befriedigend, aber man kann damit leben – bislang.

Das ILM-Konzept

Während sich auf der logischen Ebene Werkzeuge zur Wiederbeschaffung verloren gegangener Dateien finden, stehen IT-Leiter dem steigenden Datenaufkommen und dessen systemtechnischer Beherrschung bislang ziemlich machtlos gegenüber. Die Speicherindustrie bot vor etwa vier Jahren mit ILM, dem Information Lifecycle Management, ein Konzept an, das einen Lösungsweg versprach. Das Ganze war sehr technisch inspiriert und ließ sich kaum von HSM, dem Hierarchischen Speichermanagement, unterscheiden.

Mit letzterem lassen sich Daten auf den Primärspeichern nach äußerlichen Kriterien, also Metadaten wie Erstellungsdatum, letzte Benutzung oder Füllungsgrad des Speichers, auf preiswertere Medien verschieben. Zurück bleibt dann ein sogenannter Stub, ein Null-Byte-File identischen Namens, das den neuen Lagerort beinhaltet und dem Anwender den gewohnten, leicht verzögerten Zugriff auf die Datei erlaubt. ILM wollte diese kapazitätsorientierte Verschiebung durch eine inhaltsgesteuerte Verschiebung ersetzen: Die Daten sollten entsprechend ihrem Wert in einer Speicherhierarchie verschoben werden.

Zu viel Handarbeit

Das ganze Konzept litt daran, dass erst nach einer sorgfältigen Datenklassifikation eine Automation hätte eingeführt werden können, man also mehrere Jahre externe Mitarbeiter im Hause gehabt hätte. Das ganze Projekt hätte eine intensive Zusammenarbeit mit den Unternehmensabteilungen inklusive der Steuerung durch die Geschäftsführung benötigt.

Heute hat sich die Herausforderung, die zu erwartenden Datenfluten in den Griff zu bekommen, in zwei Aufgabenbereiche zerlegt:

- 1. Die Erzeugung eines globalen Dateiraumes. Die technischen Fortschritte, alle Dateisysteme im Unternehmen unter eine Haube - sprich Dateibaum - zu bringen, sind unverkennbar. Durch diese Dateivirtualisierung ist es möglich, Daten an beliebige Orte zu verschieben, ohne dass der Anwender davon etwas mitbekommt.
- 2. Die inhaltsgesteuerte Sortierung unstrukturierter Dateien. Nachdem alle Dateien

verschlagwortet sind, können diese nach definierbaren Regeln sortiert und auf die passenden – preiswerte oder unveränderbare – Speichersysteme migriert werden. So wären auch Compliance-Richtlinien umsetzbar.

Unter organisatorischen Gesichtspunkten passen Punkt Eins und Punkt Zwei nicht zusammen. Während die Speichersysteme im Rechenzentrum möglichst produktiv zu verwalten sind, gehört die Definition von Sortierregeln (Policies) immer noch auf die Managementebene.

Dateivirtualisierung durch Switch

Aus Amerika kommend versucht nun die Firma Acopia in Deutschland Fuß zu fassen. Mittels eines Switches umgeht der Hersteller das Problem vieler globaler Dateisysteme, die typischerweise Agenten auf dem Server und manchmal auch auf dem Client erfordern. Im Rechenzentrumsbetrieb mit mehreren Hunderten oder gar Tausenden von Clients wird die Wartung solcher Agenten ungern praktiziert. Acopias ARX-Switch beschränkt seine Tätigkeit auf unstrukturierte Dateien, die über die drei Netzwerkprotokolle CIFS (Common Internet File System) für Windows-, NFS (Network File System) für Unix-Umgebungen und das Apple Filing Protocol (AFP) auf gemeinsame Netzwerkordner (Shares) zugreifen können.

