



StorMagic verlaagt kosten shared storage substantieel

Virtual appliance is goed alternatief voor SAN

Met behulp van virtualisatie zijn tegenwoordig allerlei zogenaemde virtual appliances te creëren. Bram Dons bespreekt SvSAN, een storageproduct van het in 2006 opgerichte StorMagic.

In tegenstelling tot wat men soms denkt, wordt servervirtualisatie nog lang niet overal toegepast. Slechts 16 procent van alle werkbelastingen draait op virtual machines (VMs). Maar volgens Gartner groeit de markt voor virtualisatie wel snel en de analisten voorspellen dat tegen eind 2012 ongeveer 50 procent van de servers op basis van de x86-architectuur van virtualisatie gebruik zal maken. Tijdens een recent symposium signaleerde Gartner dat de small business-markt, bij ons MKB-markt genoemd, momenteel de snelstgroeiende sector is. Het is dan ook dit marktsegment waarop de firma StorMagic mikt met haar SvSAN virtuele SAN-oplossing.

De intrede van virtualisatietechnologie in de wereld van open systemen heeft, naast serverconsolidatie, een aantal interessante ontwikkelingen met zich meegebracht. Tegenwoordig kunnen we niet alleen een complete fysieke server nabootsen in de vorm van een virtuele machine, maar ook andere op zichzelf staande systeemcomponenten, zoals netwerkswiches, storagesystemen en appliances. Als we bijvoorbeeld op de VMware Appliance

Marketplace kijken, zien we dat er al meer dan 1.200 virtual appliances voor de VMware-omgeving zijn ontwikkeld. Allemaal keurig gerangschikt in verschillende categorieën, zoals netwerken, ERP, CRM, performance, systeembeheer, enzovoorts. Ook in de categorie 'storage' zien we zo'n dertig verschillende toepassingen, waaronder: VTL, CDP, NSS, SSH en back-up. Dat bracht de firma StorMagic op het idee om een shared storage-omgeving voor de ESX Server clusteromgeving te creëren op basis van een virtual appliance. Het door StorMagic ontwikkelde product, Storage Virtual Appliance (SvSAN) geheten, is voornamelijk bestemd voor kleine ondernemingen die zich geen relatief dure SAN kunnen veroorloven. De onderneming hoeft met de toepassing van SvSAN geen fysieke iSCSI, FC of NFS shared storage meer te implementeren om daarmee toch een VMware-cluster te kunnen verwezenlijken. Er wordt van lokale disk storage gebruikgemaakt in plaats van externe shared storage. SvSAN converteert de interne VMware-server disk storage naar een virtuele SAN. Hoe dat in praktijk uitpakt, zien we hierna.

Features

SvSAN is een storage virtual appliance (SVA) die de direct aan een VMware ESX Server gekoppelde, lokale disk drives (DAS) beschikbaar maakt voor toepassing in een shared SAN. Dit maakt het mogelijk om de VMware features VMotion en High Availability (HA) toe te passen zonder dat daarvoor een externe shared SAN op basis van iSCSI, Fibre Channel of NFS nodig is. Alle door SvSAN gecreëerde VMware datastores worden vanuit een centraal managementconsole beheerd en zijn volledig binnen vCenter geïntegreerd. SvSAN is gebaseerd op de iSCSI-standaard en het door VMware ondersteunde SCSI-2 Reservationsprotocol, op basis waarvan een interne disk via het netwerk beschikbaar wordt gesteld als shared storage. Verder ondersteunt het synchrone mirroring tussen meerdere datastores van SVAs die op verschillende ESX Servers draaien, of in combinatie met een externe StorMagic appliance. Voor beide is overigens een aparte licentie nodig. SvSAN maakt een automatische configuratie van gespiegelde disk drives mogelijk voor de ondersteuning van failover van VMware datastores. De op SCSI/iSCSI gebaseerde SVA-omgeving ondersteunt de standaard RFC 3720 Authentication met one-way/mutual CHAP en iSNS/ACLs. Per virtual appliance worden 256 target en 1.024 gelijktijdige sessies ondersteund. SvSAN ondersteunt 256 snapshots per SVA en de VSS Provider voor Win-

dows. De hoeveelheid ondersteunde datastore is onbeperkt. Een basislicentie ondersteunt 2TB en optioneel 4 en 8 TB tot een onbeperkte hoeveelheid storage. Op elke ESX Server binnen het SAN-cluster wordt de StorMagic SvSAN software geïnstalleerd. Daarvoor volstaat een virtuele cpu met 2GHz, Gigabitverbinding, 500 MB diskruimte en 20 GB diskruimte bij gebruik van de HA-optie.

