

BUYER CASE STUDY

VMware verbessert die Netzwerk-Auslastung und die BackupPerformance durch Nutzung von EMC Avamar-Deduplizierung

Laura DuBois

IDC-MEINUNG

In physisch begrenzten Rechenzentrums-umgebungen bietet die Deduplizierungstechnologie erhebliche Einsparpotenziale in den Bereichen Speicheroptimierung und benötigtem Rechenzentrumsplatz. Ein Client-seitiger Deduplizierungsansatz verspricht jedoch noch weitere Vorteile hinsichtlich einer verbesserten Netzwerkbandbreitennutzung und kleineren Backup-Fenstern. Für Unternehmen, die eine Servervirtualisierung nutzen oder Daten von entfernt gelegenen Standorten und Zweigstellen schützen müssen, bietet ein Client-seitiger Ansatz Vorteile durch kleinere Backup-Fenster und geringeren Netzwerk-I/O, den Wegfall von lokalen Backups an entfernt gelegenen Standorten und Bandinfrastrukturen sowie durch die Entlastung stark beanspruchter WAN-Verbindungen. Im Rahmen dieser IDC-Fallstudie werden folgende Themen behandelt:

- Untersuchung der Vorteile einer Client-seitigen Deduplizierungstechnologie bei VMware Inc.
 - Überlegungen zur Implementierung von EMC Avamar im VMware-Rechenzentrum
 - Darstellung der Zukunftspläne für die Nutzung der Technologie an den entfernt gelegenen Standorten und Zweigstellen von VMware
-

INHALT DIESER FALLSTUDIE

Thema dieser IDC-Fallstudie ist die Nutzung der Client-seitigen Deduplizierung im Unternehmen VMware. Die Studie basiert auf einem Interview mit Chanh Chi, IT-Architekt bei VMware, und Michael Pate, Manager of Technical Operations bei VMware, aus dem März 2009. Im Jahr 2007 hat VMware die Avamar-Lösungen für Client-seitige Deduplizierung und Replikation für eine Absicherung und eine Recovery von über 300 physischen und virtuellen Systemen in der gesamten Infrastruktur des Unternehmens evaluiert und bereitgestellt.

In diesem Dokument wird die Implementierung der Avamar-Technologie untersucht, und die organisatorischen Herausforderungen und Evaluierungsanforderungen werden dargelegt. Die Resultate der Implementierung sowie die zukünftige Implementierungsplanung bei VMware werden geschildert. Anwender können diese Informationen nutzen, um zu beurteilen, ob die Client-seitige Deduplizierungstechnologie für ihre Umgebung

eine praktikable Lösung darstellen könnte und welche Vorteile durch die Nutzung von Avamar-Lösungen zu erwarten sind.

SITUATIONSBESCHREIBUNG

U n t e r n e h m e n s ü b e r b l i c k

VMware (notiert an der NYSE als VMW) ist ein führendes Unternehmen im Bereich von Virtualisierungslösungen vom Desktop bis zum Rechenzentrum. VMware hilft anderen Unternehmen dabei, die Kosten zu senken, Business Continuity zu ermöglichen und höhere Sicherheits- und Umweltstandards zu erfüllen. VMware verzeichnete 2008 einen Umsatz von 1,9 Mrd. US-Dollar und arbeitete mit mehr als 130.000 Kunden und über 22.000 Partnerunternehmen zusammen. VMware ist eines der am schnellsten wachsenden Softwareunternehmen und befindet sich im Mehrheitsbesitz der EMC Corporation.

Die IT-Umgebung bei VMware setzt sich aus Hunderten von physischen und virtuellen Windows- und Linux-Servern zusammen, auf denen Anwendungen wie Microsoft Exchange, Microsoft SQL und Oracle laufen. Dazu kommen Verzeichnis-, Datei- und Print-Server. Auf seinen gesamten physischen und virtuellen Servern hat das Unternehmen ein Datenvolumen von circa 50 TB gespeichert. Die IT-Abteilung unterstützt über 6.000 Anwender.

Als Mitarbeiter der IT-Organisation von VMware konnten Chanh Chi und Michael Pate Einblicke in die Herausforderungen bei VMware geben und die Evaluierung einer neuen Backup-Technologie und die Implementierungsergebnisse mit Avamar beschreiben. Chanh Chin spielte aufgrund seiner Tätigkeit als IT-Architekt und seiner vorherigen Tätigkeit als Backup-Administrator eine maßgebliche Rolle bei der Konzeption einer neuen Ausrichtung und Strategie für die IT-Architektur zur Bewältigung der Herausforderungen in den Bereichen Backup und Recovery.

