

Datenschutzplattform mit einfacher, zuverlässiger und schneller Deduplizierung

White Paper von ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) im Auftrag von IBM Deutschland

Oktober 2008

Inhalt

Kurzübersicht	1
Einführung	1
Argumente für VTLs	1
Datendeduplizierung	2
Warum Cluster-Gateways?	3
Das IBM System Storage TS7650G ProtecTIER De-duplication Gateway	4
Die EMA-Perspektive	5
Informationen über Diligent, ein IBM Unternehmen	8
Informationen über Diligent Technologies:	8
Informationen über IBM:	8

Kurzübersicht

Virtual Tape Libraries (VTLs) sind mittlerweile ein grundlegender Bestandteil der Infrastruktur von Rechenzentren. Die Zunahme der zu verwaltenden Datenbestände sorgt jedoch dafür, dass die meisten Unternehmen Schwierigkeiten haben, ihre VTLs mit Plattenkapazität zu erweitern. Dies führte zu einer großen Nachfrage nach der Deduplizierung von Daten. Bei der Datendeduplizierung handelt es sich um eine Technologie, die in einem Plattenrepository redundante Daten findet und löscht. Das Ergebnis ist ein erheblicher Anstieg der verfügbaren Kapazität von Plattenpools. Wenn diese leistungsfähige Technologie für VTLs verwendet wird, können Unternehmen deutlich mehr Daten als bisher speichern und aufbewahren und gleichzeitig die Kosten für Plattenspeicher erheblich senken.

Damit eine VTL-Lösung den Anforderungen eines Unternehmensrechenzentrums gerecht wird, muss sie jedoch mehr leisten als nur den Speicherbedarf reduzieren. Sie muss beispielsweise einen gleichbleibend hohen Durchsatz bieten und den Anforderungen des Unternehmens hinsichtlich Verfügbarkeit gerecht werden.

Das neue TS7650G De-duplication Gateway von IBM ist ein Beispiel für eine VTL mit Funktionen für die Deduplizierung, die die Anforderungen eines Unternehmensrechenzentrums erfüllt. Dies ist die erste Lösung, die ein Cluster mit zwei Knoten unterstützt und gleichzeitig leistungsfähige Inline-Deduplizierung bietet.

Einführung

Die Bedenken im Zusammenhang mit der Einhaltung gesetzlicher Vorschriften, eDiscovery und der wachsenden Datenflut verstärken zwar den Druck, mit dem IT-Manager leben müssen, jedoch hat sich ihr Job in einer wichtigen Hinsicht in den letzten 30 Jahren im Grunde kaum verändert: Im Endeffekt liegt der Schlüssel zum Erfolg eines IT-Teams in der Frage „Sind die Unternehmensdaten im notwendigen Umfang geschützt und ist die Business-Continuity unter allen Umständen sichergestellt?“.

Der Schutz geschäftskritischer Daten bleibt oberste Priorität, unabhängig davon, ob ein Rechenzentrum seine geschäftskritischen Informationen in Gigabyte oder Petabyte misst. Das Schritthalten mit neuen Anforderungen und das gleichzeitige Bewältigen extrem umfangreicher Datenmengen stellen eine große Herausforderung dar. Neue Technologien für Sicherung und Wiederherstellung können kaum mit den Anforderungen für schnellen Schutz und schnelle Wiederherstellung von Daten mithalten. Langfristige Archive müssen jetzt den Anforderungen im Zusammenhang mit der Einhaltung gesetzlicher Vorschriften, den Verordnungen für unternehmensweite Governance, rechtlichen Offenlegungspflichten und den Anforderungen vieler Unternehmensbereiche für den schnellen Zugriff auf Langzeitdaten zur Unterstützung komplexer Geschäftsanalysen gerecht werden.

In diesem Dokument wird erörtert, wie eine Virtual Tape Library (VTL), die mit einem Cluster-Gateway und Funktionen für die Datendeduplizierung arbeitet, Rechenzentren mit diesen Anforderungen unterstützen kann und wie Manager von Rechenzentren mit diesen Einheiten problemlos Sicherungs- und Wiederherstellungsprozesse beschleunigen, die Nutzung bestehender Hardware effizienter gestalten und das Management von Datenrepositories selbst in Petabytegröße vereinfachen können.

Argumente für VTLs

Virtual Tape Libraries werden verwendet wie physische Bandeinheiten. Die VTL wird als Zwischenspeicherbereich zwischen den Platteneinheiten der ersten Ebene und den Bandarchiven eingesetzt. So erscheint sie für die Software für Sicherung und

Wiederherstellung als physisches Bandarchiv. Dadurch findet ein Paradigmenwechsel bei Sicherungen von Platte auf Band (Disk-to-Tape, D2T) in Platte auf Platte auf Band (Disk-to-Disk-to-Tape, D2D2T) statt, der einige interessante Folgen hat.

