Microsoft Exchange auf NetApp Storage
Snap Manager for Exchange und "Thin Provisioning"

Wiederholt wurden wir gebeten, Exchange-Datenbankserver auf NetApp-Storage zu konfigurieren. Besonders problematisch waren dabei Konfigurationen für Exchange 2007-CCR-Lösungen mit vielen Speichergruppen auf mehreren CCRs, die zur Konsolidierung sehr großer bestehender Exchange- oder Lotus-Mail-Systeme vorgesehen waren.

Leider stießen wir dabei wiederholt auf Konfigurationen, die nicht zielführend und nachvollziehbar vorgenommen worden waren. Vor allem die Parametrierung der NetApp-Instanzen ist für jede Art Exchange-Server (Exchange-Server 2007 ohne CR, mit LCR, als CCR, als SCC oder in der SCR-Version) relevant. Hierbei wurden häufig die Möglichkeiten des „Thin Provisioning“ falsch betrachtet.

Im Folgenden geben wir eine Übersicht über die relevanten Größen und die Möglichkeiten, die bei der Verwendung von SnapDrive und SnapManager für Exchange (SME) in Frage kommen.

Idee des „Thin Provisioning“

Grundsätzlich sollte bei der Entscheidung für NetApp als Storage-System klar sein, dass nicht nur eines von vielen SAN Arrays ausgewählt wurde, sondern ein System, welches eine Reihe von integrierten Lösungen und Diensten beinhaltet. Im Falle von Exchange sind dies die Möglichkeiten schneller Backups, schneller und bequemer Restores, die Option des Single Mail(box)Restores und anderer mehr. Allerdings benötigen diese Funktionen und Dienste zusätzliche technische Unterstützung durch die Bereitstellung von Speicherplatz, der über den Bedarf der aktiven Exchange-Instanzen hinausgeht.

Damit verschlechtert sich jedoch auch das Verhältnis von benötigtem und tatsächlich durch die Applikation genutztem Speicherplatz (brutto-netto-Verhältnis). Gleichzeitig gilt für viele Applikationen oder Speicherplatz nutzenden Funktionen, dass nicht sofort der kalkulierte Bedarf an Speicherplatz abgerufen wird. In den eingangs genannten Beispielen der zu konsolidierenden CCR-Server wird mit der vollen Belegung der Exchange-Server in Bereichen von Monaten bis Jahren kalkuliert: die Migration der alten Systeme erfolgt nicht auf einen Schlag; anschließend dauert es noch einmal, bis die Mailboxen die nach oben angepassten neuen Limits erreicht haben.

Hier nun setzt ganz allgemein die mit dem Begriff „Thin Provisioning“ bezeichnete Überlegung für die Konfiguration an: man stellt den Nutzern der Speichersysteme den erwarteten Platz nur scheinbar zur Verfügung. Tatsächlich verfügt das Storagesystem nur über eine Kapazität, die dem momentan benötigten Platzbedarf entspricht zuzüglich einer Reserve, die bei Erhöhung des Speicherplatzbedarfs sofort abgerufen werden kann. Ist die Reserve erschöpft, dann muss entweder der Bedarf der Storage-Nutzer nach unten korrigiert oder aber das System erweitert werden. Normalerweise wird man sich für den zweiten Weg entscheiden, da dieser unter Beibehaltung der vereinbarten Dienstleistung ein dynamisches Wachsen des Storage-Systems nach Bedarf gestattet. Beim Thema „Thin Provisioning“ geht es kurz gesagt darum, den vorhandenen Storage so effektiv wie möglich zu nutzen und „keine leeren Platten drehen zu lassen“.

Parameter, mit denen „Thin Provisioning“ allgemein realisiert wird

NetApp stellt eine Reihe von Whitepapern bereit, in denen „Thin Provisioning“ beschrieben wird. Alle weisen darauf hin, dass es keine festgelegte Vorgehensweise im Sinne von Step 1 bis Step „n“ zur Umsetzung von „Thin Provisioning“ gibt, sondern immer eine Reihe von verschiedenen Maßnahmen zu Einsatz kommt, um möglichst viele Dienste mit dem vorhandenen Storage anzubieten.

Leider wird in der allgemeinen Betrachtung des Themas der Fokus hauptsächlich auf den Platzbedarf gelenkt, den die Applikationen unmittelbar benötigen und nicht auf den Bedarf zur Bereitstellung zusätzlicher Dienste. Dazu später weitere Überlegungen.

Am Beispiel der im NETAPP TECHNICAL REPORT „Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment" von Rick Jooss, Jan 2008 | TR-3483 vorgestellten möglichen Konfigurationen läßt sich feststellen:

Thin Provisioning wird gesteuert über die Volume-Optionen *guarantee*, das *Volume autosizing*, das (automatische) Löschen von SnapShots und die Lun-Space-Reservation (enable/disable). Dabei wird teils der verfügbare freie Platz im Aggregat als gemeinsamer Pool genutzt, teils mit festen Volume- und Lun-Sizes gearbeitet und bei Bedarf werden Snapshots gelöscht. Von den vorgestellten Beispielkonfigurationen sind jedoch nicht alle realisierbar oder sinnvoll.

