

Link: <https://www.computerwoche.de/a/datensicherung-im-virtuellen-umfeld,2360745>

Ratgeber

Datensicherung im virtuellen Umfeld

Datum: 22.12.2010
Autor(en): Johann Baumeister

Auch in virtuellen Umgebungen sind Backup und Restore ein Muss. Dabei ändern sich die Datensicherungs-Konzepte in der Theorie nicht. Sie müssen in der Praxis aber anders umgesetzt werden.

Das oberste Ziel für den Einsatz von IT liegt in der Bereitstellung der Daten und der Applikationen zu deren Bearbeitung. Durch unterschiedliche Konzepte zur Absicherung soll dabei eine möglichst hohe Verfügbarkeit von beiden, Daten und Programmen, erreicht werden. Um diese zu gewährleisten, werden allerdings unterschiedliche Techniken angewandt.

Damit die Daten stets verfügbar sind, greift man heute noch meist auf die traditionellen Backup-Verfahren zurück. Doch die Vorkehrungen zur Datensicherung sind weitaus vielfältiger. Sie beginnen beim Aufbau eines **RAID-Plattenverbund**¹, ziehen sich fort über die unterschiedlichen Sicherungstechniken und enden heute bei den hierarchischen Speichersystemen und der Spiegelung der Daten mittels geeigneter Storage-Subsysteme.

Die Verfügbarkeit der Applikationen wird durch eine eigene Gruppe an Storage-Techniken erreicht. Der traditionelle Weg ist dabei, im Fehlerfall als ersten Schritt einen neuen Server aufzusetzen, um im zweiten Schritt die Datenbestände zu synchronisieren. Dieser Weg ist in jedem Fall mit einem Ausfall des entsprechenden Dienstes verbunden. Will man einen Ausfall der Dienste vermeiden, werden häufig **Cluster**² als probates Mittel eingesetzt. Hinzu kommen Techniken, die die Server-Last bei einem Fehlerfall auf andere Systeme transferieren.

Das Absichern der Daten und Programme in virtuellen Umgebungen kann der Anwender durch völlig unterschiedliche Techniken erreichen. Gleiches gilt auch für die Frequenz der Datensicherung. Da sich die Daten beständig ändern, muss deren Sicherung weitaus öfter erfolgen. Programme und Rechnerprozesse hingegen sind relativ statisch. Eine laufende Sicherung erscheint daher überflüssig.

Virtualisierung verknüpft Daten und Prozesse als eine Einheit

Auch wenn die beiden Aspekte - Sicherung der Daten gegen Verlust oder Änderung und Absicherung der Server-Prozesse gegen Ausfälle - auf den ersten Blick als zwei getrennte Anforderungen betrachtet werden können, sind sie es aber keineswegs. Denn in Wirklichkeit verhält es sich so wie mit Yin und Yang, dem Gegensatzpaar aus der chinesischen Philosophie: Das eine Element kann ohne das andere nicht existieren beziehungsweise jedes der beiden Elemente ist ohne das andere nutzlos. Von daher erscheint es ziemlich unverständlich, warum dieses Axiom in der IT so wenig beachtet wird.

Der Grund dafür liegt in der Regel in der erwähnten Dynamik der Daten beziehungsweise der Statik der Applikationen. Warum sollte man einen Applikations-Server mit seinen Programmen regelmäßig sichern, wenn diese Informationen sich kaum ändern?

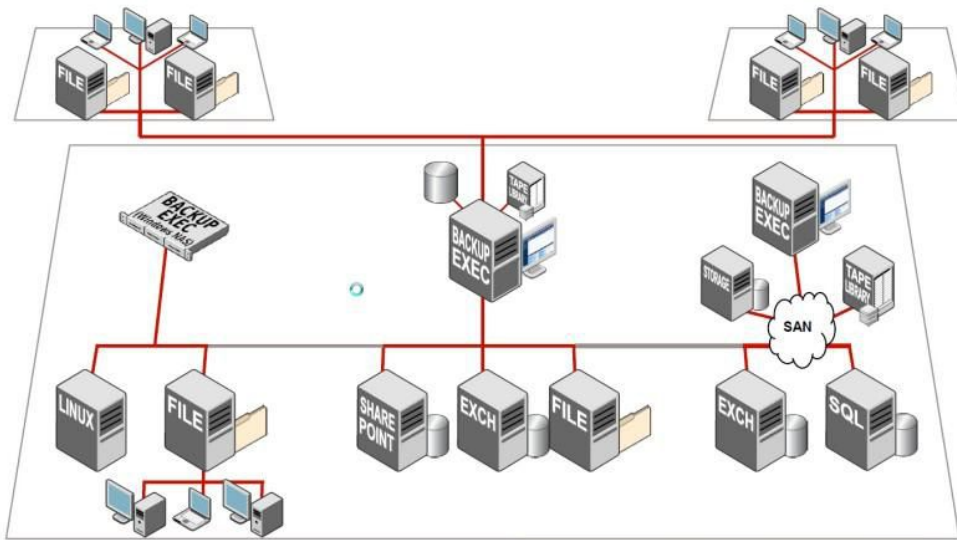
Die Technologie der Virtualisierung verändert den Blick auf die gesamte IT und die entsprechenden Konzepte. Besonders die Sicherung von Daten und Prozessen wächst in diesem Zusammenhang mehr und mehr zusammen.