Alle VTLs führen dieselben grundlegenden Funktionen aus, dienen als plattenbasiertes Ziel für den Sicherungsprozess und bieten schnelle Wiederherstellung von lokalem, plattenbasiertem Speicher. Trotzdem sind nicht alle VTLs gleich. Benutzer von VTLs der ersten Generation, die mit dem Sicherungskonzept Platte auf Platte (Disk-to-Disk, D2D) arbeiten, haben festgestellt, dass die Plattenkapazität bei steigenden Datenschutzerfordernissen sehr schnell verbraucht war. Die Möglichkeit bei VTLs der zweiten Generation zur Migration älterer Daten auf Band (das D2D2T-Modell) ist von einem wirtschaftlichen Standpunkt aus sinnvoller, da Manager ihre eigenen Richtlinien für einen Austausch zwischen Platte und Band festlegen können. Darüber hinaus bieten viele VTLs Nutzen durch zusätzliche Funktionen, wie z. B. Verschlüsselung, Komprimierung und Datenduplizierung. Diese Funktionen sind zwar alle recht nützlich, doch bietet die Deduplizierung das beste Beispiel für eine schnelle Amortisation der Investitionen in VTLs.

Benutzer von VTLs können von Folgendem ausgehen:

- Da weniger Wartungen erforderlich sind, haben Produktionssysteme längere Betriebszeiten.
- Für Sicherungen ist eine hohe Prozessorleistung erforderlich. Da aber die Sicherungen auf Band in der VTL ablaufen, wird die Leistungsfähigkeit der Produktionssystemen kaum reduziert.
- Da Wiederherstellungen erheblich schneller sind, werden die Ausfallzeiten erheblich reduziert.
- Da für inkrementelle Sicherungen kein Bandspeicher mehr notwendig ist, sinken unmittelbar die Kosten von IT-Abteilungen im Zusammenhang mit Bandspeicher und -bearbeitung.
- Da das System für die Software für Sicherung und Wiederherstellung unverändert erscheint, können IT-Abteilungen weiterhin dieselben Sicherungsskripts und richtlinienbasierten Managementverfahren nutzen, die vor Einführung der VTLs verwendet wurden. Dies ist möglicherweise der wichtigste Vorteil.
- Da die Betriebszeiten für Geschäftsprozesse länger werden, können Service-Level-Ziele (RPOs und RTOs) einfacher erreicht werden, wodurch die Zufriedenheit aller Beteiligten steigen sollte.

Datenduplizierung

Der Begriff „Datenduplizierung“ bezieht sich auf eine Reihe von Technologien, die darauf abzielen, den insgesamt zu verwaltenden Datenbestand zu reduzieren und gleichzeitig die Integrität, Verfügbarkeit und Zugänglichkeit dieses Datenbestands sicherzustellen. Durch die Reduzierung des zu verwaltenden Datenbestands können Rechenzentren die dafür erforderliche Hardware reduzieren und damit die Betriebskosten für Strom und Kühlung senken, die Kosten für Bandspeicher reduzieren und es Speicheradministratoren ermöglichen, mit einer extremen Zunahme von Rohdaten und komplexeren Speicherinfrastrukturen zurechtzukommen.

Die verschiedenen Anbieter implementieren die Deduplizierung auf verschiedene Art und Weise. Die wichtigsten Aspekte lauten wie folgt:

- **Leistung.** Unabhängig davon, wie effektiv eine Technologie bei der Reduzierung der Kapazitätsanforderungen ist, sie ist nur dann erfolgreich, wenn Daten im vorgegebenen Zeitraum verarbeitet werden (und dadurch verschiedene Service-Level-Agreements für die Fertigstellung von Sicherungen und Offsitedatenschutz eingehalten werden). Bei der Analyse von Lösungen für die Deduplizierung ist die Leistung immer ein Hauptkriterium.

Durch die Reduzierung des zu verwaltenden Datenbestands können Rechenzentren die dafür erforderliche Hardware reduzieren.