Will man z.B. die Fractionale Reserve (FR) eines Volumes < 100% setzen (100% ist der Default-Wert), so muss die Volume guarantee auf Volume stehen. Nimmt man einer LUN die Reservierung, so wird die NetApp bei jedem Neustart des Dienstes SnapDrive am Windows-Host ihrerseits eine ungültige Thin-Provisioning-Konfiguration melden.
Natürlich ist es verlockend, die Vol guarantee auf „file" oder gar „non" zu setzen – hat man dann doch bei einem neu aufgesetzten Exchange ohne Daten scheinbar jeden (Storage-)Platz der Welt auf dem eigenen Filer. Entscheidet man sich für „file", dann meldet das Volume den von Files belegten bzw. für diese reservierten Platz als genutzten Platz. Diesen theoretischen Ansatz haben wir in der Praxis tatsächlich umgesetzt vorgefunden. Für Exchange ist dieser Ansatz jedoch unbrauchbar; vor allem dann, wenn keine Reserven vorgesehen werden.

Welchen Platzbedarf hat Exchange bei der Nutzung von SME

Wie arbeitet Exchange

Exchange arbeitet mit Transaktionsprotokoll-orientierten Datenbanken. Anders formuliert: alles, was in eine Datenbank gelangt, muß vorher durch die Log-Dateien. Bei Exchange 2007 sind die Logs je ein MB groß (bei Exchange 2003 sind es 5 MB) und werden fortlaufend hochgezählt. Fällt ein Datenbank-File aus, so stellt man es aus einer Sicherung wieder her und bringt durch Nachspielen der seit dem Backup angefallenen Log-Files die Datenbank wieder auf den aktuellen Stand. Voraussetzung ist allerdings, dass die Log-Files bis zur nächsten Sicherung aufgehoben werden, die Datenbank also nicht im Umlaufprotokoll-Modus betrieben wird. Nach einer Vollsicherung werden alle angefallenen Logs gelöscht. Der NetApp-SME gestattet es darüber hinaus, die Log-Files auch für längere Zeiträume aufzuheben.

Speichergruppen und Datenbanken

Zu jeder Datenbank (DB) gehören Log-Files. Eine Exchange-Speichergruppe (SG) beinhaltet einen Satz Log-Files und eine oder mehrere Datenbanken. Es empfiehlt sich, je Speichergruppe nur eine Datenbank zu betreiben; für Exchange 2007 CCR und SCR ist dies vorgeschrieben. Datenbank-Dateien und Log-Files sollten auf unterschiedlichen Datenträgern untergebracht werden, so dass der Ausfall eines Datenträgers nicht zum Verlust aller Daten der Speichergruppe führt (siehe oben). Nutzen wir die NetApp als Speichersystem, so benötigen wir also je Speichergruppe zwei LUNs: eine für die DB und eine für die Logs. Da NetApp Snapshots auf Volume- und nicht auf LUN-Ebene erstellt, verlangt der SME die Platzierung dieser zwei LUNs auf unterschiedlichen Volumes. Nur so kann auf Snapshot-Ebene ein Restore unterschiedlicher (zeitlicher) Kombinationen von DB- und Log-Dateien erfolgen.

Bei den von uns angepassten CCR-Clustern wurde mit jeweils rund 50 SG gearbeitet. Das heißt: Mit mehr als 100 LUNs, also über 100 Volumes je ClusterNode und damit insgesamt mit über 200 LUNs bzw. Volumes.

Für die Datenbanken- und für die Log-File-Volumes wurden jeweils eigene Aggregate bereitgestellt.

Größenberechnung

Grundlage für die Größenberechnung der oben beschriebenen LUNs ist das Mailaufkommen: Die (maximale) Größe einer DB kann recht genau über das Mailbox-Limit, die Zahl der Mailboxen in der DB und die Aufbewahrungszeit gelöschter Objekte ermittelt bzw. eingegrenzt werden. Für den Platzbedarf der Log-Files ist das Mailaufkommen pro Zeiteinheit entscheidend. Hier können Erfahrungswerte bestehender Systeme oder eine Hochrechnung zu einer Größenermittlung verwendet werden. Dabei ist zu beachten, dass das Mailaufkommen über eine Woche variiert, z.B. ist es Montagvormittag höher als Sonntagnachmittag.

Je mehr Datenbanken und Storagegroups betrieben werden, umso schwieriger und mühseliger wird das Sizing für jede SG/DB, so dass der Einfachheit halber mit pauschalierten Werten gearbeitet wird. So auch in unseren Beispiel-Konfigurationen.

Kleines Rechenbeispiel für den SME

Für die weitere Betrachtung unterstellen wir eine Datenbank-Größe von max. 100 GB und ein Aufkommen von LOG-Files von max. 20 GB pro Tag und Speichergruppe. Nun gestaltet sich die weitere Betrachtung des benötigten Storage schon interessanter: Im Bereich der LOG-Files wird eine 20 GB große LUN also pro Tag einmal überschrieben (Änderungsrate 100% / Tag). Die NetApp weiß jedoch nicht, dass nach einem Backup fast alle Blöcke dieser LUN aus Windows-Sicht zum Überschreiben freigegeben bzw. gelöscht sind – sie sieht nur, dass sich in der LUN Daten befunden haben, die mit einem Snapshot auch in diesem enthalten sind.