Gebruik van interne disks

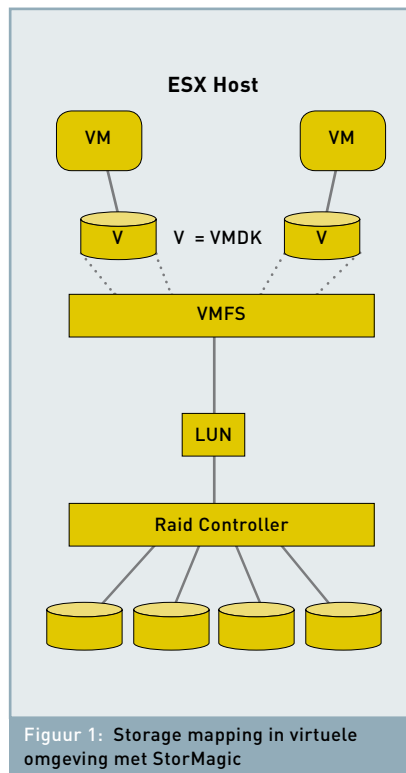
De storage virtual appliance transformeert interne disks naar een virtuele SAN. Bij een optimale configuratie worden alle interne disks in de ESX Server aan het SVA toegewezen, met uitzondering van het device waar vanaf de ESX Server wordt geboot. De laatstgenoemde bevat ook de datastore om de SVA zelf te booten. Een ESX-Server met een hardware RAID controller, waarop een aantal disk drives zijn aangesloten, bijvoorbeeld, kan in de RAID BIOS als ESX boot drive worden geconfigureerd. De overgebleven disks blijven dan unassigned, zodat bij de initiële configuratie van de SVA deze als RAID 5 of 6 automa-

DE MIRRORING FEATURE BESCHERMT TEGEN UITVAL VAN DE SHARED STORAGE

tisch aan de SVA als een storage pool kunnen worden toegevoegd. De RAID hardware 3Ware 96xx, Intel SRCxx, en alle LSI MegaRAID SAS controllers, worden ondersteund met RAID 0, 1, 5 en 6. Als de server geen van de genoemde controllers bevat, kunnen ook virtuele disken (VMDKs) in een storage pool worden toegepast. In figuur 1 zien we schematisch weergegeven hoe, met behulp van de op de ESX-Server geïnstalleerde StorMagic-software, een via een RAID controller verbonden fysieke drive wordt gekoppeld aan een storage pool in een SAN.

Mirroring

Net zoals bij alle andere toepassingen van shared storage in een clusteromgeving vormt ook de shared storage binnen VMware's clusteromgeving een



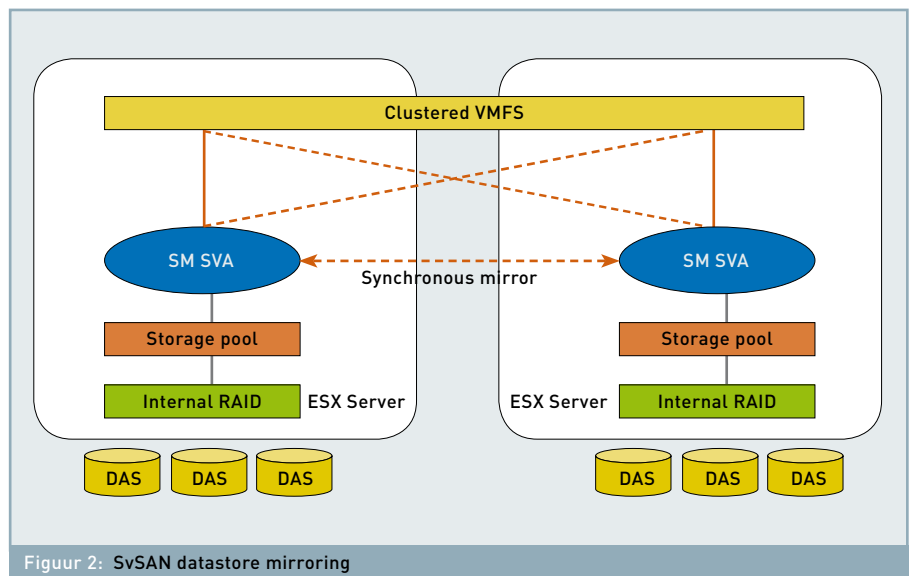
Figuur 1: Storage mapping in virtuele omgeving met StorMagic