Im Rahmen einer kleinen Artikelreihe werden daher beide Elemente betrachtet. In diesem Beitrag geht es um die Daten sowie um deren Sicherung und Wiederherstellung in virtuellen Umgebungen. In einem weiteren Artikel gehen wir dann auf die Absicherung der Prozesse beziehungsweise der Applikationen und das **Disaster Recovery**³ ein. Dabei werden auch die Unterschiede zwischen den virtuellen und physischen Umgebungen eine Rolle spielen.

Der traditionelle Weg der agentenbasierten Sicherung

Beim traditionellen Weg der Datensicherung werden nur die Daten gesichert. Dabei erzeugen Agenten in den Zielsystemen Kopien der Daten und senden diese an einen zentralen Backup-Server, der die Daten auf die Backup-Medien, meist Bänder oder ein Virtual Tape Libraries (VTL), sichert. Diese Datensicherung erfolgt in planmäßigen und regelmäßigen Abständen, meist nachts oder am Wochenende.

Nach einem initialen Full-Backup begrenzen differentielle und inkrementelle Methoden das Sicherungsvolumen. Das Verfahren wird meist ausschließlich für die Daten angewandt. Da eine virtuelle Maschine ja ein Abbild eines physischen Servers ist, kann das Verfahren natürlich auch in virtuellen Umgebungen eingesetzt werden. Dies macht es in der Migrationsphase einfach und erspart die Neuanschaffung von Backup-Software oder -Hardware, die auf die Besonderheiten der virtuellen Systeme ausgelegt ist.



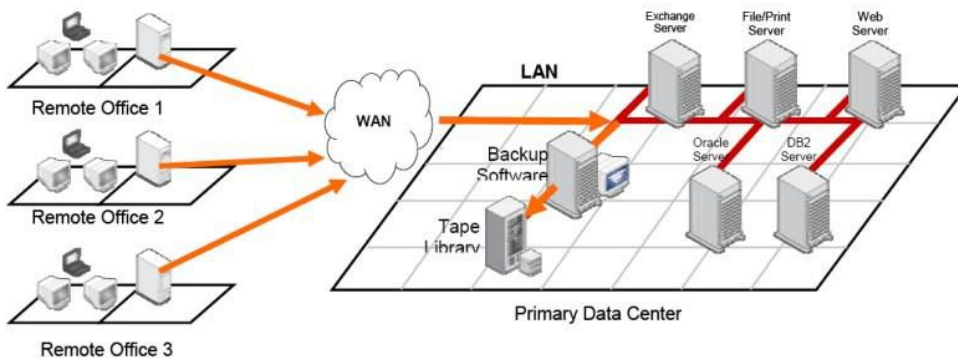
Kleine Spione: Backup Exec von Symantec basiert auf lokalen Agenten zur Datensicherung. (Quelle: Symantec)

Bei einem Fehler des Produktiv-Servers wird dieser - oder eben die virtuelle Maschine - neu gestartet und mit den notwendigen Daten des Backups versorgt. Diese Form der Sicherung ist etabliert und erprobt. Allerdings bringt der Restore - wenn nicht geübt - oftmals Überraschungen in Form von leeren Bändern oder falschen Inhalten mit sich. Darüber hinaus nimmt das Einspielen der Inkremente zu viel Zeit in Anspruch.