- **Granularität.** Deduplizierung kann auf Datei-, Block- oder Byteebene durchgeführt werden. Je differenzierter die Analyse, desto effizienter der Prozess.
- **Factoring** (Identifizierung redundanter Daten). Deduplizierung nutzt proprietäre Algorithmen, die entweder auf Hashing (wobei zuerst ein „Fingerabdruck“ für jeden einzelnen Datenblock erstellt wird und diese dann miteinander verglichen werden) oder auf dem Vergleich tatsächlicher Daten basieren. Factoring ist die wichtigste Bestimmungsgröße für die Geschwindigkeit der Deduplizierung.
- **Architektur.** Architekturen mit Funktionen für die Deduplizierung können in drei Kategorien eingeteilt werden.
 - Bei der *Parallelverarbeitung* werden Rohdaten auf Platte geschrieben, diese Daten werden dedupliziert und gleichzeitig erneut geschrieben. Die parallele Ausführung dieser drei Operationen erfordert sowohl zusätzliche Verarbeitungsleistung als auch zusätzlichen Plattenspeicherplatz.
 - Bei der *Nachbearbeitung* werden Rohdaten zuerst auf Platte geschrieben, diese Daten werden dedupliziert und dann erneut geschrieben; die Rohdaten werden schließlich gelöscht. Für dieses Verfahren mit mehreren Schritten ist weniger Verarbeitungsleistung erforderlich, da das Factoring verzögert wird, es ist jedoch zusätzlicher Plattenspeicherplatz notwendig und das Verfahren dauert länger. Dies kann Sicherungsintervalle gefährden.
 - Bei der *Inlineverarbeitung* werden die Daten dedupliziert, bevor sie auf Platte geschrieben werden. Dafür ist die höchste Verarbeitungsleistung notwendig, jedoch werden Plattenspeicherplatz und Sicherungsintervalle optimal genutzt.
- **Wo findet dieser Prozess statt?** Deduplizierung kann auf Hosts oder Virtual Tape Librarys durchgeführt werden. Zur Durchführung auf einem Host ist eine Rekonfiguration der Hostumgebung erforderlich, was häufig umfangreiche Upgrades und neue Scripts sowie Tests vorhandener Sicherungsprozeduren nach sich zieht. Bei der Durchführung auf einer VTL ist die Deduplizierung für bestehende Prozesse jedoch transparent und die neue Funktionalität lässt die bestehenden Best Practices unberührt.

Die zwei wichtigsten Messgrößen zur Beschreibung der Deduplizierung sind *Durchsatz* und *Factoringrate*. Der Durchsatz ist eine Kennzahl für die Zeit, die zum Schreiben der deduplizierten Daten in den Speicher der VTL benötigt wird. Die Factoringrate (oder Deduplizierungsrate) ergibt sich durch die Division der Gesamtdatenmenge vor der Reduzierung durch die Gesamtdatenmenge nach der Reduzierung. Es ist wichtig, zu wissen, wie die einzelnen Anbieter diese Zahlen verwenden, da einige Zahlen trügen können.

Warum Cluster-Gateways?

Ein Speichergateway verbindet Backup-Arrays auf der einen Seite des Gateways mit Backup-Servern auf der anderen Seite (vgl. Abbildung 1). Ein wichtiger Aspekt des Nutzenpotenzials dieser Einheiten ist folgender: Da ein Gateway selbst keinen Speicher hat, können Speichermanager verschiedene Kombinationen von Einheiten verschiedener Anbieter und verschiedener Art auf einer oder beiden Seiten des Gateways zusammenstellen. Da gatewaybasierter Speicher nicht proprietär ist, können Manager *Geschäftsbeziehungen mit ihren derzeitigen Anbietern beibehalten* und *weiterhin die bestehende Speicherhardware nutzen*. Wenn dann horizontales (oder vertikales) Skalieren mit neuen Arrays ansteht, haben Manager normalerweise die Wahl, ob sie weiterhin mit den

derzeitigen Anbietern zusammenarbeiten oder eine Geschäftsbeziehung mit anderen Anbietern eingehen möchten. Ganz gleich, worauf die Wahl fällt, mit Gateways kann bei Bedarf problemlos zusätzlicher Speicher hinzugefügt werden.

Ein Cluster-Gateway trägt zu hoher Verfügbarkeit bei.

Unternehmen schätzen Cluster-Gateways aus denselben Gründen, aus denen sie auch Cluster-Anwendungsserver schätzen. Durch Clustering kann die *Leistung* verbessert werden – an vielen Standorten das Kaufargument schlechthin. Ein Cluster-Gateway trägt zu *hoher Verfügbarkeit* bei, denn sollte ein Knoten im Cluster ausfallen, stellt die Funktionsübernahme durch den zweiten Knoten sicher, dass es keine Betriebsunterbrechung gibt, bis der ausgefallene Knoten wieder aktiviert werden kann. Wie auch Cluster-Server unterstützen Cluster-Gateways also eine IT-Strategie, in deren Rahmen Single Points of Failure in einem Unternehmensrechenzentrum vermieden werden.