single point of failure (spof). Bij uitval zijn de VMs immers niet meer beschikbaar. Alle reden dus om binnen de ESX-Server-omgeving een voorziening op te nemen die deze spof opheft. Een van de meest gebruikte methoden, afgekeken uit de enterprisewereld, is synchroon mirroring van disk volumes, waarbij de mirroring meestal plaatsvindt vanuit de firmware van de storage array. Het principe van de StorMagic's SvSAN mirror is daar ook op gebaseerd, alleen gebeurt dit nu door middel van soft-

ware die binnen een virtual appliance draait. Zoals we zagen, een eenvoudige SvSAN datastore maakt gebruik van de interne disk(s) van een ESX Server. Dit impliceert dat als de server of het locale disksubstelsysteem uitvalt, de datastore niet meer beschikbaar is. Door nu de datastore te spiegelen naar een of meer SVAs, die op verschillende andere fysieke ESX Servers draaien, kan binnen een VMware-omgeving deze storage spof worden opgeheven. Elke gespiegelde disk heet in StorMagic-termen een mirror 'plex'. Als een van de servers uitvalt, dan behouden de andere actieve clusternodes evengoed via hun SVA toegang tot de datastore. Elke SvSAN gebruikt opslagcapaciteit van de interne storage pool om een mirror plex te huisvesten. Met het oog op redundantie is het aan te bevelen om hiervoor een RAID array te gebruiken, maar het is niet strikt noodzakelijk. Zelfs kan in een mirror plex een opslagsysteem gebruikt worden met verschillende prestatiekenmerken. In de hierna beschreven test wordt een SATA II disk met een SAS disk gespiegeld.

Synchrone replicatie

De SVA mirror werkt op basis van de veel toegepaste synchrone datareplacatiemethode, waarbij elk datablock dat op de host veranderd is eerst door beide kanten van de mirror moet worden geconformeerd voordat de data definitief op de host wordt geschreven. De synchronisatie van een mirror kan full of fast worden uitgevoerd. Bij full wordt de totale inhoud van de gesynchroniseerde plex gekopieerd naar de