Der Dateibaum im Switch

Acopia erweitert mit seinem Konzept auf einfache Weise das Dateisystem zu einem häufig als Global Namespace bezeichneten Dateibaum. Im Cache des Switches befinden sich - vereinfacht dargestellt - die Inhaltsverzeichnisse der diversen Dateisysteme sowie Einträge, wohin eine Datei durch den Administrator migriert wurde. Eine Sicherheitskopie wird auf einem beliebigen Datei-Share abgelegt, sodass ein Austausch ohne große Mühen stattfinden kann. Nach dem LAN (Netzwerk) und dem SAN (Speichernetz) wird damit auch noch ein FAN (File Area Networking) notwendig. Das bedeutet, im Zugriffspfad der Anwendung auf den Speicher befindet sich nun noch eine zusätzliche Netzwerkschicht, die zwischen Dateisystem- und Speicherschicht liegt.

Weg mit der Universal Name Convention

Die Einführung der zusätzlichen Netzwerkschicht verbirgt die Speichersysteme vor dem Anwender. Selbst absolute Pfade auf dem Client nach der UNC-Konvention \\server\share sind virtualisiert. Der ARX-Switch kann diesen Zeiger auf ein neues Ziel „umbiegen“. Durch die Dateivirtualisierung erhält der Administrator ein neues Mittel, Kategorien zu erstellen und die Dokumente nach Alter, Größe oder Art (MP3, Doc) in einem ersten Schritt auf verschiedenen Speichersystemen zusammenzufassen.

Beliebtes Beispiel der Hersteller sind ältere Dateien auf dem Primärspeicher, die nicht nur teuren Speicherplatz belegen, sondern diesen auch noch durch die hohe Auslastung langsamer machen als nötig. Mittels Schwellwerten (Watermarks) kann der ARX-Switch auch Regeln umsetzen, die z.B. dafür sorgen, dass nie mehr als fünf Millionen Dateien in einem Share liegen.

Wenn sich alle Dateien transparent, also ohne dass der Anwender etwas davon mitbekommt, verschieben lassen, dann lässt sich natürlich auch ein älteres Speichersystem leer räumen und schlussendlich gegen ein anderes austauschen.

Wissen, was im Unternehmen läuft

In einem ILM-Konzept würde der Acopia-Switch die Rolle des Datamovers anhand der oben beschriebenen Dateiattribute übernehmen. Was noch fehlt ist die inhaltliche Analyse der Dokumente, die zur Bewertung „wichtig, vertraulich, Wissensbasis und unwichtig“ führen könnte. Acopia arbeitet in diesem Bereich intensiv mit Kazeon zusammen, die mit ihrer Appliance die Unternehmenssuchmaschine in das ILM-Konzept einbringen. Ein Sachbearbeiter, der eine Schnittstelle zur Kazeon-Appliance hat, kann in diesem Index suchen, vorselektieren, eingrenzen und die Information letztlich auf seinen Desktop holen.

Google ist hierzu nicht geeignet, da der Internet-Suchmaschine mit Sicherheit kein Zugriff auf Archivsysteme, WORM-Tapes, E-Mail-Archive oder CAS-Speicher wie EMCs Centera gewährt wird. Außerdem beherrscht die Kazeon-Appliance zurzeit 340 Dateisysteme und ist darauf optimiert, Kreditkarteninformationen und Auftragsnummern also Muster herauszufiltern. Auf dieser Ebene wird die Business-Produktivität erhöht.

Einer Einschränkung der Kazeon-Appliance muss man sich aber bewusst sein: Firmen, die 100 Millionen Dateien indizieren wollen, müssen mit einem Zeitaufwand von über 1.000 Stunden rechnen.