TS7650G ProtecTIER™ De-duplication Gateway, Cluster

Cluster mit zwei Knoten

Ein Datenimage

Backup-Server, FC-Switch (optional), TS7650G x 2, Backup-Arrays

- Aktiv-aktiv-Clusterknoten
- Hohe Leistung (900 MB/s, gleichbleibend)
- Hohe Verfügbarkeit
- Hohe Kapazität, hohe Skalierbarkeit
- Problemloses Management
 - Eine Instanz für die Deduplizierung
 - Ein Repository

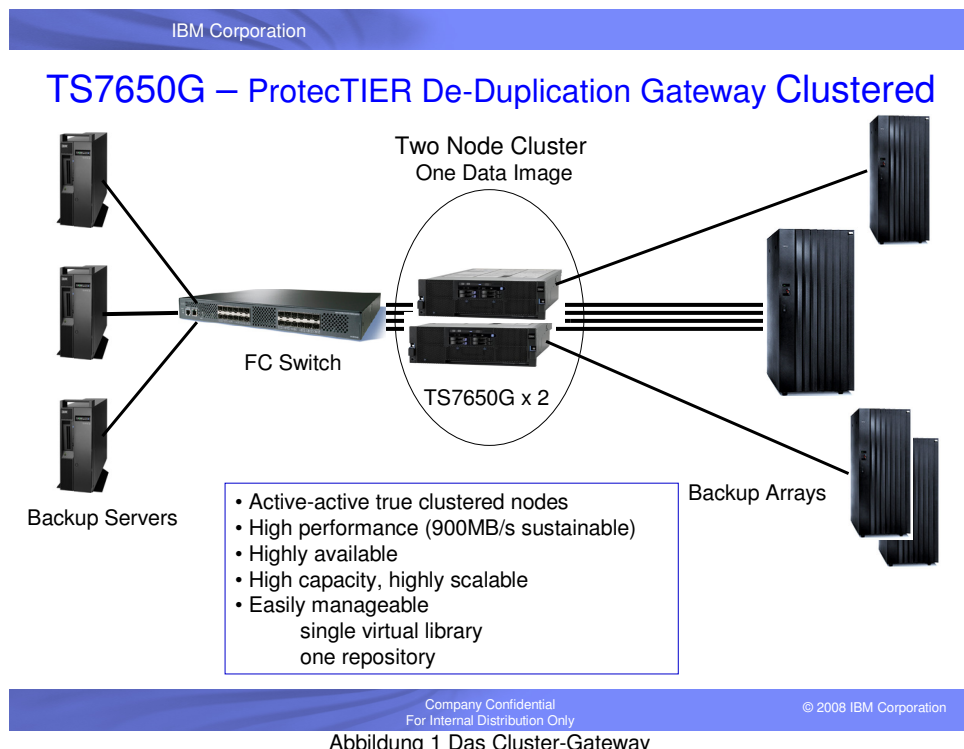


Abbildung 1 Das Cluster-Gateway

Das IBM System Storage TS7650G ProtecTIER De-duplication Gateway

Bei dem Gateway IBM TS7650G, das auf der bewährten Softwarelösung ProtecTIER von Diligent basiert, handelt es sich um ein VTL-Gateway auf einem 4U-Linux-Standardserver. Das Gateway wird zwischen den Back-End-Speicherarrays und den Backup-Media-Servern implementiert und kann ein Petabyte Rohdaten speichern. Es ist standardmäßig mit vier Quad-Core-Xeon-Prozessoren, 32 Gigabyte RAM, Dual-Gigabit-Ethernet-Ports mit TCP/IP-Offload-Engines und vier Dual-Port-PCI-Express-Fibre-Channel-Karten mit 4 GB ausgestattet. Jeder Knoten unterstützt bis zu einem Petabyte physische Kapazität und umfasst die patentierte IBM HyperFactor-Engine für die Datendeduplizierung. Für verschiedene Arrays (z. B. IBM DS4000, DS8000 und XIV) und Platteneinheiten anderer Hersteller ist Plattenunterstützung verfügbar.

Die Einheit ist einzeln und in Clustern mit zwei Knoten erhältlich. Clusterkonfigurationen sind aktiv-aktiv. Dies ermöglicht hohe Leistung und hohen Durchsatz sowie schnelle, halbautomatisierte Failover- und Failbackfunktionalität und minimiert dadurch Clusterausfallzeiten. Wenn ein Knoten ausfällt, ermöglicht Clustering die gemeinsame Nutzung des gesamten Repositories durch beide Knoten. Jeder Knoten hat Lese- und Schreibzugriff auf das Datenrepository und kann auf alle virtuellen Einheiten zugreifen.