Der Prozess die Daten wieder herzustellen, kann je nach Systemumgebung Minuten, aber auch Tage dauern. Doch in modernen IT-Szenarien mit 7 x 24 Stunden Geschäftsbetrieb, Home Office mit freier Zeiteinteilung oder weltweiter Internetpräsenz geraten diese Konzepte mehr und mehr unter Druck. Eine Absicherung gegen den Serverausfall bietet es nicht. Aufgrund dieser Einschränkungen ist das beschriebene Backup-Verfahren für virtuelle Instanzen nicht die erste Wahl.

Besonderheiten bei virtuellen Systemen

Die Besonderheiten bei virtuellen Systemen in Hinblick auf die Datensicherung liegen gerade in den Vorteilen der Virtualisierung. So ist ein zentrales Ziel der Virtualisierung, die Auslastung der Server-Hardware zu erhöhen. Ein Host, der dank Virtualisierung allerdings zu 90 Prozent ausgelastet ist, hat kaum mehr Reserven für Backups, Server hingegen, die nur mit 15 Prozent CPU-Auslastung laufen, haben genügend.



Auf Nummer sicher: Die Absicherung der Außenstellen erfolgt durch spezielle Backup-Tools und -Verfahren. (Quelle: Veritas)

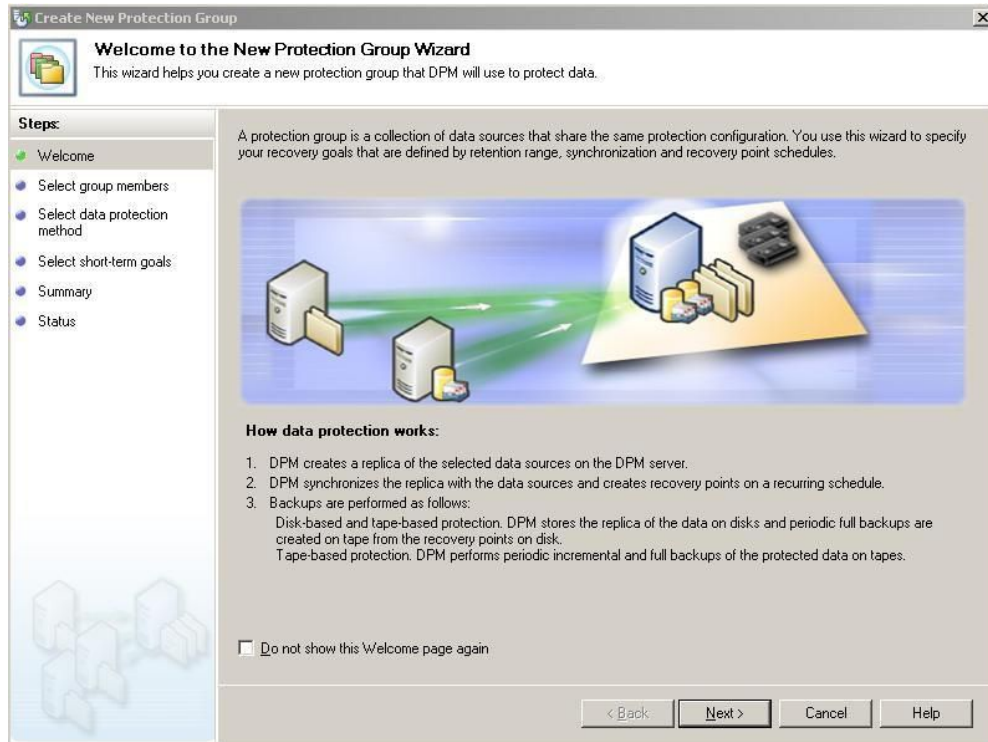
Bei den Zugriffen auf die Datenpfade und die Netzwerkpfade verhält es sich meist ähnlich. Da alle Daten in dem Szenario über das Netzwerk zum zentralen Backup-Server geschaufelt werden müssen, wird auch dieser häufig zum Nadelöhr. Einen weiteren Engpass der nächtlichen Sicherungsläufe stellt das Backup-Windows dar, also das Zeitfenster, in dem die Backup-Läufe erfolgen. Dieses wird für viele Unternehmen häufig zu klein bemessen, und bei Unternehmen, die im 7-x-24-Betrieb operieren, entfällt mitunter gänzlich.