Die Herausforderungen und die Lösung

Die IT-Organisation von VMware sah sich mit mehreren Herausforderungen konfrontiert. Die wachsende Infrastruktur brachte das Rechenzentrum an die Grenze der physischen Kapazität. In der vorhandenen Umgebung des Unternehmens wurde EMC NetWorker für Backups sowohl auf physische Bänder als auch auf CIFS-verbundene Data Domain-Deduplizierungs-Appliances eingesetzt. Dadurch wurde die Anzahl physischer Bänder um 200 Stück pro Woche reduziert, und deren externe Aufbewahrung konnte entfallen. Darüber hinaus wurde auf ein Savvis-Colocation-Rechenzentrum in Palo Alto, Kalifornien, als Disaster-Recovery-Standort repliziert. Es war offensichtlich, dass ein größeres, stabileres und energieeffizientes Rechenzentrum dringend erforderlich war. Gleichzeitig benötigte man eine Lösung, die aktuelle und zukünftige Anforderungen an die Absicherung von Daten erfüllen musste. Aus diesem Grund entschied sich die IT-Abteilung dafür, die etablierte Backup-Strategie neu zu bewerten und die Umgebungsarchitektur neu zu gestalten.

Während der Evaluierung der Backup-Umgebung identifizierte die IT-Abteilung von VMware Netzwerkengpässe als einen zentralen Problembereich. Natürlich nutzt

das Unternehmen die eigenen VMware-Lösungen zur Konsolidierung und Maximierung der Nutzung der verfügbaren Server. Das Unternehmen war zu über 95 % virtualisiert, und alle VM-Gastsysteme nutzten dieselbe zugrunde liegende physische Netzwerkschnittstelle auf den physischen Servern. Die bei VMware etablierten Backup-Methoden beanspruchten während des während des rund um die Uhr verfügbaren Backup-Fensters das verfügbare Netzwerk zu 100 %. Dies führte zu schwerwiegenden Backup-Engpässen. Eine zu große Menge Daten, die abgesichert werden musste, und immer kleinere Backup-Fenster machten es erforderlich, einmal pro Woche gestaffelte, vollständige Backups für jeden Client durchzuführen. Den Rest der Woche musste man sich – bestenfalls – mit inkrementellen Backups behelfen. „Wir wussten, dass diese Methode bei unserem erwarteten Datenwachstum nicht weiter skalierbar war“, so Chanh Chi.

Im Jahr 2006, vor der Akquisition durch EMC, erwog VMware drei Optionen zur Adressierung der Backup-Herausforderungen. Die Erweiterung der Backup-Umgebung mit einer neuen deduplizierungsfähigen Anwendung für die virtualisierte Umgebung wurde in Erwägung gezogen. Eine weitere Option war die Verwendung von EMC Disk-Library-Lösungen in der Backup-Umgebung. Da jedoch die hauptsächlichen Probleme das immer kleiner werdende Backup-Fenster und Netzwerkengpässe in der virtualisierten Umgebung waren, befasste man sich primär mit Client-seitigen Deduplizierungslösungen. Infrage kommende Systeme waren Symantec NetBackup PureDisk und EMC Avamar. Da zum Zeitpunkt der Evaluierung Symantec NetBackup PureDisk nicht alle Clients in der Umgebung von VMware (wie z. B. Oracle-Server und NAS-Filer) unterstützte, schied diese Lösung aus.

Das IT-Team von VMware führte vor Ort einen Proof-of-Concept zum Vergleich der Performance aller drei Upgrade-Szenarien durch: Backup auf Band mit EMC NetWorker, Backup auf Data-Domain-Deduplizierungs-Appliances mit EMC NetWorker sowie Backup auf generische Festplatten mit EMC Avamar. Nach Bereitstellung eines Avamar-Systems zur Evaluierung wurde die Lösung etwa zwei Monate in der Umgebung des Unternehmens getestet. Das IT-Team von VMware führte mehrere Tests durch, in denen die Backup-Ergebnisse derselben zwölf Windows- und Linux-Dateiserver und eines Exchange-Backups mit einem Gesamtvolumen von etwa einem halben Terabyte Daten verglichen wurden. Die Tests zeigten eindeutig, dass die Backup-Performance von Avamar trotz einer langsameren Netzwerkverbindung den Alternativen überlegen war. Chanh Chi beschreibt das Ergebnis folgendermaßen: „Die Deduplizierungsraten schwankten zwar, aber durchschnittlich kamen wir beim ersten Backup der meisten Clients auf eine Optimierungsrate von 3:1. Bei einigen Windows-Dateisystem-Clients ergaben sich bei den folgenden Backup-Durchläufen sogar Optimierungsraten von 500:1.“