Das Gateway TS7650G bietet einen gleichbleibend hohen *Inline*-Durchsatz mit über 450 MB/s pro Knoten. Ein Cluster mit zwei Knoten kann daher in realen Kundenumgebungen Daten mit Durchsatzraten von über 900 MB/s aufnehmen und deduplizieren. Keine andere Lösung für die Inline-Deduplizierung erreicht diese Geschwindigkeit.

Die HyperFactor De-duplication Engine

Die IBM HyperFactor-Engine nutzt proprietäre Algorithmen von Diligent Technologies für die ProtecTIER-VTL für eine schnelle Mustererkennung, um duplizierte Daten zu identifizieren und zu löschen. Dadurch kann der Gesamtbestand gespeicherter Daten (und damit der Bedarf an zusätzlichem Speicher) je nach Umgebung um bis zu 25:1 reduziert werden. HyperFactor unterscheidet sich in mehreren wichtigen Aspekten von anderen Technologien für die Deduplizierung. Erstens: Da der gesamte HyperFactor-Index speicherresident ist, bietet der Prozess einen gleichbleibend hohen, auf Unternehmen abgestimmten Durchsatz von über 450 MB/s pro Knoten. Bei Hash-basierenden Technologien sind häufige Plattenlesevorgänge erforderlich, die entweder mangelhafte Inbandübertragungsgeschwindigkeiten oder eine Außerbandnachbearbeitung, die nicht nur teuer, sondern auch mühsam ist, nach sich ziehen. Zweitens: Während andere Technologien für die Deduplizierung erhebliche Kapazitätsbeschränkungen mit sich bringen, kann die HyperFactor-Lösung Datenrepositories mit bis zu einem physischen Petabyte speichern. Drittens: Da HyperFactor redundante Daten erst nach einem Vergleich der neuen und der vorhandenen Daten auf Byteebene filtert, kann die Lösung auf Unternehmen abgestimmte Datenintegrität unterstützen. Bei anderen Methoden werden Hashing-Codes statt Daten verglichen, was ein gewisses Risiko birgt.

Die EMA-Perspektive

Deduplizierung. Datendeduplizierung kann zu erheblichen Kosteneinsparungen beim Kauf von Hardware und Banddatenträgern führen. Durch die Reduzierung des insgesamt zu verwaltenden Datenbestands können darüber hinaus die tagtäglich anfallenden Kosten für den Datenschutz gesenkt werden. Jedoch unterscheiden sich die Technologien für die Deduplizierung der verschiedenen Anbieter erheblich und einige bieten deutlich mehr Flexibilität und möglichen Nutzen als andere. Jede der drei in diesem Dokument

beschriebenen Technologien für die Deduplizierung – Inlineverarbeitung, Parallelverarbeitung und Nachbearbeitung – hat ihre Fürsprecher und viele Anbieter geben für ihre Produkte extrem hohen Deduplizierungs- und Durchsatzraten an. Der Vorteil der Inline-Deduplizierung ist, dass die Deduplizierung gleichzeitig mit der Datenaufnahme abläuft. Für viele Unternehmenskunden sind VTLs, die sehr hohe Inline-Deduplizierungsraten unterstützen können, interessant. Das Team von ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) empfiehlt eine genaue Prüfung der Leistungsdaten und der Art und Weise, wie diese Daten zustande kommen, bevor eine Investition in Technologien für die Deduplizierung in Betracht gezogen wird (vgl. Seitenleiste).

Gateways und Clustering. EMA ist der Ansicht, dass ein richtig konzipiertes Speichergateway auf zwei verschiedene Arten Nutzen bringt. Erstens können diese Einheiten mit vielen auf Standards basierenden Speichersystemen zusammenarbeiten, wodurch bereits getätigte Investitionen in Speicherhardware geschützt werden können. Wenn also vorhandene Arrays und JBODs hinter dem Gateway miteinander verbunden werden, bieten sie einen Nutzen, solange sie betrieben werden. Zweitens können bei Bedarf an neuen Speichereinheiten Einheiten von verschiedenen Anbietern verwendet werden. Damit kann die Auswahl des Anbieters im gleichen Maße zu einer geschäftlichen und einer technologischen Entscheidung werden. IT-Manager können den Anbieter auswählen, der ihnen am meisten zusagt, Anbieter gegeneinander ausspielen und einen höheren Nutzen aus ihren Investitionen im Speicherbereich ziehen.

