

Virtuelle Tape Libraries, Überblick und Merkmale



Martin Mrugalla
martin.mrugalla@empalis.com



Inhalt

- Was ist eine Virtuelle Tape Library (VTL)?
- Mögliche Gründe für eine VTL im TSM Umfeld
- Klärung der Begriffe – VTL, VTA
- Was ist Dedup, was ist zu beachten?



Was versteht man unter einer VTL?

Was ist eine VTL?



- Eine VTL ist eine einfache Sichtweise auf ein grosses Plattensystem das sich als sequentielle Band Volumes und Laufwerke darstellt.
- Hinter allen VTL Produkten steckt der gleiche Grundgedanke: 'Backup to Disk' Geschwindigkeit mit der Funktionalität der Bandnutzung
- Manche Hersteller bieten die Migration der gesicherten Daten vom Platten-cache auf Band (destaging).
- Einige bieten auch De-Duplizierung
- Viele bieten auch Tape-Cloning
- Die meisten Lösungen sind vollkommen transparent für die eingesetzte Sicherungssoftware

Wo liegen die Stärken einer VTL?



- Sie kann das Backup beschleunigen weil die Daten nicht auf Band, sondern auf ein schnelles Disk-System geschrieben werden.
- VTLs in der Lage, auch sehr große Datenmengen innerhalb des nächtlichen Backup-Zeitfensters zu sichern.
- Sie bieten, besonders bei der Wiederherstellung von vielen kleinen Dateien, einen Geschwindigkeitsvorteil gegenüber Bandbibliotheken
- Sie bieten erhebliche Vorteile bei gestreuten Lesevorgängen
- Sie erlauben eine nahtlose Integration in die vorhandene Infrastruktur. Da VTLs eine Bandbibliothek emulieren, sind keine Änderungen an der vorhandenen Backup-Struktur erforderlich. Die Server verwenden wie bisher ihren Backup-Client



Welche Gründe sprechen für die Anschaffung einer VTL im TSM Umfeld?

Gründe für eine VTL im TSM Umfeld



TSM arbeitet mit seinen vorgelagerten Diskpools schon ähnlich wie eine VTL. Es gibt trotzdem Gründe, die für die Anschaffung einer VTL im TSM Umfeld sprechen.

- Die administrativen Aufgaben wie Migration, Backup Storagepool und Reclamation werden innerhalb des zur Verfügung stehenden Zeitfensters kaum noch fertig
- Viele Clients sollen gleichzeitig lan-free gesichert werden
- Die bestehenden Diskpools bieten keine ausreichende Performance
- Zur Bewältigung der täglichen Aufgaben ist eine erhebliche Investition in Speicherhardware notwendig



VTL vs VTA

VTL vs VTA



Bei virtuellen Tape Libraries unterscheidet man grundsätzlich zwischen zwei Typen, VTL und VTA.

Merkmale eines VTA:

- Es gibt immer eine physikalische Library
- Relativ kleiner Disk Cache
- FC Disk Cache
- Die virtuellen Volumes sind meist klein
 - Sind meist kleiner als die physikalischen Bänder im 'Backend'.
- Band Stacking
- Migrierte Volumes müssen in den Disk Cache zurückgeladen werden bevor sie genutzt werden können

VTL vs VTA



Vorteile eines VTA:

- integrierte „Disk-to-Disk-to-Tape“-Datensicherung (D2D2T)
- Vollständige Virtualisierung der Bandlösung
- Volumegrößen sind unabhängig von der eingesetzten Bandtechnologie frei wählbar
- Kann TSM von den administrativen Aufgaben Migration und Erstellen von Copypools komplett entlasten. Reclamation kann bei geeigneter Wahl der virtuellen Volumegrößen deutlich verringert.

VTL vs VTA



Merkmale einer VTL:

- Eine physikalische Library ist nicht zwingend notwendig, bei einigen Herstellern aber möglich
- Sehr grosser Plattenspeicher
- Meist große SATA Platten
- Die virtuellen Volumes sind frei wählbar
 - Sie entsprechen normalerweise der von physikalischen Bänder im 'Backend' wenn eine reale Tapelibrary zum Klonen der Volumes eingesetzt wird.

VTL vs VTA



Vorteile einer VTL:

- Es wird keine physikalische Library benötigt
- Hoher Durchsatz, da das Filesystem für Backup optimiert ist
- Wenn virtuelle Volumes auf Band 'gecloned' werden kann die Sicherungssoftware bei Ausfall der VTL direkt auf diese Bänder zugreifen



De-Duplikation

De-Duplikation



Wie arbeitet De-Duplikation?