Continuous Data Protection befreit vom Backup-Zeitfenster

Eine Lösung für das Dilemma des zu kleinen Backup-Fensters sind die Verfahren der Continuous Data Protection (CDP). Diese Sicherungsverfahren haben mit den herkömmlichen Backup-Konzepten so gut wie keine Gemeinsamkeiten. Dies beginnt bei der Frequenz der Sicherung. Meist liegen ihr weitaus kürzere Intervalle zugrunde. Dies können Stunden oder auch noch geringere Zeitabstände sein.

Änderungen ergeben sich auch beim Sicherungsverfahren. Statt wie bisher über das Dateisystem zu operieren, wird hierbei meist die Snapshot Technik angewandt. Auch die Sicherungsmedien sind anders als beim traditionellen Backup. CDP verwendet in der Regel Festplatten als Sicherungsmedium. Diese können über beliebige IP-Strecken angebunden werden.

Die Restore-Funktionen laufen bei den CDP-Verfahren ebenfalls anders ab. Die Wiederherstellung von Dateien kann der Anwender meist selbst direkt vornehmen. Sie erfordert keine Unterstützung der Administratoren. Komplexere Datenbestände wie eine Datenbank müssen aber auch bei CDP durch den Administrator wiederhergestellt werden.



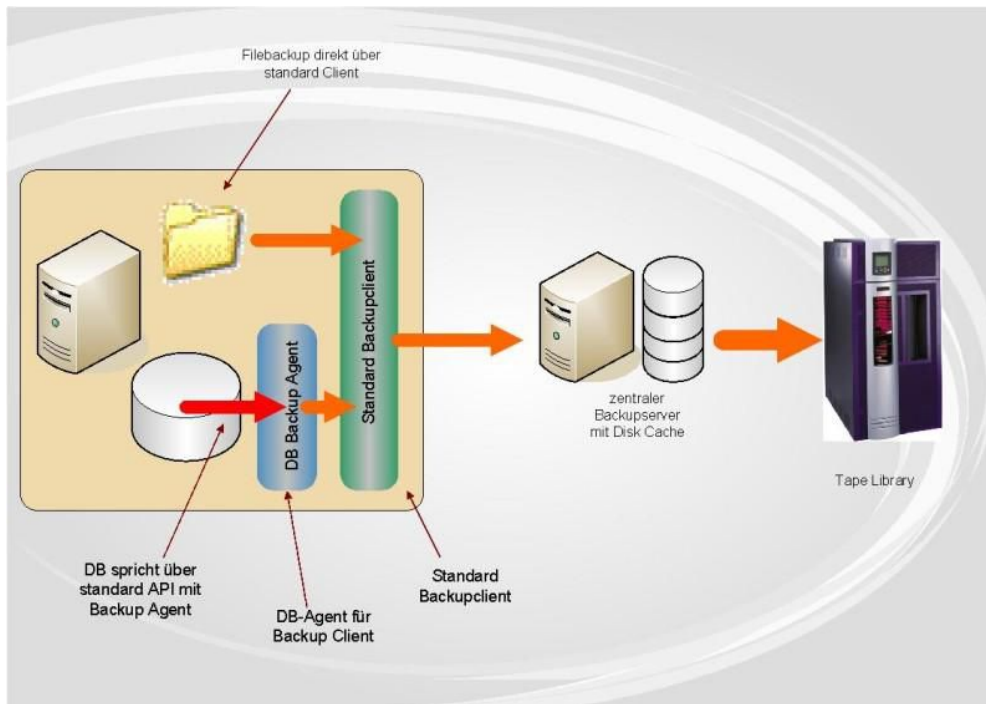
Dienstleistung: Der Data Protection Manager sorgt für eine laufende Sicherung der Daten nach dem CDP-Verfahren.

Die CDP-Sicherung kann auch in virtuellen Umgebungen eingesetzt werden. Allerdings erfordern die speziellen Sicherungen in virtuellen Umgebungen in jedem Fall die entsprechenden Tools und das Know-how der Administratoren. Die kontinuierliche Sicherung nivelliert die Last für die beteiligten Komponenten. Sie erzeugen dafür aber ein Grundrauschen an Last für alle beteiligten Baugruppen. Dazu gehören die CPU, das Storage-System durch die Speicherzugriffe und die entsprechenden Netzwerke. Da der Host den virtuellen Maschinen diese Ressourcen anteilig zur Verfügung stellen muss, wird er durch die CDP-Prozesse zusätzlich belastet.