Cluster-Gateways haben im Vergleich zu Standardgateways eine höhere Verfügbarkeit. Sie bieten sowohl hohe Leistung als auch Möglichkeiten für die horizontale Skalierung und verbessern gleichzeitig die Business-Continuity. Es sind jedoch nicht alle Cluster-Gateways gleich. Die Funktionalität für eine schnelle Funktionsübernahme von Aktiv-aktiv-Konfigurationen bietet einen Schutz, den Aktiv-passiv-Systeme normalerweise nicht erreichen. Im Falle des in diesem Dokument beschriebenen Systems werden Failover und Failback (sowie Funktionen für die horizontale Skalierung) praktisch ohne Bedienereingriff erreicht. Der Schutz, den automatisierte Failover-/Failbackfunktionalität bietet, bedeutet kontinuierliche Datenverfügbarkeit und sichere Betriebszeiten für kritische Geschäftsprozesse.

Warnung: Das falsche Spiel mit den Deduplizierungs-Benchmarks

Eins ist klar: Je effizienter die Deduplizierung, desto höher der Nutzen einer Lösung für die Deduplizierung. Doch was bedeutet bei der Deduplizierung eigentlich „effizient“? Die Deduplizierung erfolgt im Wesentlichen in zwei Phasen. In der ersten Phase werden Daten in den Prozess aufgenommen (d. h. die Daten werden in die VTL oder eine andere Einheit eingespeist, in der die Deduplizierung stattfindet). Während der zweiten Phase werden diese Daten dann mit dem vorhandenen Datenbestand verglichen, duplizierte Daten werden identifiziert und die nicht duplizierten Daten werden auf Platte geschrieben. Oberflächlich betrachtet mag es den Anschein haben, dass die Deduplizierung einfach zu messen ist – und das ist sie eigentlich auch. Doch die Zahlen können trügen. Anbieter bewerten ihre Prozesse anhand der Stärken ihrer Produkte und geben Durchsatz- und Deduplizierungsraten getrennt an. Dies mag durchaus rechtens sein, jedoch kann diese Darstellung trügerisch sein. Einige Anbieter messen den Durchsatz als Zeitaufwand für die Datenaufnahme und die Deduplizierung, andere nicht. Für einige Anbieter (meist diejenigen, die bei der Deduplizierung auf Nachbearbeitung setzen) ist der Durchsatz die Geschwindigkeit, in der Daten in den Prozess aufgenommen und in die VTL geschrieben werden. Dabei wird jedoch nicht der tatsächliche Zeitaufwand eingerechnet, der für die Deduplizierung der Daten erforderlich ist und bei allen Techniken für die Nachbearbeitung erst beginnt, wenn die Daten auf die Platten der VTL verschoben wurden.

Käufer sollten im Zusammenhang mit der Dateneduplizierung berücksichtigen, dass eine wichtige Messgröße „die Zeit bis zum gewünschten Ergebnis“ ist, also der Zeitaufwand bis ein vollständig deduplizierter Datenbestand auf Platte geschrieben ist.

EMA ist der Ansicht, dass insbesondere hinsichtlich Durchsatzraten der einzige legitime Vergleich der zwischen einem endgültigen Ergebnis mit einem anderen ist. Der Zweck einer Investition in Technologie für die Deduplizierung ist schließlich die Deduplizierung von Daten. Daher ist es wenig sinnvoll, nur einen Teil dieses Prozesses zu bewerten. Käufer sollten im Zusammenhang mit der Dateneduplizierung berücksichtigen, dass eine wichtige Messgröße „die Zeit bis zum gewünschten Ergebnis“ ist, also der Zeitaufwand bis ein vollständig deduplizierter Datenbestand auf Platte geschrieben ist.

Nutzen von Virtual Tape Libraries. EMA hat VTLs schon immer befürwortet, da sie einen erheblichen und einfach zu berechnenden Return-on-Investment (ROI) bieten. VTLs stellen eine bewährte Technologie mit folgenden Hauptvorteilen dar: Beschleunigung der Prozesse für Sicherung und Wiederherstellung, Vermeidung schwerwiegender Leistungseinbußen bei Produktionseinheiten durch die Unterstützung von Sicherungen auf physischem Bandspeicher, Reduzierung der Anzahl zu verwaltender Bänder und relativ einfache Inbetriebnahme ohne zusätzliche IT-Kosten. VTLs bieten jedoch noch einen weiteren Vorteil – einen, der immer wichtiger wird, je größer die Standorte werden.