SME ist ein von NetApp angebotenes Backup-Programm, welches das Erstellen, Aufbewahren und Wiederherstellen von Exchange-Backups ermöglicht. Möchte ich also Backups nicht nur erstellen, sondern auch eine Zeit lang verfügbar haben, so, muss dafür der entsprechende Platz vorgehalten werden.
Für die weitere Betrachtung unterstellen wir eine Datenbank-Größe von max. 100 GB und ein Aufkommen von LOG-Files von max. 20 GB pro Tag und Speichergruppe.

Das bedeutet: Im Bereich der LOG-Files wird eine 20 GB große LUN einmal pro Tag überschrieben (Änderungsrate 100% / Tag). In unserem Beispiel werden also für die LOG-Dateien – grob betrachtet – für jeden Tag Backup, das aufbewahrt werden soll, 20 GB an Platz je SG benötigt.

Bei den Datenbanken ist die Änderungsrate weitaus geringer. Sie hängt vom Nutzungsverhalten der Anwender ab. So wurden von uns auf 35 MB limitierte Mailboxen beobachtet, deren Inhalt nach einer Woche komplett überschrieben war – aufgrund der geringen Größe sind die Benutzer gezwungen, alte Mails zu löschen, um Platz für neue zu schaffen. Bei einer 5-Tage-Arbeitswoche liegt die Änderungsrate damit bei 20%. Bei größeren Mailboxen werden mehr Mails aufbewahrt, so daß die Änderungsrate geringer ausfällt. In unserem Beispiel gehen wir von 2% pro Tag aus.

Wollen wir also Backups für fünf Tage aufheben, so benötigen wir für die Datenbankdaten 100 GB + 5x2% = 110 GB und für die Log-Files 20 GB x 5 = 100 GB. Und dies je Speichergruppe und je CCR-Node und unabhängig davon, ob wir unsere Volumes und Luns für „Thin-Provisioning“ konfigurieren oder nicht.

Der falsche Ansatz

Ein Beispiel: Die DB-Luns werden erst einmal nicht näher betrachtet, da ihre Größe bei einem neuen System tatsächlich über einige Tage hinweg bis auf die Zielgröße anwachsen wird. Bei den Logs ist das etwas anderes: wir schieben den Inhalt von nur fünf je 2 GB großen Mailboxen während einer Migration in die Datenbank. Diese ist dann maximal 20 GB (von 100 GB) groß – und das Verzeichnis der Log-Files ist durch das Verschieben („Alles“ muss durch die Log-Files !) ebenfalls mit 20 GB, also

der maximal vorgesehenen Größe, gefüllt. In der Beispielkonfiguration wurde die LUN der Logs auf 25 GB Größe definiert, so daß der Füllstand der LUN 80% beträgt. Dies entspricht zum einen der Microsoft-Empfehlung und bietet zum anderen 5 GB Reserveplatz innerhalb der Lun (Luns können nicht automatisch vergrößert werden).

Es gibt 50 Volumes im Aggregat, dessen Snapreserve aus Gründen der Platzersparnis auf 0% gesetzt wurde. Das die Lun hostende Volume ist auf 35 GB Größe mit der Möglichkeit des automatischen Wachsens konfiguriert. Die Vol guarantee steht auf „file“, der Platz für die LUN ist garantiert. Das Aggregat ist 2 TB groß, so daß zusätzlich zu den bestehenden Volumes entweder weitere sieben Volumes angelegt werden können oder aber jedes Volume auf max. 40 GB anwachsen kann.

Im Ausgangszustand meldet das Aggregat 1250 GB belegten Platz (50 x 25 GB pro LUN) – wir haben also mehr als ein Drittel freien Platz (37,5 %) – der will ja erst einmal belegt werden!

Was passiert hier ?

Die Volumes melden die LUN-Größe als eigene Größe (=25 GB).

Am ersten Tag wird die LUN komplett gefüllt – mit 20 GB. Ein Snapshot wird erzeugt. Die vom Snapshot und vom aktiven Filesystem genutzten Blöcke sind identisch. Im Volume sind noch 15 GB verfügbar.

Am zweiten Tag werden erneut fünf 2 GB große Mailboxen migriert. Das aktive Filesystem belegt erneut 20 GB – nachdem 75% der neuen Daten migriert sind, muss das Volume auf 40 GB erweitert werden, da sonst im aktiven FS nicht genügend Platz zur Verfügung steht.

Am dritten Tag wiederholen wir den Migrationsvorgang – unsere DB ist nun gerade mal zu 60 % gefüllt. Bei den Logs geht es jedoch nicht weiter: die Volumes können wir nicht weiter vergrößern, denn mehr als 40 GB pro Volume sind nicht verfügbar. Also bleibt nur, Snapshots zu löschen: wir müssen also den Snapshot vom Vortag löschen, um Platz für den heutigen Snapshot zu schaffen.