Das Problem der Daten- und Applikations-Konsistenz

Bei der Sicherung von Daten muss immer deren Konsistenz beachtet werden. Dateien im Dateisystem müssen geschlossen sein, das Datenhaltungssystem muss in einem konsistenten Zustand gebracht werden, und die Puffer müssen geleert werden.

Beim Backup von Datenbanken und ähnlichen komplexen Datenstrukturen ist im besonderen Maße auf deren Integrität zu achten. Typischerweise bringen Datenbanken große Datenmengen mit sich, die dazu führen, dass der Backup-Vorgang entsprechend lange dauert. Zur selben Zeit finden jedoch Änderungen auf der Datenbank statt, die mit hoher Wahrscheinlichkeit dazu führen, dass Datenbereiche, die bereits auf das Backup-Medium kopiert sind, geändert werden. Dagegen enthalten andere Bereiche der Datenbank, die noch nicht gesichert wurden, bereits die aktuellen gültigen Metadaten. Ein solches Backup wäre nutzlos und könnte nicht zu einer funktionierenden Datenbank wiederhergestellt werden.



Backup als Dienst: BT bietet einen Backup-Dienst an, der die Unternehmen vollständig von der Datensicherung entlastet. (Quelle: BT)

Um dem entgegenzuwirken, haben die Datenbankhersteller Online-Backup-Schnittstellen entwickelt. Zusammen mit Schnittstellen zur Backup-Client-Software ist es so möglich, die Daten direkt von der Datenbank auf das Backup-Medium zu sichern und gleichzeitig alle Daten mitzusichern, die notwendig sind, um eine Konsistenz im Backup zu gewährleisten.

Applikationsintegration für Backup-Läufe

Die Integration von Applikationen in Backup-Läufe funktioniert nicht ohne die jeweiligen Applikationen oder das Betriebssystem. Bei der nächtlichen Sicherung mit Backup-Fenster hilft man sich mitunter damit, dass in der Zeit des Backup-Laufs die Applikation gestoppt wird. Beim **CDP-Verfahren**⁴ ist dies nicht mehr möglich. Folglich müssen sich die CDP-Varianten mit den Applikationen oder dem Betriebssystem abstimmen. Diese Interaktion gibt es nicht für alle Applikationen, daher scheidet CDP für diese Anwendungen, die durch das CDP-Tool nicht unterstützt werden, gänzlich aus.

All diese bis dato genannten Verfahren wurden ursprünglich in physischen Umgebungen eingesetzt, sind aber prinzipiell auch in virtuellen Umgebungen anwendbar. Sie beruhen auf Agenten in den zu sichernden Systemen. Neben den erwähnten Besonderheiten müssen aber noch weitere Aspekte beachtet werden. Dies betrifft die Aktualisierung der Agenten und deren Kommunikation mit ihren Backup Server.

Virtuelle Maschinen zeichnen sich dadurch aus, dass sie häufig auch deaktiviert, suspendiert oder verschoben werden können. Dies hat natürlich auch weitreichende Auswirkungen auf die Backup-Jobs. Denn ein Backup-Job, der einen Server periodisch sichern soll, erwartet diesen aber aktiv im Netz. Die große Dynamik der Virtualisierung kann in diesem Fall leicht zu Verwirrungen führen.

Sicherung durch den Host

Die Sicherung von Systemen durch Agenten in den virtuellen Umgebungen ist machbar, wird meist aber nicht empfohlen. Stattdessen geht man dazu über, die Sicherungen der virtuellen Maschinen durch den Host zu bewerkstelligen. Dabei erfolgt der Zugriff auf das Host-Betriebssystem durch das Backup-Tool, also durch die Virtualisierungs-Engine, wie den ESX-Server, den Hyper-V oder den XenServer. Nutzbar sind aber auch Virtual Appliances, die direkt als fertige virtuelle Maschinen auf den Host gelegt werden und dann die Sicherung der weiteren VMs übernehmen.

Der Vorteil dieser Backup-Varianten besteht in einer fertig abgeschlossenen und vorkonfigurierten Backup-Hilfe, die direkt und einfach nutzbar ist. Über die Virtualisierungsschicht erfolgt dann der Zugriff auf die virtuellen Maschinen. Dabei ändert sich allerdings auch das Sicherungsobjekt. Nunmehr erfolgt nicht mehr die Sicherung von ausgewählten Benutzerdaten im Dateisystem, einer Datenbank oder eines Exchange-Mailstores, sondern es wird stattdessen die virtuelle Maschine wie ein Snapshot gesichert.