In Unternehmensrechenzentren werden erhebliche Beträge in Prozesse für die Sicherung und Wiederherstellung investiert und man bemüht sich, diese Prozesse zu automatisieren, um Einsparungen bei den tagtäglich anfallenden Betriebskosten für Sicherung und Archivierung zu erzielen. Unabhängig davon, ob diese Prozesse auf Scripts oder Richtlinien basieren – sie stellen das gesammelte IT-Wissen an den einzelnen Standorten dar und stehen für die Kenntnis der eigenen Best Practices. In großen Rechenzentren wurden möglicherweise über Jahrzehnte Prozesse entwickelt und die Investitionen in Entwicklung, Test und Pflege der Scripts, die diese Prozesse steuern, sind nicht zu unterschätzen. Da eine Virtual Tape Library alle Prozesse so wahrnimmt wie das physische Bandarchiv, das sie repräsentiert, sind die Investitionen in Scripts und Richtlinien geschützt. Wenn eine VTL in das Rechenzentrum eingeführt wird, besteht eine wichtige Änderung im Rahmen der Prozesse darin, dass Sicherungen und Wiederherstellungen schneller durchgeführt werden können, da Daten jetzt auf Platte geschrieben werden und nicht mehr auf physischen Bandspeicher. Physischer Bandspeicher kann weiterhin als erstes Sicherungsziel für Daten verwendet werden, auf die selten zugegriffen wird. Er bleibt eine wichtige Ebene für die Datenaufbewahrung in der Speicherhierarchie und erfüllt die Anforderungen einer langfristigeren Datenaufbewahrung, bei der Überlegungen hinsichtlich Gesamtbetriebskosten und Energieeinsparungen Priorität haben.

Das IBM TS7650G ProtecTIER De-duplication Gateway. EMA hat die Märkte für VTLs und Deduplizierung seit ihrer Entstehung genau beobachtet. Wir empfehlen unseren IT-Kunden VTLs als einfache Möglichkeit zur Einhaltung von Sicherheitsintervallen und zur Bereitstellung einer schnellen Datenwiederherstellung. Ferner empfehlen wir Deduplizierung als effektivste Möglichkeit zur Datenreduzierung. Wir begannen uns mit der HyperFactor-Technologie für die Deduplizierung zu beschäftigen, als Diligent Technologies diese vor mehreren Jahren als integralen Bestandteil von ProtecTIER und VTL vorstellte. Schon damals waren wir von Geschwindigkeit und Effizienz beeindruckt, heute sind wir es noch mehr. Die Durchsatzraten liegen derzeit in realen Kundenumgebungen bei über 450 MB/Sekunde (pro Knoten) – und das gleichbleibend. Durch das Hinzufügen eines zweiten Knotens (Clustering) wird diese Zahl verdoppelt und Durchsatz, Kapazität und Verfügbarkeit steigen. Wenn IBM (das Unternehmen kaufte Diligent Technologies Anfang 2008 auf) die

Clusteringfunktionalität auf mehr als zwei Knoten erweitern kann, verbessern sich die Zahlen weiter. (Die ProtecTIER-Architektur unterstützt künftige Erweiterungen.)

TS7650G bietet erstklassige Funktionalität für die Deduplizierung – mit ausgezeichneten Durchsatzraten.

TS7650G bietet erstklassige Funktionalität für die Deduplizierung – mit ausgezeichneten Durchsatzraten, Dies ist direkt auf die HyperFactor-Technologie zurückzuführen, die dieses Produkt von anderen abhebt. Bei Hashing-Technologien sind häufige Plattenlesevorgänge während des Factoringprozesses erforderlich, was zu einem Leistungsengpass führt. HyperFactor arbeitet jedoch nicht mit Hashing. Alle erforderlichen Metadaten sind in den 4 GB RAM des TS7650G gespeichert, unabhängig von der Größe des Datenrepositorys. Da der Zugriff auf den Factoringindex innerhalb von Mikrosekunden erfolgt (für einen Plattenlesevorgang sind es Millisekunden), ist der Durchsatz bei der Deduplizierung erheblich schneller als bei Technologien anderer Anbieter.

Speicheradministratoren werden das einfache Management großer Sicherungsvolumen beim TS7650G zu schätzen wissen. Alle deduplizierten Daten hinter dem Gateway (ein ganzes Petabyte) werden als einziges Ziel für alle Richtlinien und Media-Server betrachtet. Da das System skalierbar ist, können kürzlich durchgeführte Sicherungen unter Einhaltung der Vorschriften lokaler IT-Richtlinien online bleiben und Wiederherstellungen (von denen der Großteil aus den aktuellsten Sicherungsdaten erfolgt) sehr schnell durchgeführt werden. Darüber hinaus können aufgrund der Datenreduzierung durch HyperFactor gesicherte oder replizierte Daten problemlos elektronisch an ferne Standorte übertragen werden. Die Disaster Recovery-Planung wird dadurch ebenfalls erheblich einfacher.