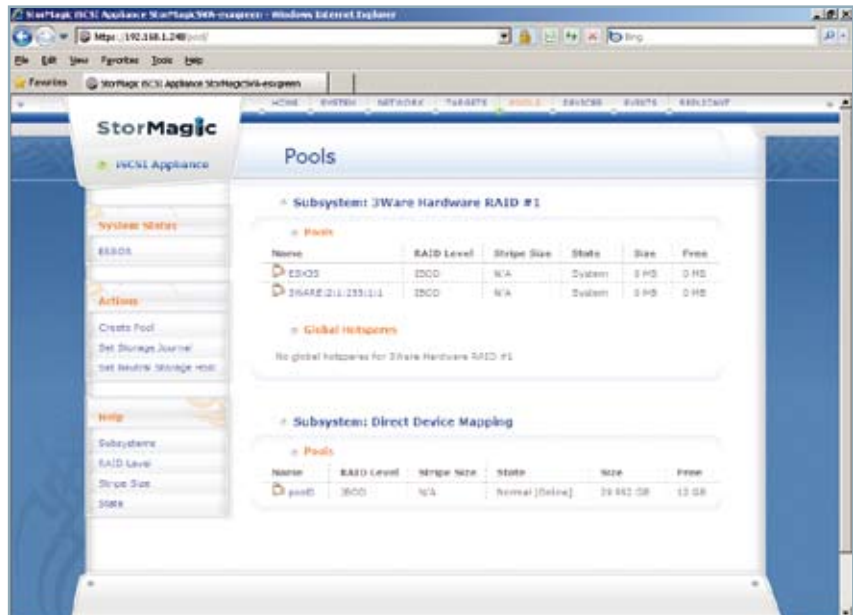


Figuur 2: SvSAN datastore mirroring

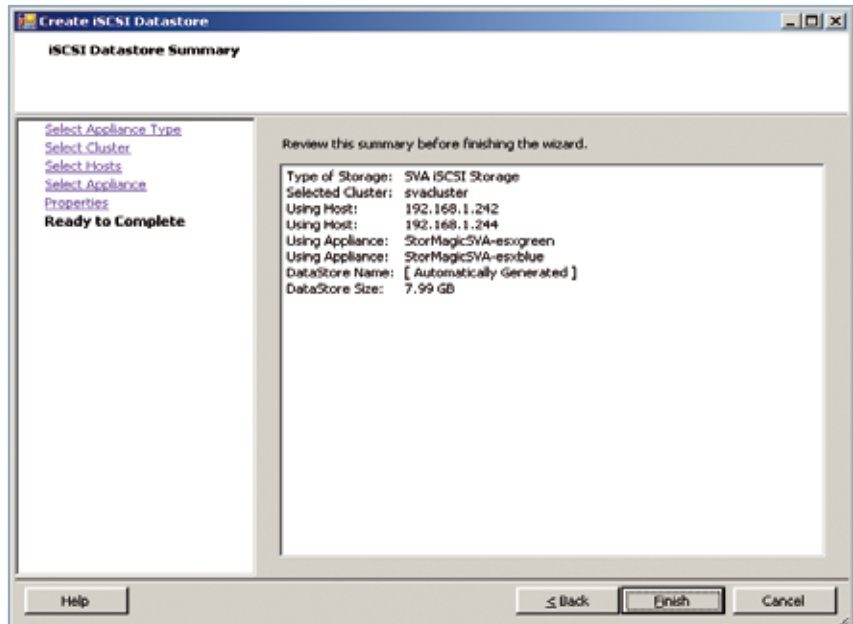
ongesynchroniseerde plex. Met fast houdt de overgebleven plex de veranderde data bij en kopieert tijdens de hersynchronisatie alleen deze wijzigingen. SvSA ondersteunt fast hersynchronisatie. Dat betekent, als een plex slechts voor een korte tijd offline was, de mirror weer snel gesynchroniseerd is. Een volledige hersynchronisatie is nodig bij creatie van een mirror (of bij toevoeging van een plex), bij uitval van het storagestelsel, als een plex voor langere tijd offline is geweest, of als er een grote hoeveelheid data is gewijzigd op de overgebleven plex. Tijdens een hersynchronisatie tussen twee plexen kan er natuurlijk geen failover plaatsvinden en ontstaat een spofsituatie. Als de gesynchroniseerde overgebleven plex om de een of andere reden uitvalt, zal de mirror offline gaan en zal dat blijven, totdat de gesynchroniseerde plex weer online komt. Elke andere gebeurtenis zal leiden tot verlies van data, omdat per definitie een ongesynchroniseerde plex niet up-to-date is. De meest risicovolle momenten voor een mirror zijn dus wanneer deze ongesynchroniseerd is. Na de creatie van de mirror worden de mirror plexen gesynchroniseerd, zodat ze dezelfde data bevatten.

GRAAG NOG EEN STREAMER GRAAG NOG EEN STREAMER

Als een van de kanten van de server uitvalt, dan worden de veranderingen op de overgebleven plex bijgehouden. De uitgevallen plex is dan ongesynchroniseerd en zal bij het opstarten weer eerst een synchronisatieproces moeten doorlopen voordat de mirror weer beschermd is tegen uitval. Normaal bevatten alle mirror plexen dezelfde data, maar als gevolg van uitval van een netwerkverbinding met verlies van quorum, serveruitval of lokale opslagsysteem, kan een plex niet meer synchroon met de hoofdmirror raken. Om de betrouwbaarheid van een mirror te waarborgen, maakt StorMagic van een quorumstelsel gebruik, waarbij het merendeel akkoord moet gaan voordat een zogenoemde state-verandering plaatsvindt. Op elke SVA draait daartoe een Mirror State Daemon mirror die communiceert met de peer mirrors op alle



Figuur 3: Storage Pool menu



Figuur 4: SVA iSCSI Datastore

andere servers die deel uitmaken van een mirror. Er draait ook nog een additionele instance van een mirror state op een externe server. Deze fungeert als neutral storage en werkt als een tie breaker. Het draait als een Neutral Storage Daemon nsd en draait als een Windows Service op een Windowsstelsel. Binnen een VMware-omgeving is dit doorgaans het systeem waarop vCenter draait. Een gespiegelde SvSAN datastore wordt op dezelfde wijze gecreëerd als een normale datastore waarbij de ESX host via een iSCSI-verbinding inlogt. Bij een

gespiegelde datastore echter heeft de ESX Server meerdere toegangspaden tot de datastore. Als een kant van de mirror uitvalt, dan neemt ESX een path failure waar en maakt van haar path failover-mechanisme gebruik om naar het andere path over te schakelen. Multi-pathing is automatisch geconfigureerd door de SvSAN vCenter plug-in tijdens de creatie van een datastore.

