

Gebaseerd op het 'write once, consume many'-principe

ParaScale geeft cloud storage boost

Cloud computing is bezig aan een flinke opmars. Ondanks de populariteit van het concept heerst er nog veel begripsverwarring. Inmiddels weten we dat cloud computing veel meer behelst dan alleen maar het inhuren van storagecapaciteit. Bram Dons beoordeelt de door een Californische start-up op de markt gebrachte ParaScale Cloud Storage Software (PCS).

Voor een goed begrip: wat verstaan we eigenlijk onder cloud storage? Cloud storage is een natuurlijke uitbreiding van SaaS (Software-as-a-Service) en cloud computing. Slaan we Wikipedia erop na, dan lezen we: "de architectuur achter cloud computing is een massaal netwerk met cloudservers die als een grid zijn verbonden en parallel werken". Cloud computing is in essentie een uiterst krachtig concept, waarbij commodity hardware wordt gebundeld tot een grote pool die bestaat uit computer- en opslagcapaciteit. Voorheen was een dergelijk krachtige computeromgeving alleen mogelijk op basis van dure, zogenoemde Massive Parallel Systems (MPP)-componenten. Cloud storage is een relatief nieuw concept binnen de IT-industrie. Het is oorspronkelijk als een service toegepast en bekend geworden door de opkomende start-up internetondernemingen, de zogenoemde Web 2.0-bedrijven, die besloten tot outsourcing van hun opslagomgeving. Sindsdien is het concept uitgebreid en is er sprake van twee typen cloudomgevingen. Zo worden er tegenwoordig public en private clouds onderscheiden.

Google en Amazone

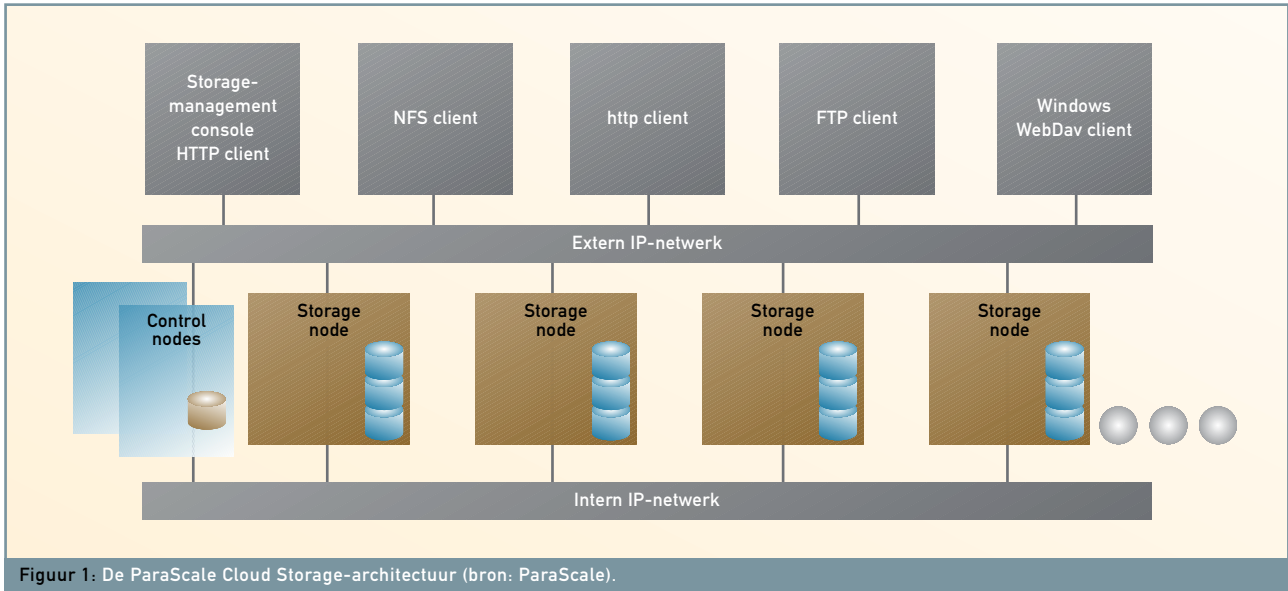
Vandaag de dag wordt cloud computing voornamelijk voor in-house processing toegepast, zoals bij Google, of extern als een service aangeboden, zoals de zogenoemde Elastic Compute Cloud van Amazon (Amazon EC2). Het cloudconcept is in eerste instantie door Amazon (S3) ook in de vorm van een service naar de storageomgeving gemigreerd. Achter de Amazon-schermen wordt van commodity hardware devices gebruikgemaakt waarmee een grote storage pool is gecreëerd. Opkomende webondernemingen hebben dankbaar van dit aanbod gebruikgemaakt. Dit heeft helaas niet kunnen voorkomen dat er in het begin nogal wat begripsverwarring ontstond rondom de begrippen cloud computing en cloud storage. Inmiddels weten we dat cloud computing niet uitsluitend een service betreft en meer behelst dan alleen maar het inhuren van storagecapaciteit. In het algemeen gaat cloud storage over schaalbare storagecapaciteit en prestaties door desgewenst standaard server hardware toe te voegen en deze via een standaard IP-netwerk te laten delen door meerdere gebruikers.