Die Resultate

Wegen der überzeugenden Ergebnisse der durchgeführten Proof-of-Concept-Tests erwarb VMware Avamar und stellte die Lösung im Dezember 2007 bereit. Zwar gehörte VMware zu diesem Zeitpunkt bereits zu EMC, doch die Avamar-Bereitstellung erfolgte nach dem Börsengang von VMware, sodass VMware das Produkt kaufen musste. Das neue Produktionsrechenzentrum in Santa Clara, Kalifornien, nahm gegen Ende 2007 den Betrieb auf. Laut Chanh Chi „war ein wichtiger Aspekt bei unserer Entscheidung für Avamar die Fähigkeit zur Remote-Replikation, die uns ermöglicht, Backup-Daten zwischen dem neuen

Produktionsrechenzentrum und dem Disaster-Recovery-Rechenzentrum zu replizieren, wodurch das Band-Handling entfällt.“

VMware erwarb zwei Avamar-Systeme, bestehend aus jeweils zwölf Daten-Nodes. EMC Professional Services unterstützte VMware bei der Avamar-Installation, die zwei Tage in Anspruch nahm. Ein Avamar-System befindet sich in Palo Alto, Kalifornien, und jeden Tag werden Daten bidirektional zwischen den beiden Avamar-Systemen repliziert. VMware begann mit zwölf Daten-Nodes in jedem Avamar-Grid. Das erste Grid diente der Absicherung von etwa 12 TB primärer Daten, das zweite von etwa 7 TB primärer Daten. Grid-übergreifend sichert die Avamar-Lösung über 300 virtuelle und physische Dateiserversysteme ab. EMC NetWorker wurde für das direkte Backup der EMC-Celerra-Filer auf Band und das Backup der Exchange-Umgebung auf vorhandene Data-Domain-Devices beibehalten. Um der gestiegenen Anzahl zu unterstützender Clients gerecht zu werden, hat das IT-Team von VMware pro Grid vier weitere Speicherknoten mit einer Kapazität von 16 TB pro Grid hinzugefügt; zusätzlich wurde ein weiteres Grid mit höherer Dichte mit einer Kapazität von 32 TB gekauft.

- ☒ **Backups virtueller Maschinen.** VMware stellt innerhalb jedes VM-Gasts Avamar-Softwareagenten bereit. Avamar kommt auch in der eigenen VMware-Consolidated-Backup-(VCB)-Umgebung für einige wenige Clients, die eine besondere Absicherung erfordern, zum Einsatz. In allen Fällen ermöglicht Avamar schnelle, vollständige tägliche Backups.
- ☒ **Backup-Fenster.** „In den ersten zehn Minuten haben 50 % der 360 Clients das tägliche vollständige Backup abgeschlossen, Nach einer Stunde fehlen nur noch wenige Clients“, sagt Chi. Vor der Avamar-Implementierung dauerten Backups einen ganzen Tag oder länger.
- ☒ **Backup-Performance.** Bei nur einem virtuellen Client mit zwölf LUNs, jede mit 415 GB Daten, würde ein Backup mit NetWorker eine Woche dauern. Jetzt nimmt das Backup von 5 TB und 40 bis 50 Millionen Dateien mit Avamar etwa elf Stunden in Anspruch.
- ☒ **Netzwerk-Performance.** Insgesamt führt die Organisation jede Nacht Backups mit einem Volumen von 18 TB bei einem Netzwerk-Traffic von 200 GB durch. Dies entspricht etwa einem Prozent der Netzwerkbandbreite, die für das herkömmliche Backup-Verfahren erforderlich war.
- ☒ **Recovery-Performance.** Die Daten-Recovery mit Avamar ist wesentlich einfacher und schneller als mit früheren Lösungen. „Wenn wir früher für einzelne fehlerhafte oder beschädigte Dateien eine Recovery durchführen mussten, nahm dies für einen 200-KB-File einige Stunden in Anspruch“, so Chi. Mit Avamar ist eine Recovery schon nach wenigen Augenblicken abgeschlossen. Die fehlerhaften Daten werden auf dem Client durch die fehlerfreien ersetzt. Wir haben schon 8-GB-Dateien in nur 30 Sekunden wiederhergestellt.“
- ☒ **Benutzeroberfläche.** Chi erklärt: „Die Avamar-Benutzeroberfläche ist sehr intuitiv nutzbar, es ist keine wirkliche Backup-Erfahrung notwendig. So können manche Backup-Aufgaben auf einen größeren Kreis von IT-Mitarbeitern verteilt werden. Release-Engineering-Teams können zum Beispiel Backups und die Recovery von Quellcode selbst durchführen.“

VMware bewahrt derzeit Backup-Daten von etwa 30 Tagen auf den festplattenbasierten Avamar-Systemen auf. Wöchentlich werden Backups auf Band für die langfristige Aufbewahrung sowie Drittkopien für den Fall der Fälle erstellt.