Installatie en configuratie

Voorafgaande aan de installatie van de StorMagic-software op de ESX Servers dient de VMware iSCSI initiator voor

de ondersteuning van shared storage te zijn geactiveerd. Bovendien dient de firewall op de ESX Server voor de iSCSI-initiator te worden geopend. Als eerste installeren we de SM Manager Suite op de vCenter Server, waarna we vervolgens de Virtual Infrastructure Client opstarten en de SM Manager Suite Plug-in activeren. De StorMagic tab heeft twee tabbladen: vSAN en SVA. De vSAN view dient voor het beheren van de virtual SAN, tonen van de datastores, SVAs en fysieke appliances en de creatie en het verwijderen van iSCSI datastores. De SVA view dient voor het beheer van iSCSI storage. Na de configuratie van de Ethernetpoorten die SVA gebruikt voor iSCSI, een eventuele iSNS en NTP server, wordt als eerste een storage pool gecreëerd. De SVA heeft automatisch gedetecteerd of er fysieke disks aanwezig zijn en formatteert elke unassigned drive die als een RAID 0, 5 of 6 array kan worden gebruikt in een storage pool. Van daaruit kunnen iSCSI targets worden gecreëerd en als VMFS datastores worden gebruikt. Als de aan SVA toegewezen storage meer dan een VMDKs beslaat, is er sprake van een directe mapping, in plaats van een random disk mapping. Het enige type dat voor direct mapping in aanmerking komt is een JBOD.

Creatie iSCSI datastore

Na de installatie van SVA en configuratie van een storage pool, kan worden

begonnen met de creatie van nieuwe iSCSI datastores. Een datastore kan aan een of meer ESX hosts in een cluster worden toegewezen op basis waarvan naderhand VMs met VMotion kunnen worden gemigreerd. Via de 'Create New iSCSI Datastore' uit de vSAN iSCSI Datastores view wordt het proces gestart. Als eerste moet de naam van een cluster of host worden ingevoerd die toegang heeft tot de datastore en alle hosts in het cluster die shared access hebben. Dan wordt de maximale grootte van de datastore en het type: 'Single' of 'Mirrored' ingevoerd. De StorMagic SVA voert vervolgens een aantal taken uit. Het stelt de Access Control List samen, die de geselecteerde ESX Servers toegang verschaft tot de datastore.

SVSAN ONDERSTEUNT 256 SNAPSHOTS PER SVA

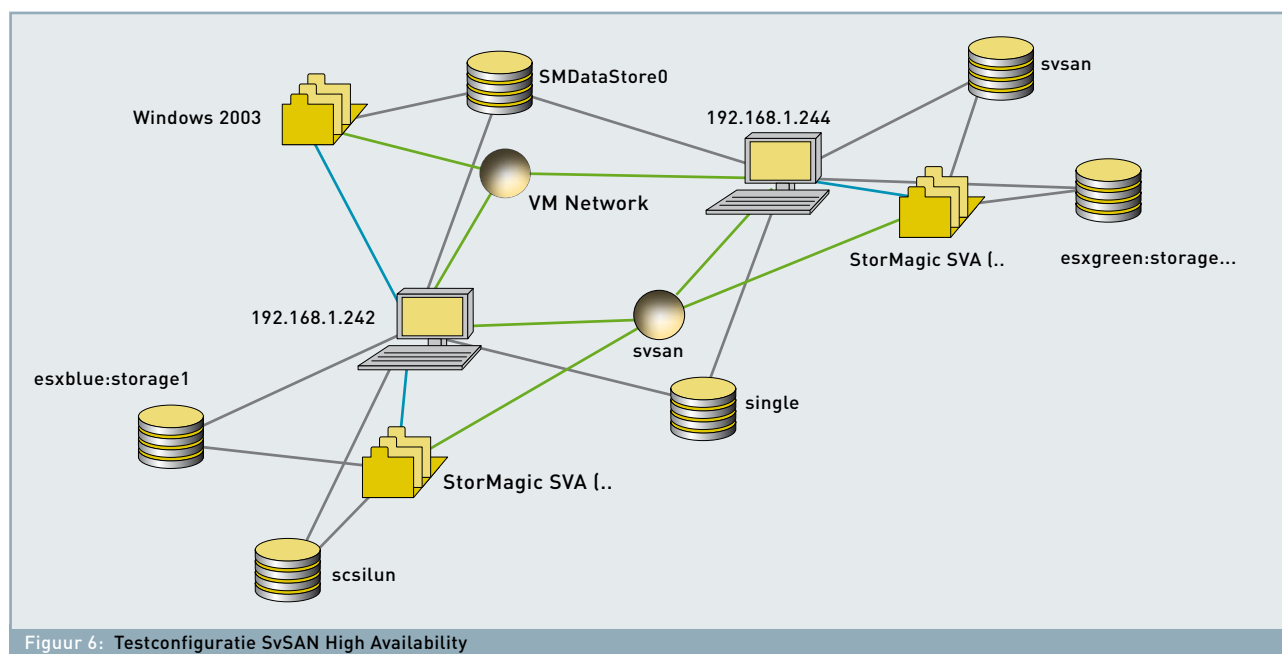
Daarna configureert hij deze, zodat het volume kan worden gebruikt, voert hij een iSCSI bus rescan commando uit ter verzekering dat de server het nieuwe volume herkent en registreert. De SVA formatteert het volume met VMFS en voert tenslotte een laatste rescan uit, zodat de geselecteerde hosts toegang krijgen tot de datastore. Met StorMagic SVA hoeven de ESX-configuratie tools niet te wor-

den gebruikt om een datastore in te stellen; dat doet de SVA dus automatisch.

