

Alternatief voor Fibre Channel ?

Windows Clustering met iSCSI

Sinds de ratificatie van het iSCSI-protocol in 2003 is er in de Windows-omgeving meer belangstelling gekomen voor iSCSI-gebaseerde opslagsystemen. Wij bekeken hoe betrouwbaar en stabiel iSCSI presteert in een geclusterde Windows Server 2003-omgeving.

BRAM DONS

Volgens sommige criticasters is iSCSI niet geschikt als opslagnetwerk, hoofdzakelijk vanwege de 'onveilige' IP-netwerktechnologie. Dat is in principe juist, maar vaak wordt vergeten dat een IP-netwerk met behulp van het TCP-protocol toch redelijk betrouwbaar te maken is. Een iSCSI-gebaseerd opslagnetwerk zal nooit de betrouwbaarheidsgraad van Fibre Channel kunnen bereiken, maar de iSCSI-technologie wordt langzaam maar zeker volwassen. De verwachting is dat iSCSI voor de meeste mid-range toepassingen zal uitgroeien tot een goed en betaalbaar alternatief voor de nog altijd te dure en lastig te implementeren Fibre Channel-technologie. Dat Microsoft vertrouwen heeft in de iSCSI-technologie blijkt uit de ondersteuning daarvan binnen de Windows 2003 cluster-omgeving. Voorlopig wordt dit nog slechts op twee nodes ondersteund, maar met de komst van Windows Server 2003 Service Pack 1 worden acht nodes ondersteund. Wij onderwerpen een twee-node clustersysteem aan diverse tests.

Testconfiguratie

Alhoewel Microsoft de hardwarecomponenten voorschrijft voor een iSCSI-gebaseerd clustersysteem, maken we voor deze test toch gebruik van niet-gekwalificeerde componenten. Voor de goede orde: dat dingt niets af van de kwaliteit van de afzonderlijke componenten. De twee cluster-nodes bestaan uit een

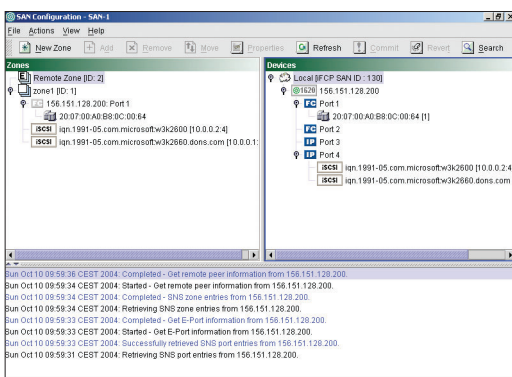
Pentium IV 2.4 GHz-systeem met elk 1 GB intern geheugen. Daarop draait Windows Enterprise Server 2003 en Microsoft's iSCSI-initiator versie 1.05a. Verder is het absoluut aan te raden om voor het 'IP Storage Network' een apart 1 gigabit netwerk te gebruiken, dat bovendien gescheiden is van het locale LAN en/of WAN. Microsoft stelt als absolute eis dat een 1 Gbps-netwerk wordt gebruikt. Bovendien moet het data-path in de switch (routing) volledig hardwarematig zijn uitgevoerd. Verder moet de switch van het type 'non-blocking' zijn. Beide nodes zijn in de test echter via een eenvoudige (laag 2) 1 Gbps 3COM Superstack 3 switch verbonden met een zogenaamde 'SAN router'. Voor omgevingen zonder Fibre Channel-gebaseerde opslagsystemen kun je van een iSCSI-target gebruik maken, die rechtstreeks op het IP-gebaseerde opslagnetwerk wordt aangesloten. Voor de toepassing van FC-gebaseerde opslagsystemen is een router nodig, die de koppeling met een IP-netwerk mogelijk maakt. Om het IP-gebaseerde opslagnetwerk (waarop het iSCSI-protocol draait) te koppelen aan een FC-gebaseerd opslagsysteem, wordt een router gebruikt om het iSCSI-protocol te vertalen naar het Fibre Channel-protocol en vice versa. Daarvoor zijn er van verschillende leveranciers routers beschikbaar. Wij hebben voor de test de beschikking over de nieuwe Eclipse 1620 SAN Router van de firma McDATA.