Vom Ziel, einige Tage Backup vorzuhalten, ist somit gerade einmal ein Tag geblieben. Der Zugriff auf die Daten im Bereich der Log-Files hat sich stark verlangsamt – klar, das Aggregat ist zu 100% (gearbeitet wird ohne Snapreserve !) belegt; die Volumes melden ebenfalls 100% Füllstand.

Nicht einmal Exchange ist zufrieden – der SME verschiebt alle aufzubewahrenden Logs in das Snapinfo-Verzeichnis unterhalb des Log-Verzeichnisses, so dass auch das aktive Filessystem (aus Windows-Sicht) gefüllt ist. Aber das können wir ändern: wir löschen durch den SME nicht mehr benötigte Logs und verzichten auf die Möglichkeit des „Up-to-the-Minute-Restores“ vom Backup des Vortages, denn dieses wird ja ohnehin gelöscht.

Übrig bleibt: Möglichkeit der Wiederherstellung des letzten Backups mit Nachfahren der seit diesem Zeitpunkt aufgelaufenen Log-Files. Außerdem können wir gegen unseren einzigen verbliebenen Snapshot mit Single-Mailbox-Restore arbeiten.

Das Ergebnis steht in keinem Verhältnis zu den Erwartungen, den Möglichkeiten und dem Preis der Software. Dabei hat die Konfiguration am ersten Tag noch den Eindruck erweckt, wir hätten ein Drittel des Platzes als Reserve.

Noch ein falscher Ansatz

Bei dem bisher vorgestellten Versuch, „Thin Provisioning“ für Exchange zu implementieren, wurde die Aufgabe des Überwachens der Speicherplatzbelegung, also das „Space Monitoring“, an Ontap delegiert. Ontap kann Volume für Volume auf knapp werdenden Platz reagieren und bis zu einem

definierten Maße Volumes vergrößern um anschließend Snapshots des betreffenden Volumes zu löschen – auch die umgekehrte Reihenfolge ist möglich.

Damit „hintergehen“ wir jedoch den SnapManager für Exchange (SME). Für diesen besteht ein Backup aus drei Teilen: dem Snapshot auf dem DB-Volume, dem Snapshot auf dem Log-File-Volume und dem zum Backup gehörenden Snapinfo-Verzeichnis im aktiven Filesystem der Log-File-Lun. Löscht SnapManager ein Backup, dann werden alle drei Teile dieses Backups entfernt und der Datenbestand ist konsistent. Löscht hingegen Ontap, so wird nur genau ein Snapshot gelöscht. Die übrigen Snapshots der anderen Volumes belässt Ontap unverändert.

Für den SME ist solch ein „halb“ entferntes Backup jedoch nicht mehr gültig und wird nicht mehr angezeigt. Der Administrator muss nun manuell ermitteln, ob Snapshots in einem der Datenbereiche existieren, für die es keine Zuordnung im anderen Bereich mehr gibt und hat diese dann manuell zu entfernen. Tut er dies nicht, so bleiben die Snapshots erhalten; die Dateiblöcke sind zum überschreiben gesperrt und der verfügbare Platz auf dem betreffenden Volume geht schneller zur Neige als erwartet.

Bei mehr als 200 Volumes – wie in unserer Beispielumgebung, kann sich der Abgleich zwischen Ontap-Snapshots und SnapManager-Backups – als äußerst aufwendig bis nicht zu bewältigen gestalten. Hier wäre eine zusätzliche Funktion wünschenswert, die die SME-Backups mit den Snapshots auf den betroffenen Volumes vergleicht und bei Bedarf überzählige Snapshots entfernt.

Das Löschen von Backups mit dem Ziel der Bereitstellung freien Speicherplatzes sollte also nicht durch Ontap, sondern durch den SME erfolgen.

Die etwas andere Betrachtung des Themas „Thin Provisioning“

Nach den vorangegangen Ausführungen wird deutlich, dass die üblichen Methoden des "Thin Provisioning" auf Exchange-Umgebungen kaum anwendbar sind. Trotzdem ist "Thin Provisioning" für Exchange möglich. Im SME erfolgt dies über die Festlegung von „Fractional space reservation policies“.

Dabei muss uns klar sein, dass der von Exchange benötigte Platz tatsächlich bereitgestellt werden muss. Die Volume guarantee steht dazu auf „volume“ und die LUN reservation ist aktiviert. Wo lässt sich nun der Platz effektiver nutzen und dynamisch zuteilen ?

Wenn nicht im Bereich des von der Applikation genutzten Storages, dann muss es im Bereich der von NetApp angebotenen Dienste (Backup und Restore, Single Mailbox Restore) sein.

In diesem Bereich finden wir eine Lösung, die uns hilft, die Zahl der aufzubewahrenden Backups an den verfügbaren Platz anzupassen – und dies dynamisch.

Schreibbare Zugriffe mit Flexclone

In früheren Snapmanager-Versionen (bis zur Version 4) wurden von NetApp Formeln zur Größenberechnung der Volumes und Luns im SME-Administration-Guide aufgelistet. Außerdem gab es die Empfehlung, zusätzlich zum durch die aktive Lun und die Backups benötigten Speicherplatz noch einmal den Platz für ein schreibbares Snapshot der Lun (rws-Snapshot) vorzuhalten. Der Sinn dieser Empfehlung liegt darin begründet, dass bei Bedarf ohne Einschränkung auch schreibend auf den Snapshot der Lun zugegriffen werden kann.