Dies ist auch der Übergang von der Sicherung der Daten hin zur Sicherung der Applikationsprozesse beziehungsweise des Server-Images. Damit ändern sich aber auch alle Verfahren des Disaster Recoverys und der Wiederherstellung eines Server-Dienstes samt seiner Daten. Nun genügt es, nur die virtuelle Maschine wiederherzustellen. Sie umfasst im Idealfall neben den Daten auch die Programme. Das separate Aufsetzen eines Applikation-Servers entfällt dabei. Die Sicherung der virtuellen Maschine durch den Host hat damit eine Reihe von Vorzügen - aber auch Nachteile. Diese liegen erneut in der Last, die auf dem Host erzeugt wird. Aber auch um die Konsistenz der Daten muss man sich beim Host-basierten Backup kümmern.

Snapshots vereinfachen die Wartung?

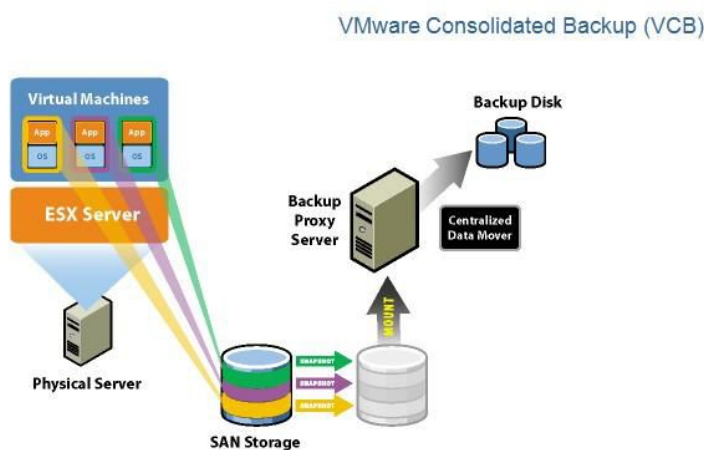
Bei der Sicherung durch Snapshots müssen einige Fallstricke beachtet werden. Wenn der Snapshot nur die Deltas zu dem vorherigen Full Backup oder vollständigen Snapshot umfasst, operiert dieser im Prinzip wie ein inkrementelles Backup. Der Snapshot ist damit, so wie die Inkremente des Backups, allein nicht zu gebrauchen, vielmehr kann er nur im Kontext der vorherigen durchgeführten Sicherungen eingesetzt werden.

Das Betriebssystem Windows speichert beispielsweise bei Snapshots nur die Änderungen (Deltas) einer virtuellen Maschine in einer Differencing Disk. Die eigentliche virtuelle Maschine gehört nicht zum Umfang des Snapshot. Diese Snapshot sind daher nicht eigenständig verwendbar oder lauffähig, sondern können immer nur in Verbindung mit ihrer eigentlichen virtuellen Maschine benutzt werden.

Der Anwendungszweck dieser Snapshot liegt in der Wiederherstellung eines früheren Zustands einer virtuellen Maschine, wie etwa nach einer Änderung der Konfiguration. Als Backup-Ersatz sind diese Snapshots nicht zu verwenden. Beachtet werden muss aber auch, dass sich die Funktionen der Snapshots zwischen den Herstellern unterscheiden und es Snapshot-Techniken gibt, die den Bezug zu den Basissicherungen selbstständig herstellen.

VMware Consolidated Backup entlastet den ESX-Host

Mit VMware Consolidated Backup stellte VMware eine Möglichkeit zur Datensicherung vor, die einen Großteil der Probleme und Engpässe der bisherigen Verfahren löst. Das Verfahren verlangt allerdings, dass die Daten (vmdk) an einen Proxy Server (VCB Proxy) gemountet werden. Der VCB Proxy Server trägt dann die Last der Sicherung, und zusätzlich wird das Netzwerk entlastet, da die zu sichernden Daten nicht über das LAN transferiert werden müssen.