Vorbereiding

Bij een shared SCSI-bus of FC-gebaseerd clustersysteem moeten alle nodes direct fysiek toegang hebben tot de aangesloten opslagsystemen. Bij toepassing van iSCSI is dat niet anders. De cluster-nodes hebben via het IP-netwerk met behulp van een iSCSI-initiator of iSCSI-gebaseerde Host Bus Adapter (HBA), wel of niet voorzien van TCP Offload Engine

(TOE), direct toegang tot het opslagsysteem. Als opslagsysteem dient een zogenaamde 'iSCSI-target'. Dat kan een speciale 'storage appliance' zijn, die aan de 'front end'-kant is voorzien van een 100/1000 Mbps ethernet-aansluiting voor de koppeling met een IP-gebaseerd opslagnetwerk. Aan de 'back end' moet hij aangesloten zijn op een SCSI-, Fibre Channel-, of SATA-gebaseerd disksysteem. De meeste storage-appliances werken intern op basis van het Linux besturingssysteem. Zoals we al zagen in onze testopstelling, is het IP-opslagnetwerk via een SAN router verbonden met een FC-gebaseerd opslagsysteem, waarbij als 'initiator' Microsofts 'iSCSI Initiator' dient. Die is overigens gratis vanaf de Microsoft-website te downloaden. Voordat met de installatie van de Microsoft Cluster-software kan worden begonnen, moet eerst gecontroleerd worden of beide cluster-nodes toegang hebben tot het gedeelde opslagsysteem. Als voorbereiding daarop, wordt eerst op beide nodes de iSCSI-initiator-software geïnstalleerd en daarna wordt de configuratie van de McDATA 1620-router gedaan.

De 1620-router heeft vier poorten, twee voor Fibre Channel en twee voor IP. We configureren een van de IP-poorten voor iSCSI (iFCP is ook mogelijk) en we sluiten het 9176 opslagsysteem aan op een van de FC-poorten. De FC-poort is zelfconfigurerend. Daarna wordt het opslagsysteem en de beide nodes in een zone opgenomen.



Abbeelding 1 De configuratie van de 1620 SAN router

Na de configuratie van de 1620 moeten nog twee diskvolumes worden aangemaakt voor de installatie van een clustersysteem: een Quorum- en een Datadisk. De StorageTek 9176 is een disk array-systeem dat bestaat uit twaalf SCSI-schijven. Met behulp van StorageTek's SANtricity Storage Manager is de disk array op te delen in verschillende Virtuele Volumes, die elk opgebouwd kunnen zijn uit een of meer disks in een bepaalde RAID-configuratie. Wij

kieszen voor twee volumes op basis van RAID 0. Voor het opslagnetwerk wordt een apart gigabit-ethernet IP-netwerk gebruikt.

We starten het opslagsysteem en een van de twee cluster-nodes op. Via de iSCSI-initiator wordt op de iSCSI-targets ingelogd en worden beide diskvolumes met NTFS geformatteerd. Daarna wordt de tweede node opgestart, en controleren we of beide NTFS-partities via de iSCSI-initiator bereikbaar zijn. Op beide cluster-nodes moeten verder Active Directory Services (ADS) worden geïnstalleerd, waarin beide nodes tot hetzelfde domein moeten behoren. We installeren daartoe op de eerste node ADS, waarbij tegelijk een Directory Name Service (DNS) wordt geïnstalleerd. Op de eerste node wordt voor 'Domain controller for a new domain' gekozen, en op de andere node 'Additional domain controller for an existing domain'. Daarna kiezen we 'Domain in a new forest', de DNS-naam voor het nieuwe domein en de Domain NetBIOS-naam. Na het doorklikken van de default-instellingen (zoals installatie-directories), wordt de Active Directory Installation Wizard gestart die ADS installeert. Na de verplichte reboot van het systeem kan worden begonnen met de installatie van het clustersysteem op de eerste node (de andere blijft nog uitgeschakeld).

Installatie clustersysteem

Vanuit het menu Administrative Tools → Cluster Administrator wordt de installatie van de eerste cluster-node gestart. Je kunt kiezen uit 'Open connection to cluster', 'Add nodes to cluster' en 'Create new cluster'. Voor de eerste node kiezen we uiteraard de laatste mogelijkheid, waarna het 'Welcome to the New Server Cluster Wizard'-scherm verschijnt. In het volgende scherm wordt de gevonden domeinnaam automatisch weergegeven en voeren we een unieke cluster-naam binnen dat domein in. In het daaropvolgende menu wordt de computer-naam ingevuld van de eerste node binnen de nieuwe cluster. Op basis van deze gegevens wordt automatisch geanalyseerd of de clusterconfiguratie aan alle voorwaarden voldoet.

De eerste controle is nu of er een remote-verbinding met de gekozen naam 'iSCSIcluster' in het domein kan worden aangemaakt. Nadat de verbinding met de server tot stand is gekomen, wordt gecontroleerd of de gewenste cluster-node ook geschikt is als cluster-node (bijvoorbeeld door de versie van het OS). Vervolgens wordt gekeken welke cluster-resources beschikbaar zijn, waaronder een Quorum-disk en verschillende typen netwerken. Belangrijk