Im Tagesgeschäft tritt dieser Bedarf jedoch nicht auf, da SME beim Verify der Exchange-Daten nur lesend auf den Snapshot zugreift. Auch beim Single-Mail(-Box)-Restore wird nur lesend auf dem Snapshot gearbeitet. Außerdem sollte berücksichtigt werden, dass SME zu einem Zeitpunkt immer

nur auf einen Snapshot zugreift und – sowohl beim Backup als auch beim Verify – die Speichergruppen nacheinander abarbeitet.

Benötigt man den vollen Platz der Lun schreibbar im Snapshot (z.B., um offline zu defragmentieren), so ist es geschickter, mit einer FlexClone-Lizenz zu arbeiten. Der Einsatz dieser Lizenz erlaubt es, den Platz für einen schreibbaren Snapshot nur einmalig im Aggregat statt pro Volume vorzuhalten.

Für die Möglichkeit, bei Bedarf einen schreibbaren Snapshot bereitzustellen, empfiehlt sich also das Vorhalten der Größe der größten LUN als SnapReserve im Aggregat unter Nutzung von Flexclone.

Lun und Volume Sizing

Exchange-Datenbanken

Die Planung der Lun- und Volume-Größen für Datenbanken ist relativ einfach. Die Datenbanken haben eine tatsächliche oder geplante Größe. Die Lun sollte 20% größer als die Datenbank sein. Im Volume muss die Lun Platz finden zuzüglich des Platzes für aufzubewahrende Backups, der aus der Änderungsrate und der Zahl aufzubewahrender Backups ermittelt werden kann. Dazu ein Beispiel: unsere Datenbank sei 100 GB groß und beinhaltet 50 Mailboxen zu je 2GB. Single-Instanz-Ratio und Dumpster bleiben der Einfachheit halber unberücksichtigt (nehmen wir an, sie rechnen sich gegeneinander auf). Das Volume muss damit wenigstens 125 GB groß sein, so daß die DB 80% Füllstand entspricht. Die Änderungsrate sei 2% pro Tag; 10 Tage Backups sollen vorgehalten werden: 0,02 * 100 GB * 10 = 20 GB. Insgesamt sollte uns also für die 100 GB große Mailbox-Datenbank ein Volume von 145 GB die Möglichkeit bieten, Backups für 10 Tage vorzuhalten und eine ausreichende Reserve bereitzustellen.

Dieses Verhältnis ist akzeptabel: Für 45% „Overhead“ erhalten wir ein schnelles Storage-System und Backups für zehn Tage. In unserer Beispielumgebung mit 50 Datenbanken in aktiver und passiver Instanz ergeben sich damit also folgende Werte: 50 Datenbanken à 100 GB (200 Mailboxen mit je 500 MB Inhalt versorgen 10.000 Mitarbeiter) benötigen 50 * 2 * 145 GB = 14,5 TB auf dem Storagesystem.

Verzichten wir auf das Vorhalten von Backups, dann reduziert sich der Bedarf um 20 GB je Speichergruppe, wir benötigen also 50 * 2 * 125 GB = 12,5 TB.

Anders formuliert: mit 2 TB zusätzlichen Platz erkaufen wir uns für 10.000 User, die jeweils eine 500 MB große Mailbox nutzen, das Backup über 10 Tage hinweg. In Anbetracht der Tatsache, dass ein externes Medium 12,5 TB für ein Vollbackup benötigt, ein effektiver Ansatz.

Dass die Datenbanken oftmals nicht in ihrer kalkulierten Größe vorliegen, sondern diese erst nach einer Zeit des Anwachsens erreichen, lässt an die Möglichkeiten des „Thin Provisioning“ denken – aber Vorsicht: Luns können derzeit nicht automatisch erweitert werden. Damit sind also in jedem Fall manuelle Eingriffe notwendig.

Exchange-Logdateien

Was geschieht in den Volumes, in denen die Log-Dateien verwaltet werden?

Die Betrachtung der Abläufe in diesem Bereich ist etwas aufwendiger, was in der Arbeitsweise des SME begründet liegt. Mit dem SME lassen sich entweder alle Logs des gesamten Aufbewahrungszeitraumes vorhalten oder nur diejenigen, die während der Erstellung des Backups gerade „in Abarbeitung“ sind. Im ersten Falle lassen sich Up-to-the-Minute Restores ausführen; im zweiten Fall nur Point-in-Time-Restores. Die Festlegung für die Art der Aufbewahrung von Logs gilt jeweils pro Backup-Management-Gruppe (Standard, Daily, Weekly).

Welche Art des Restores benötigen wir in welcher Situation ?

Geht eine Datenbank verloren oder wird inkonsistent, dann benötigen wir ein konsistentes (=verifiziertes) Backup und alle danach aufgelaufenen LogFiles. Es gehen keine Informationen verloren. Dies ist die übliche Art des Restores von Exchange-Datenbanken.