Administrationsaufgaben trennen

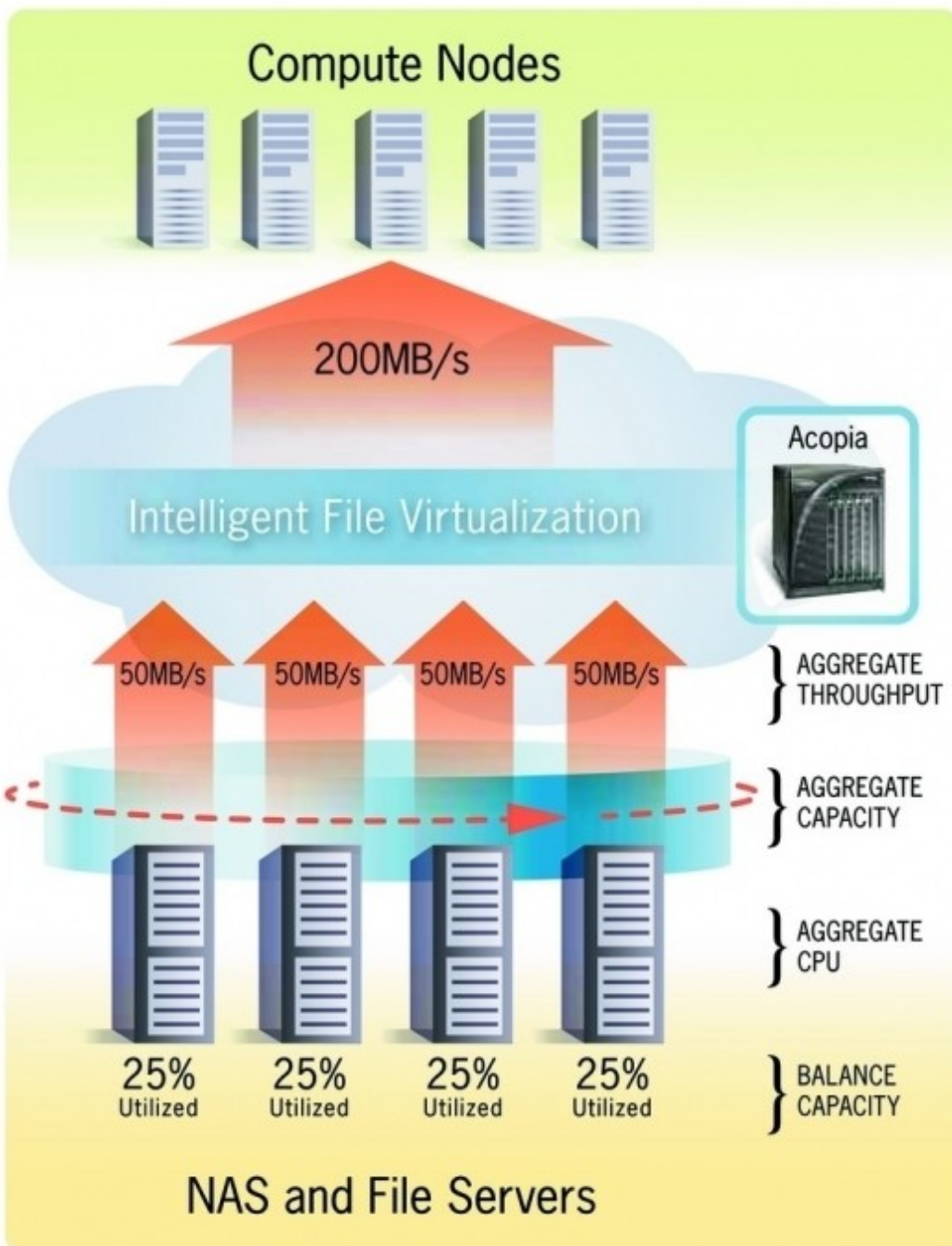
1. Applikationsschicht: ILM-Lösungen schreiben ihre Daten in das Dateisystem, das über den Global Namespace zur Verfügung gestellt wird.
2. Dateisystemschiicht: Die Dateisysteme CIFS und NFS bilden den Global Namespace. Benutzer und Anwendungen legen hier ihre Daten ab.
3. NEU: Netzwerkschicht: Acopias ARX-Switch präsentiert den Global Namespace zum Server und den Arbeitsrechnern. Administratoren können hier die diversen Anwendungen individuell betreuen und Migrationsregeln zu den Speicherklassen und zum CAS-Speicher definieren.
4. Speicherschicht: Die komplette Speicherhierarchie von den primären NAS-Fileern über Content Adressable Storage (CAS) bis hin zum Bandroboter befindet sich hier. Mit Hilfe der Dateivirtualisierung lässt sich diese Infrastruktur in ein gestuftes Speichermodell (Tiered Storage) verwandeln.