is de controle op aanwezigheid van een disk op dezelfde storage-bus als de bootdisk. Denk bijvoorbeeld aan de C-schijf, die niet aanmerking komt als disk voor een cluster-resource. Bij de netwerkverbinding wordt gekeken of deze van DHCP gebruik maakt. DHCP wordt namelijk niet ondersteund en de voor clustering gebruikte netwerk-interfaces moeten daarom van statische adressen worden voorzien. Vervolgens wordt gekeken of er een 'sharable quorum resource' kan worden gecreëerd. Zo niet, dan wordt er een 'local quorum cluster' gecreëerd. Daarna wordt er gecontroleerd of er geen 'drive letter collisions' tussen nodes in een bestaande cluster bestaan. Verder worden ook de resources tussen de bestaande cluster en de toe te voegen nodes vergeleken, evenals de netwerken binnen de cluster met de netwerken op de nodes. Als laatste wordt gecontroleerd of alle nodes toegang hebben tot de quorum-resource. Na het invullen van het IP-adres voor de cluster management-tools en de Cluster Service Account-gegevens, wordt het 'Proposed Cluster Configuration'-menu getoond, waarin je de instellingen nogmaals kunt controleren.

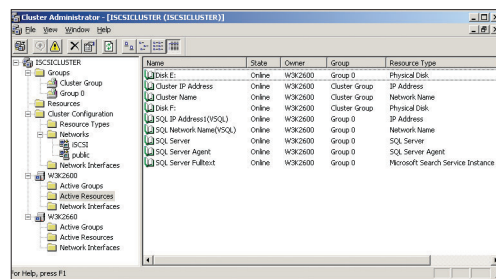
Na akkoord van de bevindingen wordt begonnen met de creatie van de cluster, waarbij de 'cluster services' worden gecreëerd en opgestart. Daarna volgen de resource-types (Volume Shadow Copy Service Task, Majority Node Set, Generic Script) en de configuratie van de resources (quorum, groups). Als laatste worden de cluster-resources opgestart.

Voor de clusterinstallatie op de tweede node, moet de eerste node worden uitgeschakeld. Als eerste worden de ADS geïnstalleerd. Vanuit het menu Administrative Tools → Cluster Administrator kiezen we nu de optie 'Add nodes to cluster'. De installatie verloopt verder bijna identiek aan de eerste node. Nadat de installatie met succes is afgerond, kunnen we de tweede cluster-node weer opstarten. We zien vervolgens alle resources in het 'Cluster Administrator'-menu online komen op een van de twee clusternodes. Om te controleren of de cluster failover-functie werkt, verhuizen we de resources naar de andere cluster-node, wat binnen enkele tientallen seconden is gebeurd (afhankelijk natuurlijk van het aantal te verplaatsen resources). We schakelen vervolgens een van de nodes uit en zien dat de node na inschakeling weer keurig online komt.

Installatie SQL 2000 Enterprise Editon

Nu gaan we SQL 2000 Enterprise Edition op beide cluster-nodes installeren. Tijdens de SQL 2000-setup

wordt automatisch gedetecteerd dat er een cluster-systeem aanwezig is, waarop de SQL Server wordt geïnstalleerd. De installatieprocedure verloopt verder vrijwel identiek aan een standaard SQL-installatie. Wel volgt er een waarschuwing dat SQL Server 2000 niet geschikt is voor Windows 2003, maar die kan genegeerd worden. Na afloop van de installatie moet wel Service Pack 3a worden geïnstalleerd. Nadat we SQL Server op de andere node hebben geïnstalleerd, brengen we beide nodes online en verschijnen de diverse SQL resources in de 'Active Resources'-lijst van de Cluster Administrator. Als laatste voeren we een failover-test uit en zien we dat alle SQL resources keurig van de ene naar de andere node verhuizen (en omgekeerd).



Afbeelding 2 ↑ Actieve resources Cluster Administrator menu

Toekomstplanning

Op dit moment ondersteunt Windows server 2003 iSCSI op twee nodes, die gebruik maken van SCSIport, Storport en MS Software Initiator. De bedoeling van Microsoft is om met Windows 2003 Server SP1 acht nodes te gaan ondersteunen onder Storport en Microsoft Software Initiator stack. SP1 zal daarvoor een aantal substantiële verbeteringen voor de afhandeling van failover bevatten. Op dit moment ondersteunt SCSIport namelijk nog geen afzonderlijke reset van LUN's, waardoor de ondersteuning van SCSIport beperkt blijft tot twee nodes. De afzonderlijke reset van LUN's is zeer belangrijk in multi-node clusteromgevingen, waarbij een disk-failover moet kunnen plaatsvinden met zo min mogelijk verstoring voor de andere disks.

Een verbeterde versie van clusdisk zal gebruik maken van de SCSI Unique ID, waarmee een afzonderlijke LUN-reset kan worden uitgevoerd. Gebruikers met acht-node iSCSI-clustersystemen moeten echter meer dan een half jaar wachten na de release-datum van SP1, voordat ze daarvoor ondersteuning krijgen van Microsoft. Deze tijd is overigens nodig om leveranciers de tijd te geven om hun clustersystemen met SP1 te kunnen testen en ratificeren. ✕