Erstellung einer Kosten-Nutzen-Analyse. Managern, die in Technologie für die Deduplizierung investieren möchten, sollte es nicht schwerfallen, eine Kosten-Nutzen-Analyse zu erstellen, um eine Investition in TS7650G zu rechtfertigen. TS7650G mit HyperFactor:

- Zögert den Bedarf an kurzfristigen Hardwareinvestitionen hinaus, da mit der HyperFactor-Deduplizierung Speicherplatz auf vorhandenen Arrays frei wird.
- Reduziert den langfristigen Bedarf an zusätzlichem Plattenspeicher, da beim Schreiben neuer Daten in die VTL nur geänderte Datenblocks erfasst werden.
- Wirkt sich positiv auf die Budgets für Strom und Kühlung aus, da der Kauf zusätzlichen Plattenspeichers hinausgezögert wird.
- Ermöglicht den direkten Zugriff auf große Datenmengen und kann daher problemlos genutzt werden, wenn Vorstandsmitglieder eines Unternehmens auf Problemstellungen im Zusammenhang mit interner Governance oder der Einhaltung gesetzlicher Vorschriften reagieren müssen.
- Kann bis zu ein Petabyte physischer Speicherkapazität pro Virtualization Engine unterstützen (dies sind bis zu 25 Petabyte deduplizierte Kapazität).
- Bietet einen gleichbleibend hohen Durchsatz bei der Inline-Deduplizierung von 900MB/s bei zwei Clusterknoten.
- Optimiert die Verarbeitungsanforderungen, da das Produkt mit Inline-Deduplizierung arbeitet. Die redundanten Daten werden schon bei der Aufnahme reduziert. Aus diesem Grund ist auch kein spezieller Zeitraum für eine Nachbearbeitung erforderlich wie bei vielen anderen Produkten für die Deduplizierung.
- Sorgt für hohe Datenintegrität.

Informationen über Diligent, ein IBM Unternehmen

Informationen über Diligent Technologies:

Diligent wurde im April 2008 von der IBM Corporation aufgekauft. Das Unternehmen stellt bewährte skalierbare plattenbasierte Datenschutzlösungen speziell für Großunternehmen her. Die ProtecTIER™-Plattform von Diligent mit HyperFactor™ ist eine revolutionäre Technologie für die Inline-Deduplizierung von Daten, mit der Kunden mehr Daten schützen können, während sie weniger Daten aufzeichnen – ohne bestehende Richtlinien, Verfahren und Prozeduren in ihren Rechenzentren ändern zu müssen. Wenn Sie mehr über Diligent erfahren möchten, besuchen Sie folgende Website: <http://www.diligent.com>

Informationen über IBM:

Wenn Sie mehr über IBM System Storage erfahren möchten, besuchen Sie folgende Website: <http://www.ibm.com/storage>

Informationen über Enterprise Management Associates, Inc.

Das im Jahr 1996 gegründete Unternehmen Enterprise Management Associates (EMA) ist ein führendes Analysten- und Consultingunternehmen, das sich auf den Markt für IT-Management konzentriert. Das Unternehmen bietet IT-Anbietern und IT-Experten in Unternehmen objektive Informationen über den realen geschäftlichen Nutzen etablierter und neuer Technologien unter anderem in den Bereichen Sicherheit, Speicher und IT Service Management (ITSM) sowie Configuration Management Database (CMDB), Virtualisierung und serviceorientierte Architektur (SOA). Trotz des schnellen Unternehmenswachstums hat EMA nie den Kunden aus den Augen verloren und bietet weiterhin personalisierte Unterstützung und einfache Kontaktaufnahme mit seinen Analysten. Weitere Informationen über das umfangreiche Archiv des Unternehmens im Bereich IT Management Research, das gebührenfreie und online verfügbare IT Management Solutions Center sowie IT-Consultingangebote erhalten Sie unter www.enterprisemanagement.com.

Dieser Bericht darf ohne die vorherige schriftliche Zustimmung von Enterprise Management Associates, Inc. weder vollständig noch teilweise kopiert, vervielfältigt, in einem Datenabfragesystem gespeichert oder weitergegeben werden. Alle in diesem Dokument enthaltenen Meinungsäußerungen und Einschätzungen stellen die gegenwärtige Einstellung von EMA dar und können sich jederzeit ändern. Die in diesem Dokument genannten Produktnamen können Marken und/oder eingetragene Marken der jeweiligen Unternehmen sein.

©2008 Enterprise Management Associates, Inc. Alle Rechte vorbehalten. EMA™, ENTERPRISE MANAGEMENT ASSOCIATES® und das Mobius-Symbol sind eingetragene Marken oder Marken von Enterprise Management Associates, Inc.

Corporate Headquarters:

5777 Central Avenue, Suite 105
Boulder, CO 80301
Tel.: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com