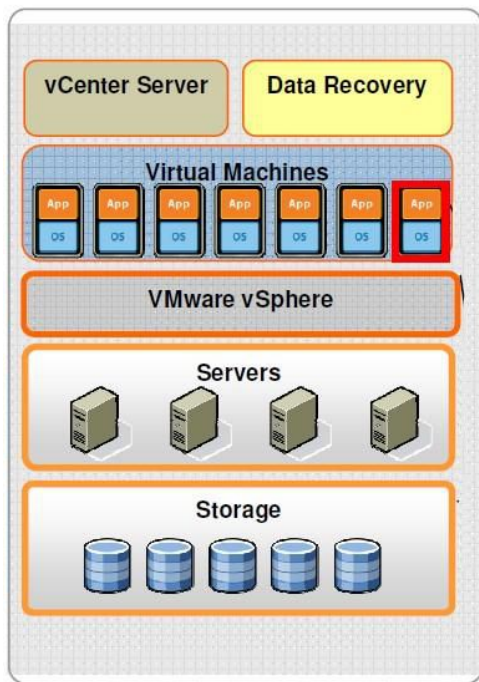


VCB-Verfahren: Bei VMware Consolidated Backup übernimmt der Host die Sicherung der Daten. (Quelle: VMware)

Diese Technik ermöglicht sowohl die Sicherung einer vollständigen virtuellen Maschinen (Full Backup) und für Windows VMs als auch die Sicherung einzelner Dateien im File Level Backup. Somit wird die Problematik des begrenzten Backup-Zeitfensters mittels dieser Technik relativiert, da der erste Schritt zum Vorbereiten des Backups ein Snapshot der VM mit sich bringt und anschließend das vmdk gesichert werden kann; währenddessen "läuft" die VM weiter. Die Wiederherstellung läuft dabei übers Netzwerk.

Backup-Software sorgt für Applikationskonsistenz

Trotz aller Vorzüge des VCBs bleibt weiterhin ein Wermutstropfen. Auch die Sicherung mit VCB sorgt nicht für die geforderte Konsistenz aus der Sicht der Applikation beziehungsweise deren Daten. Allerdings ist aus der Warte des Betriebssystems die Konsistenz gegeben. Daraus folgt, dass man weiterhin die Integration von Backup-Software braucht, welche die notwendigen Agenten für eine konsistente Sicherung aus Applikationssicht mitbringt. Beim Hochfahren oder Wiederanlaufen der virtuellen Maschine muss daher diese Übereinstimmung (Konsistenz) durch einen Restore der Daten wiederhergestellt werden.

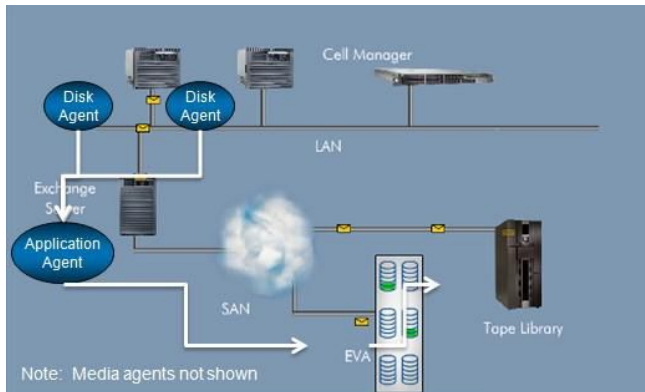


Zusatzdienste: VMware liefert mit dem Data Recovery nun auch eine eigene Backup-Software. (Quelle: VMware)

Zusammen mit der Hinwendung zur Cloud und der Umbenennung der Virtualisierungsprodukte bringt der Hersteller **VMware**⁵ eine neue Backup-Schnittstelle, die vStorage API for Data Protection. In **vSphere**⁶ besteht damit eine weitere Methode zur Sicherung von Daten in virtuellen Umgebungen. Die bisherige Technik über das VMware Consolidated Backup bleibt bestehen und wird auch in vSphere unterstützt. Die neue Methode operiert über die vStorage API for Data Protection. Das Tool dazu nennt **VMware Data Recovery**⁷.