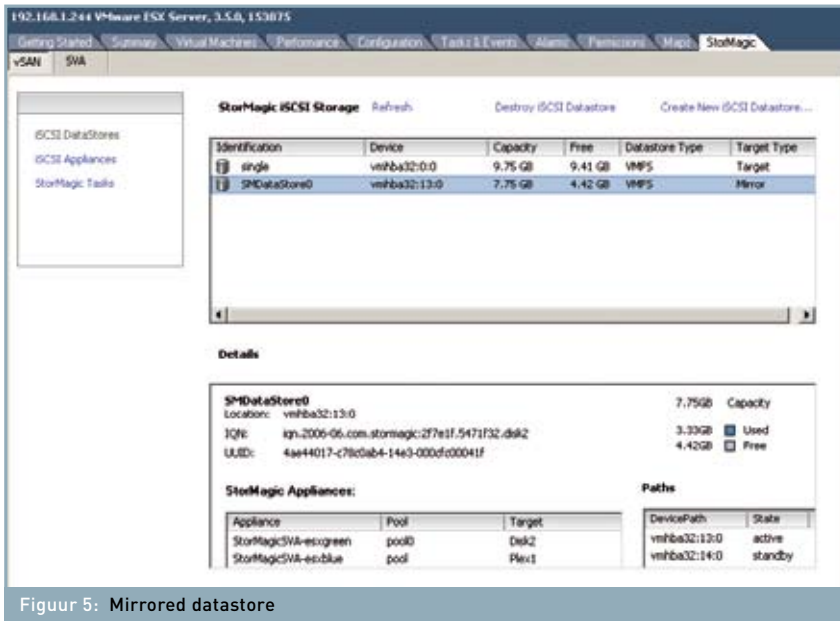
Testconfiguratie

Onze testomgeving bestaat uit twee ESX 3.5 Servers waarop elk een StorMagic SVA draait. Beide ESX Servers zijn via een mirrored shared storage 'SMDataStore0' verbonden. Voordat een mirrored datastore kan worden gecreëerd moet een neutral storage host op elke SVA worden gedefinieerd die deel van een mirror gaat uitmaken. De neutral storage host is de host waarop de StorMagic Suite wordt geïnstalleerd, dat normaal de vCenter host is. Bij selectie van een paar iSCSI appliances zal de datastore worden gespiegeld: een kopie van de datastore voor elke appliance. De SVAs zorgen er zelf voor dat de twee datastores identiek zijn, ook al verschillen de onderliggende disken in capaciteit, type en snelheid. We creëren een VM network, activeren VMware's VMotion, creëren een VM en installeren daarop een Windows 2003 Server.

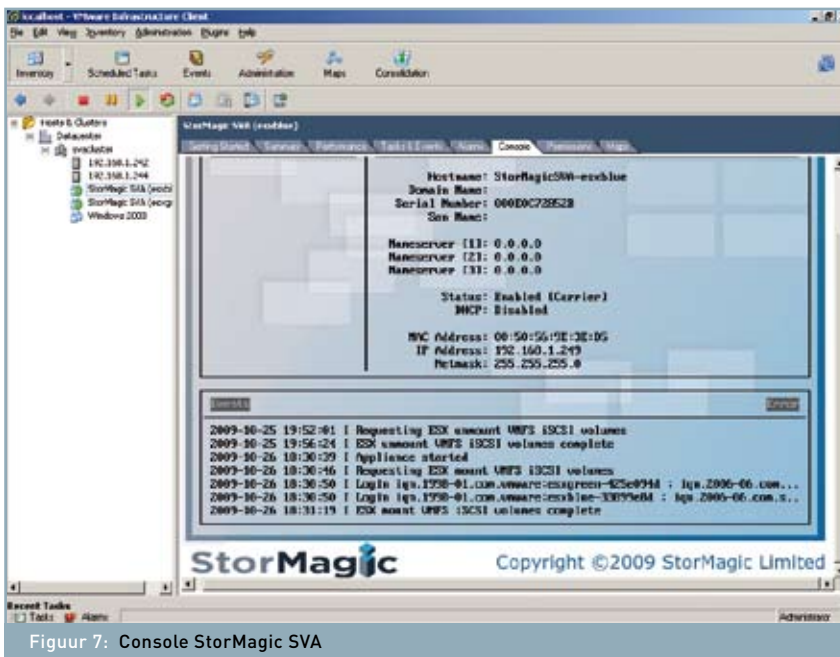
Als eerste testen we VMotion door de actief draaiende Windows 2003 Server van de ene naar de andere ESX Server te migreren. Voor het testen van de mirroringfunctie schakelen we een van de SVA's op de ESX Server uit. We zien dat de Windows Server en de daarop actieve applicaties keurig blijven doordraaien. Vervolgens bren-