Sollen Ereignisse, die sich auf die DB ausgewirkt haben, ungeschehen gemacht werden, so stellen wir das letzte Backup wieder her, ohne jedoch alle LogFiles nachzupielen. Damit gehen allerdings auch Mails verloren. Situationen für diese Art des Restores können sein: Virenbefall oder ein falsch konfigiurierter Virenscanner, der alle Anhänge entfernt oder andere ungewöhnliche Situationen.

Was muss also aufbewahrt werden? Nun, auf jeden Fall alle seit dem letzten Backup angefallenen Logfiles. Stellen wir allerdings inkonsistente Datenbanken nicht sofort fest, sondern erst nach zwei Tagen, so sollten besser auch die Logs diese zwei Tage lang zusätzlich aufbewahrt werden. Es spielt also eine Rolle, wie schnell die Administratoren auf beschädigte Datenbanken reagieren können – so lange sollten Logs aufbewahrt werden.

Im Gegensatz zu den Datenbanken beträgt die Änderungsrate bei den Log-Dateien 100%.

Wie reagiert nun die NetApp ?

Zuerst sei nochmals darauf hingewiesen, dass SME alle notwendigen Logs je Backup in einen Unterordner des SnapInfo-Verzeichnisses (dem Gedächtnis des SME) verschiebt. Deshalb ist es sinnvoll, das SnapInfo-Verzeichnis auf demselben Laufwerk zu platzieren wie auch die Logfiles – SME verschiebt in diesem Fall die Files einfach, was um ein Vielfaches schneller geht als ein tatsächliches Kopieren (ältere SME-Versionen haben prinzipiell mit Kopieren gearbeitet und dann pro SG schon mal eine Stunde oder mehr dazu benötigt).
Haben wir also 20 GB LogDateien pro Tag und wollen zwei Tage aufbewahren, dann müssen 60 GB Daten in die Log-Lun passen. Für die Aufbewahrung von 10 Tagen sind dies entsprechend 220 GB.

Verzichten wir auf die Möglichkeit des Up-to-the-Minute-Restores, so werden nur die Logs aufbewahrt, die gerade durch den Informationsspeicherdienst in die DB eingearbeitet werden. Dann müssen nur diese wenigen Logs und die Logs des aktiven Filesystems Platz in der Lun finden.
Sollen nun aber alle Logs der letzten zwei Tage und für weitere acht Tage nur die minimal benötigten Logs aufgehoben werden, um auf die Datenbanken zwecks Single-Mailbox-Restore zuzugreifen, dann ist dies nur annähernd und nur durch „gegenläufiges“ Arbeiten verschiedener Backup-Management-Groups möglich. Anregungen und Lösungen dazu bekommen Sie von uns.

Der völlige Verzicht auf die Speicherung von Log-Files ist nicht möglich, da zu einem Exchange-Backup immer der Inhalt der Datenbank und die gerade in Abarbeitung befindlichen LogDateien gehören.

Betrachten wir nun näher, wie die NetApp die Datenbewegungen im Volume mit den LogFiles wahrnimmt: Wir gehen dazu von einer schon im Betrieb befindlichen Lun aus, da während des ersten Füllens der Lun andere Effekte auftreten.

Die Lun ist 65 GB groß und dimensioniert für einen Umfang von 20 GB pro Tag bei Aufbewahrung aller Logs für die Dauer von drei Tagen. 5 GB sind Reserve. Damit liegen vor der letzten Sicherung des Tages 20 GB aktuelle Logs im Speichergruppen-Verzeichnis. 20 GB befinden sich im Snapinfo-Verzeichnis des Vortages und 20 GB im Snapinfo-Ordner des Tages davor. Alle Logs und deren Inhalte sind eindeutig, d.h. für die NetApp sind für die 60 GB auch entsprechend viele Blöcke im WAFL-Filesystem zugewiesen.

Mit dem Backup des aktuellen Tages löscht Windows den Ordner der Sicherung von vor zwei Tagen und verschiebt die aktuellen Logs in einen neuen Snapinfo-Ordner (daher finden wir auch pro SME-Backup zwei Snapshots auf dem Volume der Log-Lun aber nur eines auf dem der Datenbanken: das

erste Snapshot entspricht dem Zustand bei Erstellung des DB-Backups und das zweite dem nach dem „Aufräumen“ in den Snapinfo-Verzeichnissen). Für Windows ist nun alles klar: es gibt erneut 20 GB freien Platz zur Aufnahme neuer LogFiles.

Für den Filer stellt sich dies aber ganz anders dar: für ihn sind alle Blöcke belegt; er muss auch die Blöcke aufheben, die Windows nach dem Backup neu beschreibt. Ontap weiß nicht, dass ein Teil der im Snapshot liegenden Blöcke nicht mehr benötigt werden (mit dem Space-Reclaimer könnte zwar eine Anpassung erfolgen; jedoch ist dies ein aufwendiger Prozess, den man nicht täglich für insgesamt 100 Luns ausführen kann).