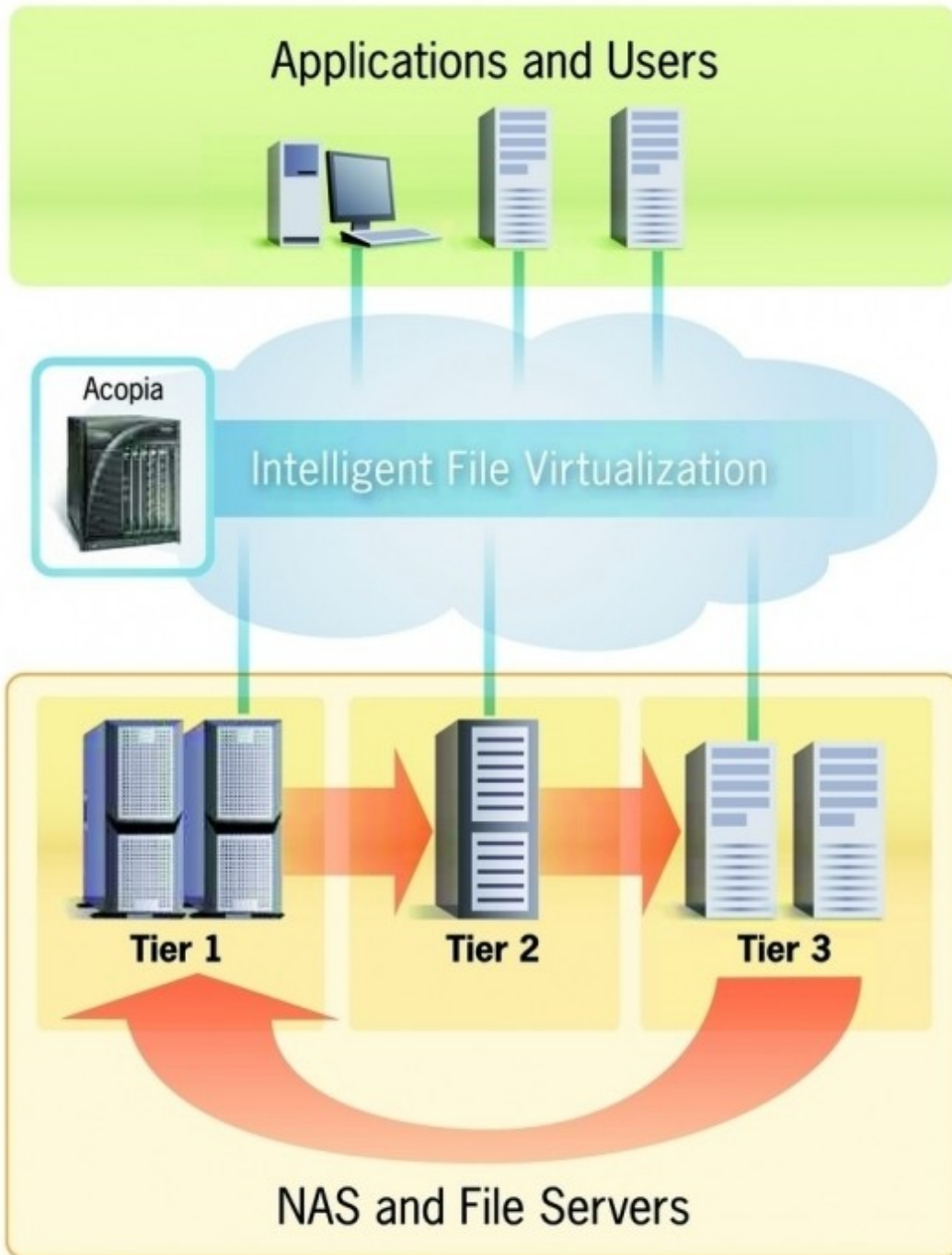
Redakteur: Rainer Graefen

Die Beiträge auf dieser Website sind urheberrechtlich geschützt. Bei Fragen zu den Nutzungsrechten wenden Sie sich bitte an manuela_maurer@vogel-medien.de oder Tel.: 0931-418-2888.

Bildergalerie







Dieses PDF wurde Ihnen bereitgestellt von <http://www.searchstorage.de>