Weitere Planung

Michael Pate meint dazu: „Wir haben in der letzten Woche unsere Avamar-Kapazität verdoppelt. Wir haben zusätzlich 52 TB Avamar erworben, um die Daten unserer internationalen entfernt gelegene Standorte für Forschung und Entwicklung sowie Engineering sichern zu können. Wir haben ein Grid in einem neuen Rechenzentrum in Washington State platziert, das Backups von zehn entfernt gelegenen Standorten und Zweigstellen übernimmt.“ VMware nutzt den lokalen Avamar-Softwareagenten zur Durchführung von Backups auf ein größeres Rechenzentrum. Hier erleichterte die Avamar-Lösung den Umgang mit der beschränkten Anzahl von WAN-Verbindungen zum größeren regionalen Rechenzentrum. Dieser Ansatz hat es VMware ermöglicht, Backups von primären Dateiservern an Remote-Zweigstellen auf ein Avamar-Datenspeicher-Grid innerhalb eines größeren Rechenzentrums durchzuführen. Der Einsatz von Bandwechslern an entfernt gelegenen Standorten und Zweigstellen entfällt dadurch, ebenso wie die Kosten für ca. 1.000 Bänder pro Quartal.

ERKENNTNISSE UND EMPFEHLUNGEN

Tipps für Anwender

Bei der Bewertung der potenziellen Vorteile, die der Einsatz von Datendeduplizierungstechnologien in ihrer Umgebung erbringen kann, sollten Anwender folgende Punkte berücksichtigen.

- Deduplizierung ist nicht gleich Deduplizierung.**
Deduplizierungstechnologien können in mehreren verschiedenen Bereichen zum Einsatz kommen, z. B. bei Backup-Software, Backup-Appliances usw. Es empfiehlt sich, dass Anwender die zu lösenden Hauptprobleme benennen, und auf Grundlage dieser Herausforderungen der Deduplizierungsansatz ausgewählt wird, der am ehesten zu deren Bewältigung geeignet ist.
- Proof-of-Concept-Tests sind erforderlich.** Deduplizierung wird heute stark nachgefragt, und viele Anbieter bieten Technologien an, die angeblich Deduplizierungsfunktionen bieten. IDC hat jedoch festgestellt, dass einige Deduplizierungslösungen in der Proof-of-Concept-Phase nicht das halten konnten, was die Anbieter versprechen.
- Tests mit den realen Daten, die gesichert werden sollen.** Die Deduplizierungsraten variieren, je nachdem, für welche Art von Daten Backups durchgeführt werden sollen. Manche Daten lassen sich gut deduplizieren, bei anderen Arten von Content können nur geringere Deduplizierungsraten erreicht werden. Die Kenntnis realistischer Deduplizierungsraten anderer Unternehmen – wie in diesem Dokument beschrieben – und der Deduplizierungsraten, die im Rahmen Ihres Proof-of-

Concept-Tests ermittelt werden, ermöglicht Ihnen, Überraschungen bei der Inbetriebnahme der Technologie zu vermeiden.

Berücksichtigung anderer Optimierungsarbeitslasten.

Optimierungstechnologien wie Komprimierung und Deduplizierung verbessern die Wirtschaftlichkeit der Nutzung von Festplatten für die Datensicherung und ermöglichen es, mehr Backups für längere Zeiträume auf Festplatte aufzubewahren. Ihr Unternehmen sollte sicherstellen, dass die Komprimierung, Deduplizierung (sowie die Verschlüsselung) sich ergänzen und in der richtigen Reihenfolge ausgeführt werden, damit von den maximalen Vorteilen der Optimierung profitiert wird.

WEITERE INFORMATIONEN

Verwandte Forschungsergebnisse

- Worldwide Storage Software 2009–2013 Forecast: Quest for Efficiency Driving Investments in Storage Software* (IDC #217529, März 2009)

- Worldwide Data Protection and Recovery Software 2007 Vendor Shares: Market Rebounds* (IDC #213320, Juli 2008)

Copyright

Dieses IDC-Dokument wird als Teil des kontinuierlichen IDC Informationsangebots veröffentlicht, das schriftliche Untersuchungen, den Austausch mit Analysten, Tele-Briefings und Konferenzen umfasst. Informieren Sie sich unter www.idc.com über das Abonnement von IDC-Publikationen und Consulting-Services. Ein Verzeichnis der weltweiten IDC-Niederlassungen finden Sie unter www.idc.com/offices. Wenden Sie sich an die IDC-Hotline unter 800 343 4952, Nebenstelle 7988 (aus den USA) oder +1 508 988 7988 bzw. sales@idc.com, um Informationen darüber zu erhalten, wie Sie den Bezugspreis dieses Dokuments beim Erwerb eines IDC-Services verrechnen können, oder für Informationen zu zusätzlichen Exemplaren oder Rechten bezüglich der Veröffentlichung im Internet.

Copyright 2009 IDC. Eine nicht autorisierte Reproduktion ist untersagt. Alle Rechte vorbehalten.