- De-Duplication-Lösungen untersuchen die zu sichernden Daten auf dem Block- oder Byte-Level und speichern identische Daten nur ein einziges Mal auf dem Backup-System ab. Zu welchen Dateien ein Block- beziehungsweise Byte-Segment gehört, merkt sich das System mithilfe von Metadaten.
- Nach den Aussagen der Hersteller soll sich das Volumen der Backup-Daten durch derartige Lösungen unter **günstigen** Umständen auf ein Fünfzigstel reduzieren lassen.

De-Duplikation



In-Line De-Duplikation

- Die De-Duplizierung erfolgt bevor die Daten auf dem Disk-Cache abgelegt werden
- Der Durchsatz wird massgeblich durch die Leistungsfähigkeit der eingesetzten Prozessoren, der Hauptspeichergröße und teilweise durch die eingesetzten Festplatten bestimmt
- Die Daten können meist unverzüglich nach dem Abspeichern in eine weitere VTL dupliziert werden
- Bei der Variante Source De-Duplizierung erfolgt der Prozess auf Client-Seite über eine Backup-Software Funktion
- Bei Destination Dedup erfolgt der Prozess im Zielsystem. Hier wird keine Dedup Funktion in der Backup Software benützt oder benötigt.

De-Duplikation



Delayed In-Line De-Duplikation

- Die De-Duplizierung erfolgt ähnlich wie bei In-Line Deduplication nach dem Zwischenspeichern der Daten. Der Zeitversatz bis zum Beginn der Deduplikation kann sich dabei auf weniger als eine Minute reduzieren.
- Der Durchsatz wird maßgeblich durch die Leistungsfähigkeit der in der VTL eingesetzten Prozessoren bestimmt
- Die Daten können sofort nach dem Abspeichern in eine weitere VTL dupliziert werden
- Transparent für die Backup Software

De-Duplikation



Was ist Out-of-the-Line De-Duplikation?

- Die De-Duplizierung erfolgt erst nach dem Backup der Daten
- Der Durchsatz wird maßgeblich durch die eingesetzten Prozessoren, die Hauptspeichergröße und hier auch stärker durch die Leistungsfähigkeit der Diskspindeln bestimmt
- Die Daten können erst nach dem De-Duplizieren in eine weitere VTL dupliziert werden
- Transparent für die eingesetzte Backup Software

De-Duplikation



Vorteile- Nachteile der De-Duplikation

•**Vorteile:**

- Einsparungen bei den Backup-Medien (Bänder, VTL)
- transparente Integration in bestehende Backup-Lösungen möglich

De-Duplikation



Vorteile- Nachteile der De-Duplikation cont.

•

Nachteile:

- Für den Vergleich beziehungsweise die Bildung des Hash-Werts ist mehr Rechenleistung notwendig
- Mitführen einer Index-DB/Referenzliste
- Dateien sind stark fragmentiert und eventuell über zahlreiche Medien verstreut
- nicht für revisionssichere Daten geeignet
- Bei Inline-Deduplikation kann die Backup Geschwindigkeit leiden

De-Duplikation



Im folgendem Umfeld arbeitet De-Duplikation nicht oder nur ungenügend:

-
- Wenn Daten vom Client bereits komprimiert wurden
- Wenn Daten verschlüsselt sind
- Wenn Multiplexing genutzt wird (Veritas, Legato usw.)
- Wenn die 'Nullen Komprimierung' bei Datenbanken genutzt wird

De-Duplikation



Mögliche Risiken der De-Duplikation

- Was passiert wenn die Hash (Metadaten) DB korrupt ist?
- Was passiert bei einem Memory Fehler?
- Was passiert wenn ein 'Block' der ersten Version geschädigt wird?
- Ein Teil der Dedup Implementierungen bergen ein kleines Hash Key Collision Risiko (d.h. sie verlassen sich darauf, daß bei gleichem Hash auch der dahinter stehende Datenbereich gleich ist). Dies kann zu einem unbemerktem Datenverlust führen.

Zusammenfassung:

- Deduplizierung kann die langjährige Robustheit der Datensicherung aufgrund der komplexen Logik kaum im vollem Maße garantieren – daher erscheint Ptape beim Einsatz von Deduplikation für eine Zweitkopie vorerst ratsam.
- Auch Geschwindigkeitseinbussen können manchen sonst sinnvollen Einsatz von Deduplikation verhindern.



Fragen ?????