Figuur 6: Testconfiguratie SvSAN High Availability



Figuur 5: Mirrored datastore



Figuur 7: Console StorMagic SVA

ONGESYNCHRONISEERDE MIRROR IS RISKANT

gen we de uitgeschakelde SVA weer online en wordt automatisch het synchronisatieproces opgestart. Na enige tijd is de mirror weer gesynchroniseerd en de spof van de shared storage opgeheven. Vervolgens activeren we VMware's High Availability-optie. Het idee achter HA is om bij uitval van een ESX Server over te schakelen naar een andere ESX Server. Tijdens de configuratie verscheen de bekende isolation address foutmelding. De VMware HA agent, wat in feite de

Legato Automated Availability Manager (AMM) is, maakt van de Service Console default gateway als zijn isolation address gebruik. Dat betekent, als een host de andere hosts in de cluster niet kan bereiken, deze aanneemt dat deze 'geïsoleerd' is. Het isolationprobleem blijkt een ondoorgroondelijke fout te zijn, die veelvuldig is beschreven in diverse blogs en community threads. Zelfs in Update 4 van ESX Server 3.5 was dit probleem nog niet opgelost! Op het web valt te lezen dat

wordt aanbevolen om van DNS gebruik te maken (niet per se noodzakelijk) en alle /etc/hosts-bestanden op de ESX Server te controleren. Nadat het isolationprobleem was opgelost door de diverse hosttabellen aan te passen, konden we een van de ESX Servers in de cluster afschakelen, zonder dat de werking van de Windows Server werd verstoord. Doordat ook een van SVA's uitvalt, geldt hier natuurlijk dat de shared storage van de resterende ESX Server weer een spof vormt.

Conclusie

Het door StorMagic ontwikkelde Storage Virtual Appliance (SvSAN) is vooral geschikt voor kleine ondernemingen die zich geen relatief dure SAN kunnen veroorloven. Met behulp van SvSAN is op eenvoudige wijze een voor het MKB betaalbare shared storage-infrastructuur te creëren zonder dat daarvoor een 'dure' SAN op basis van een fysieke iSCSI of FC disk array nodig is. Daarnaast biedt het product via de mirroring feature bescherming tegen uitval van de shared storage en heft daarmee het single point of failure op. Ook op dat punt is geen dure mirroringfunctie op een storage array meer nodig. SvSAN mirroring, in combinatie met VMware's High Availability, biedt bescherming tegen uitval van zowel de shared storage als de fysieke ESX host. Uitval van het virtuele SAN of een ESX Server heeft geen gevolgen meer voor de actief draaiende VMs. Die blijven onder alle omstandigheden, met name bij uitval van de gebruiker merkbare onderbreking, gewoon doordraaien. ■

Een bètaversie voor VMware vCenter is in de maak en er zijn ook plannen voor een SvSAN versie voor ESXi, Hyper-V en Zen. De basis 2 TB managementconfiguratie heeft een list price van \$995, die thans 'zo lang de voorraad strekt' zonder kosten met een valide Promo Key kan worden gebruikt. De High Availability-optie gaat \$995 per ESX Server kosten. De managementlicenties die betrekking hebben op de 4 TB, 8 TB en onbeperkte storagecapaciteit bevatten tevens de High Availability-optie.