Da zu einem SME-Backup – wie schon wiederholt angesprochen – immer auch Snapshots aller betroffenen Volumes gehören, ergibt sich nun für das Log-Volume folgender Datenbestand: 60 GB im aktiven Filesystem, davon 40 GB belegt mit den Logs der letzten zwei Tage(n-1 und n-2) und 20 GB verfügbar für neue Logs (n[=0]). Der SnapShot des Vortages mit den Logs des Vortages und der zwei Tage davor (also der vergangenen drei Tage: n-1, n-2 und n-3) und der Snapshot von vor zwei Tagen mit den Logs der zurückliegenden Tage (n-2, n-3 und n-4). Die Logs der Vortage sind also in mehreren Snapshots enthalten und belegen **keinen** zusätzlichen Platz. Die nicht mehr in der BackupSequenz enthaltenen Logs der beiden älteren als die von uns aufbewahrten Backups sind im Volume allerdings zusätzlich enthalten: wir benötigen also 60 GB in der Lun (für die Tage n, n-1 und n-2) und 100 GB im gesammten Volume (40 GB im Snapshot-Bereich; für die Tage n-3 und n-4), um für drei Tage Logs vorzuhalten. Im Volume ist also der Platz für die beiden nicht mehr sichtbaren Backups (2 * 20 GB) mit einzuplanen.

Die Anzeige (df-Befehl) für das Volume (bei angenommener Größe von 100 GB) sieht damit wie folgt aus: used Space=100 GB, availible Space=0 und reserved Space=65 GB. Der reservierte Platz entspricht unserer LUN-Größe, denn die Lun wurde space reserved angelegt. Der genutzte Platz setzt sich zusammen aus den Daten im aktiven Filesystem (65 GB, von denen 60 GB durch Windows beschrieben wurden) und den durch Snapshots belegtem Platz (2 * 20 GB = 40 GB). Die Snapshot-Belegung wird uns auch so (mit 40 GB) angezeigt; selbst wenn die SnapReserve auf null gesetzt worden ist.

Was geschieht nun nach dem letzten von uns erstellten Backup ?

Exchange überschreibt die für Windows als freier Bereich sichtbaren 20 GB im aktiven Filessystem. Der Filer hat diese Blöcke jedoch als im Snapshot belegt markiert und muss demzufolge für die aktuellen Zugriffe freie Blöcke im WAFL neu zuweisen. Die 20 GB „wachsen“ in den SnapShot-Bereich des Volumes hinein. Dort ist in unserer bisherigen Konfiguration jedoch kein Platz mehr, so dass der Platz dem aktiven Filesystem „weggenommen“ werden muß: nachdem 10 GB neue Logs geschrieben wurden, wird Exchange die Datenbank dismounten, da trotz der vom Windows-Filesystem gemeldeten noch freien 10 GB keine Schreibzugriffe mehr ausgeführt werden können. Derselbe Effekt tritt auch in anderen Bereichen auf: Der für Snapshots genutzte Bereich darf auch über die per SnapReserve festgelegte Größe hinaus wachsen; im CIFS-Betrieb sieht man große freie Bereiche auf einem vom Filer bereitgestellten Share – trotzdem kann darauf nicht geschrieben werden. Ein recht bizarres Verhalten, dessen Ursache einem erst bewusst werden muss.
Die 100 GB im Volume reichen noch nicht aus: wir benötigen für die Phase bis zur Erstellung des nächsten Backups weitere 20 GB – und zwar exklusiv zum Bereitstellen neuer Blöcke für das aktive Filesystem. Die beste Lösung wäre, diesen Platz zu reservieren. Der Parameter, den NetApp für solche Zwecke vorgesehen hat, ist die FractionalReserve(FR). Diese legt fest, wieviel Prozent des belegten oder für Luns reservierten Platzes zum Überschreiben vorgehalten wird. Die FR wird pro Volume festgelegt und beträgt in unserem Beispiel 20 GB von 65 GB, also ca. 31 %. Damit schaffen wir eine Overwrite-Reserve von 20 GB (von 120 GB sind dies ca.17%) – und benötigen nunmehr ein 120 GB großes Volume.

Damit ist das Volume sechsmal so groß wie die aktuell benötigte Datenmenge – der Mehrbedarf wird benötigt für die Möglichkeit, die Logs der zurückliegenden drei Tage zwecks Up-to-the-Minute-Restore innerhalb beeindruckend kurzer Zeit zurückzuspielen.

Möglichkeiten, die SnapDrive und SME bieten

Nachdem die wichtigsten Zusammenhänge dargestellt worden sind soll nun der Hinweis folgen, wie NetApp alle zu beachtenden Größen in eine Lösung implementiert hat:

Das Löschen alter Backups (ein Ansatz für Thin Provisioning) hat bei der Arbeit mit SME durch den SME zu erfolgen. Dazu muss dieser in der Lage sein, Platzengpässe zu erkennen. Die Konfiguration dazu erfolgt im Menüpunkt „Fractional Space Reservation Settings“ des SME.