Zero Downtime Backup

Um auch die letzte Hürde, die Inkonsistenz der Daten, zu vermeiden, muss eine Synchronisation der Backup-Prozesse mit der Applikation erfolgen. In dem als Zero Downtime bezeichneten Verfahren geht zum Beispiel **HP**⁸ bei seinen Speichersystemen EVA daher den folgenden Weg:



Kleine Helfer: HP erreicht durch die Integration des Backup-Verfahrens (Agenten) in die Applikationssysteme ein Zero Downtime Backup. Die Sicherung erfolgt dabei auf das Speichersubsystem EVA. (Quelle: HP)

Die Sicherungsprozesse der virtuellen Maschinen laufen direkt im Speichersubsystem der EVA. Damit ist der **ESX-Host**⁹ von VMware von diesem Prozess entlastet. Das macht allerdings auch VCB. Zusätzlich aber erfolgt bei Zero Downtime Backup eine Synchronisation des Backup-Prozesses mit der Applikation. Die Applikation bleibt damit verfügbar und muss nicht deaktiviert werden. Durchgeführt werden diese Kopierprozesse in Minutenintervalle. Damit wird der Verlust im Fehlerfall gering gehalten. Um bei dieser hohen Frequenz der Sicherung die Speichersystem nicht zu überlasten, werden diese gesicherten Snapshots auf VTLs hinterlegt. Ferner ist auch eine Sicherung auf Band oder Festplatten möglich.

Datenreduktion durch Deduplizierung

Um die Menge der zu sichernden Daten zu reduzieren, werden die Techniken der **Deduplizierung**¹⁰ eingesetzt. Hierbei wird versucht, doppelte Inhalte nur einmal zu sichern. Die weiteren Verweise auf die Daten werden durch Platzhalter abgebildet. Die Deduplikation unterscheidet prinzipiell zwei Verfahren, das Single-Instancing und eine blockbasierte Variante.

DateId	CarrierTracking	OfferID	PriceDisc
20070601	6431-403C-98	10	0.00
20070601	6431-403C-99	10	0.00
20070602	6431	10	0.00
20070602	6431-4D57-83	20	0.00
20070602	6431-4D57-84	20	0.00
20070602	6431-4D57-85	10	100.00
20070603	4E0A-4F89-AE	10	0.00



DateId	CarrierTracking	OfferID	PriceDisc
20070601	6431-403C-98	10	0.00
20070601	6431-403C-99	10	0.00
20070602	6431	10	0.00
20070602	6431-4D57-83	20	0.00
20070602	6431-4D57-84	20	0.00
20070602	6431-4D57-85	10	100.00
20070603	4E0A-4F89-AE	10	0.00

Schrumpfkur: Der SQL Server 2008 erlaubt die Kompression der Daten direkt in der Datenbank und reduziert somit das Backup-Volumen.

Beim Single-Instancing wird eine Datei immer als Einheit betrachtet. Es eignet sich daher vor allem für identische Mail-Anhänge oder mehrfach verwendete Dokumente im Dateisystem. Wird eine Daten aber nur geringfügig verändert, so stellt diese für das Single-Instancing ein separate Instanz dar, und die Veränderung führt zu einer zweiten Ablage der kompletten Datei. Diesen Nachteil will man durch die blockbasierte Variante der Deduplikation vermeiden. Hierbei betrachtet man nicht Dateien, sondern Datenblöcke. Deren Größe variiert nach Hersteller, in der Regel werden diese aber immer weitaus kleiner sein als die der Dateien.

Fazit

Die Sicherung von Daten durch Backup-Prozesse ist in virtuellen Umgebungen genauso wichtig wie in der rein physischen Welt. Durch die Verknüpfung der Daten und Applikationen in einem Image der virtuellen Maschine bieten sich aber andere Möglichkeiten. Dies bleibt nicht ohne Auswirkungen auf Techniken, Tools und Prozesse. (hal)

Links im Artikel:

- ¹ https://www.tecchannel.de/storage/extra/401665/raid_sicherheit_level_server_storage_performance_festplatten_controller/
- ² https://www.tecchannel.de/server/hardware/456463/hochverfuegbarkeit_cluster_contra_server/
- ³ https://www.tecchannel.de/storage/backup/2020371/ratgeber_backup_restore_disaster_recovery/
- ⁴ https://www.tecchannel.de/storage/backup/442468/continuous_data_protection_backup_luecken_schliessen/
- ⁵ <http://www.vmware.com/de/>
- ⁶ https://www.tecchannel.de/server/news/2018387/vmware_stellt_esx_nachfolger_vsphere_4_vor/
- ⁷ <http://www.vmware.com/products/data-recovery/>
- ⁸ <http://www.hp.de/>
- ⁹ https://www.tecchannel.de/server/virtualisierung/432353/workshop_vmware_esx_server/
- ¹⁰ https://www.tecchannel.de/storage/backup/1993739/deduplizierung_mehr_sichern_weniger_speichern_datensicherung/