De-duplicatie voor iedereen

De data volumes binnen zowel kleine als grote bedrijven blijven met rasse schreden groeien. Volgens een onderzoek van IDC is de hoeveelheid digitale gegevens die bedrijven wereldwijd aanmaken en opslaan de laatste drie jaar met 3.000 procent toegenomen. Binnen het huidige economische klimaat staan bedrijven onder steeds grotere druk om hun opslagkosten te reduceren. Tegelijkertijd zien zij zich voor de taak gesteld om hun gegevens te consolideren. Daarnaast moeten IT-organisaties ook nog eens in staat zijn om in geval van een calamiteit alle bedrijfsgegevens op snelle wijze te herstellen om de bedrijfsuitval tot een minimum te beperken. Dit verklaart waarom de-duplicatie van data in 2009 het meest besproken onderwerp op het gebied van gegevensopslag was. Grote ondernemingen hebben data de-duplicatie reeds in een vroeg stadium omarmd vanwege de mogelijkheid die deze technologie biedt om de opslagcapaciteit te ontzien, de prestatie van opslagsystemen te verbeteren en de back-uptijden te verlagen. Hoewel deze techniek tot voor kort alleen was voorbehouden aan grote bedrijven vanwege de hoge kosten die ermee gepaard gingen, begint data de-duplicatie nu de volwassenheidsfase te naderen. Daardoor wordt data de-duplicatie ook toegankelijker voor kleine en middelgrote bedrijven. Data de-duplicatie is een techniek voor gegevensconsolidatie die naar grote blokken redundante gegevens (normaliter 4KB of meer) zoekt en deze slechts één keer opslaat, ongeacht het aantal kopieën dat ervan in omloop is. Er wordt gebruikgemaakt van een zogenoemde pointer om naar de oorspronkelijke gegevensblokken te verwijzen, zodat gegevens niet telkens opnieuw moeten worden opgeslagen. Kortom, data de-duplicatie maakt een einde aan het bestaan van meerdere kopieën van dezelfde gegevens door naar identieke bestanden of gegevensblokken te zoeken en ervoor te zorgen dat er slechts één kopie van wordt opgeslagen. Data de-duplicatie kan op twee manieren uitgevoerd worden: bij de bron en op de doellocatie. Sommige oplossingen bieden ruimte voor beide methoden, terwijl andere oplossingen zich slechts tot één variant beperken. Wanneer de-duplicatie bij de bron plaatsvindt, worden gerepliceerde gegevens tijdens het back-upproces verwijderd alvorens deze naar de opslaglocatie worden verzonden. Deze methode bespaart de bandbreedte van het netwerk, omdat het gegevensvolume dat via het netwerk naar de doellocatie wordt overgedragen kan worden beperkt tot een factor 20. Deze methode blijkt met name efficiënt in situaties

waarin de opslaglimiet van het netwerk bijna is bereikt, of back-ups worden uitgevoerd binnen filialen met een netwerk met een beperkte bandbreedte. Bij deze methode is het mogelijk dat back-ups meer tijd en een groot aantal CPU-cycli in beslag nemen. Dit kan resulteren in prestatieproblemen op productiecomputers. Wanneer de-duplicatie op de doellocatie plaatsvindt, worden de gerepliceerde gegevens tijdens het back-upproces verwijderd zodra deze de opslaglocatie bereiken. Dit maakt het mogelijk om de initiële back-up sneller te voltooien, doordat CPU-intensieve ontdebellingstaken op de bron-CPU worden vermeden. Maar aangezien alle kopieën die voor het de-duplicati proces in omloop zijn via het netwerk worden overgedragen, kan dit resulteren in een bottleneck, omdat de niet-ontdubbelde gegevens de overdracht mogelijk vertragen. Het enorme potentieel van de-duplicatie schuilt in het feit dat het de gegevensopslag tot maar liefst 90 procent kan reduceren. De-duplicatie is een bewezen methode om meer gegevens met minder opslagcapaciteit op te slaan. Dit levert tastbare zakelijke voordelen op, zoals: lagere opslagkosten, doordat er op efficiëntere wijze gebruik wordt gemaakt van de opslagcapaciteit. Dit maakt het mogelijk om minder opslagsystemen aan te schaffen en te beheren. Er is sprake van een gunstig milieueffect door lager gebruik van stroom, koeling en apparatuur, met als gevolg een lagere CO₂-uitstoot. Bovendien is een snelle return on investment te realiseren dankzij minder aankoop van storage. Dit is met name van belang voor kleine en middelgrote bedrijven, omdat dit hen het vermogen en de flexibiliteit biedt om snel veranderingen door te voeren die zich in directe resultaten vertalen. Tot slot maakt de-duplicatie snellere back-ups en verbeterde gegevensopslag mogelijk, zodat bedrijven effectiever kunnen inspringen op juridische en compliance-kwesties zonder daarbij de opslagcapaciteit verder te belasten. Bedrijven die hun opslagcapaciteit willen verhogen en hun opslagkosten willen reduceren, kunnen data de-duplicatie niet meer aan zich voorbij laten gaan. Hoewel de-duplicatie traditioneel op hardwareniveau plaats vond en alleen aan grote ondernemingen was voorbehouden, bereikt deze technologie nu de volwassenheidsfase. Hierdoor wordt de-duplicatie beschikbaar in de vorm van een betaalbare softwareoplossing. ■

DAVID BLACKMAN, GENERAL MANAGER ACRONIS NORTHERN EUROPE
(DAVID.BLACKMAN@ACRONIS.COM)