Die zu überwachende Größe ist die „Fractional Overwrite Reserve“, die wir über die FR definieren können. Die FR darf dazu nicht auf null gestellt werden – mit diesem Wert wäre die „used overwrite reserve“ immer 100% und wir hätten keine Möglichkeit, eine Schwellwert zum Löschen alter SnapShots zu definieren. Eine FR von 100% ist auch nicht sinnvoll – wir wollen ja nicht die Gesamtgröße der Lun zum Überschreiben vorhalten. Unser Beispiel von 31% veranschaulicht die Zusammenhänge besser.

Nach Definition einer FR lässt sich nun pro Volume ein „Trigger point for overwrite reserve utilization“ in % (Default: 70%) festlegen, bei dessen Erreichen Snapmanager beginnt, alte Backups zu löschen. Gleichzeitig lässt sich festlegen, wieviele Backups erhalten bleiben sollen und nicht zu löschen sind (Default: 5 Backups). Sind also 70% der „Überschreibreserve“ (in unserem Beispiel: 14 GB von reservierten 20 GB) mit Daten gefüllt, dann löscht SME das älteste Backup.

Sind nur noch fünf Backups vorhanden, dann wird nicht weiter gelöscht, sondern jetzt wird auf Erreichen des zweiten Schwellwertes gewartet, den „Trigger point of overwrite reserve utilization“ für das automatische Dismounten von Datenbanken (Default: 90%).

Thin Provisioning – aus anderer Sicht

Die bisher von uns genutzte Beispielrechnung geht von 20 GB Logs pro Tag und SG aus. Dieser Wert ist allerdings der höchste anzunehmende Wert für den Umfang der LogFiles. In der Praxis wird dieser Wert überwiegend nicht erreicht. Fallen also nur 10 GB statt 20 GB Datenmenge an, so können wir auf denselben Luns und Volumes doppelt so viele Snapshots und damit Backups vorhalten als geplant.

Geben wir also vor, dass mindestens ein Backup vorgehalten werden soll und nach 10 Tagen alte Snapshots durch den SME zu entfernen sind, dann können wir die „Fractional Space Reservation Settings“ dazu nutzen, zwischen einem und zehn Backups vorzuhalten. Die Lun sollte dazu 45 GB groß sein: 20 GB im Snapinfo des letzten aufzubewahrenden Backups und 20 GB im aktiven Filesystem sowie 5GB Reserve.

Im Volume ist die FR von 20 GB sowie einmalig 20 GB für das letzte Backup vorzusehen. Damit kann mit einer Volume-Größe von 85 GB eine Backup-Aufbewahrungsfrist zwischen einem und mehreren Tagen erreicht werden.

SME wird die Zahl der Backups an den aktuellen Bedarf anpassen und uns damit eine minimale Zahl von Backups garantieren sowie alternativ nach Möglichkeit auch weitere Backups vorhalten.

Die Qualität und die Quantität der Dienste für Backup/Restore werden somit dynamisch angepasst: "Thin Provisioning" für Exchange.

Liste der für Exchange relevanten Parameter

Vol options Fractional reserve %	Legt fest, welcher Anteil des belegten oder reservierten Platzes fürs Überschreiben reserviert werden soll
Vol options Guarantee volume	Das volume meldet ans Aggregat seine festgelegte Größe als tatsächliche Größe (nötig für FR <100%)
Vol options Extend off	Nein, beim SME-Verify soll Ontap die Luns im Snapshot nicht splitten, soviel Platz haben wir nicht
Vol options Try_first_snap_delete	Beim automatischen Löschen von Snapshots sind die ältesten Snaps zuerst zu löschen
Snap sched 0 0 0	Die Snapshots erstellt und löscht der SME – nicht Ontap, daher automatische Snapshotverwaltung ausschalten
Snap reserve 0	Die Schwellwerte zur Feststellung voller Volumes holen wir aus der overwrite-Reserve; deshalb hier 0
Vol autosize Off	Da Luns nicht automatisch vergrößert werden hilft vol autosize nur, um mehr Backups aufzuheben – wir löschen stattdessen alte Backups
Snap autodelete On	Muss aktiviert warden, sonst kann SME Snapshots der Volumes nicht löschen
Snap autodelete Delete_order_oldest_first	Alte Backups zuerst löschen
Snap autodelete Trigger space_reserve	Wir überwachen, ob der reservierte Platz zur Neige geht (Trigger point of overwrite reserve utilization)
Snap autodelete Target_free_space 1	Snap autodelete „on“ verlangt einen Wert für Ontap: wie lange alte Snaps gelöscht werden: Wert klein halten, damit SME das Löschen veranlasst und nicht Ontap
Lun set reservation Enable	Exchange Luns sind im Betrieb immer gefüllt – das können wir dem System auch so mitteilen
SME: Trigger point of overwrite reserve utilization: delete Backups … %	Wenn schon % der overwrite reserve belegt sind wird SME mit dem Löschen alter Backups beginnen (70%)
SME: Trigger point of overwrite reserve utilization: backups to retain *n*	Die Zahl der von uns garantierten Backups (5)
SME: Trigger point of overwrite reserve utilization: automatically dismount dbs %	Wenn schon % der overwrite reserve belegt sind werden die Datenbanken außer Betrieb genommen, um Schäden am Datenbestand zu vermeiden (90%)