Geen wondermiddel

Wat cloud storage vooral niét is? In tegenstelling tot wat sommige 'experts' beweren, is cloud storage geen wondermiddel en het is dan ook hoogst onwaarschijnlijk dat het de SAN-omgeving in de nabije toekomst gaat vervangen. In plaats daarvan biedt het een nieuw storageconcept dat alleen voor bepaalde toepassingen en omgevingen geschikt is. Voorbeelden daarvan zijn de distributie en levering van digitale content, file archivering, videobewaking, streaming media van grote file systems en medische beelden. Al deze toepassingen hebben een aantal eigenschappen gemeen. Het gaat hierbij om grote datasets en file systems, er is een noodzaak tot parallele verwerking, er is sprake van een lange levensduur van de data en tot slot is er een grote behoefte aan systemen die weinig kosten. Als men bijvoorbeeld gebruikmaakt van een transactionele database, dan is cloud storage geen bruikbaar alternatief en blijft men ook in de toekomst aangewezen op de traditionele leveranciers van storage arrays. Ook als een tier 1 NAS-storagesysteem nodig is, voor e-mail, databases of file serving, kan men beter kiezen voor de traditionele leveranciers van NAS-systemen.

Architecturen

Voor de komst van het concept cloud storage was er voor de hierboven genoemde video, digitale content en gelijksoortige toepassingen alleen een dure NAS-cluster beschikbaar. Dergelijke clustered NAS-systemen zijn al ruim tien jaar op de markt en kunnen, afgezien van de prijs, in prin-



Figuur 1: De ParaScale Cloud Storage-architectuur (bron: ParaScale).

cipe wel worden gebruikt voor het bouwen van een cloud storage-omgeving. Deze traditionele systemen zijn gebaseerd op een Tightly Coupled Symmetric (TCS)-architectuur, die speciaal voor HPC (High-Performance Computing) zijn ontworpen en worden nu door leveranciers aan de man gebracht als een zogenaemde scale-out oplossing voor de opkomende cloud storage-markt. De volgende generatie NAS-clusters is echter gebaseerd op de Loosely Coupled Asymmetric (LCA)-architectuur, waarbij de metadata en besturing van de cluster centraal plaatsvindt. Dit type architectuur is niet geschikt voor de HPC-omgeving, maar ontworpen om aan de eisen te voldoen die een zeer omvangrijke storageomgeving stelt. We hebben het dan over petabytes (zie ook het artikel over Schaalbaar NAS op pagi-

na 46 in dit tijdschrift, alwaar nader wordt ingegaan op de twee verschillende architecturen). Binnen schaalbare NAS-platforms zijn er veel keuzes te maken, maar in het algemeen is er een onderscheid te maken op basis van het service-, hardware- en softwaremodel. Als voorbeeld van een LCA-oplossing bespreken we hier het onlangs door de firma ParaScale op de markt gebrachte product ParaScale Cloud Storage Software, afgekort tot PCS.

Toepassingen

PCS is een software-only oplossing waarmee een krachtige cloud storage-omgeving kan worden gecreëerd en beheerd. Het is een softwareapplicatie die binnen de user space op meerdere Linux-operatingssystemen draait, zoals RedHat 5.2 of

CentOS 5.2. Het voorziet in een objectgeoriënteerd, gedistribueerd, cluster file systeem dat op een standaard Linux-hardwaresysteem draait met processoren van Intel en AMD. PCS biedt cloud storage-software waarmee tien tot honderden commodity servers zijn te verenigen tot een storagecluster. PCS is ontworpen om een oplossing te bieden voor de toenemende vraag naar systemen die efficiënt met content en data-archivering kunnen

PCS IS OOK GESCHIKT VOOR HOSTED PRIVATE CLOUDS

omgaan. PCS is echter niet geoptimaliseerd voor het leveren van single write-prestaties, zoals de traditionele NAS-clusteroplossingen dat wel doen. PCS is ook geschikt voor de toepassing van zogenaamde hosted private clouds, waarbij gebruikers hun eigen private cloud kunnen installeren. Door de toepassing van thin-provisioning van virtual file systems is men niet gebonden aan de fysieke capaciteit van een opslagsysteem. Hierdoor wordt de toepassing van just-in-time aanschaf van opslagcapaciteit mogelijk gemaakt. Zo kan bijvoorbeeld een 15 TB virtual file system worden gecreëerd bij slechts 5 TB fysiek aanwezige opslagcapaciteit. De software is niet alleen geschikt voor private clouds, maar ook serviceproviders kunnen gebruikmaken van de ParaScale eco-structuur voor de creatie van een externe cloud. Een andere toe-

(Advertentie)

passing is videobewaking. In grote steden met duizenden bewakingscamera's worden continu beelden naar een storage-systeem gestreamd, waarbij iedere camera zijn eigen set met image files creëert. Door bijvoorbeeld een ParaScale storage cloud met veel storage nodes uit te rusten, kunnen de van de tientallen camera's afkomstige datastreams naar een enkele storage node worden gestreamd, om daarna naar de overige honderden

CLOUD STORAGE ZAL SAN-OMGEVING NOOIT VERVANGEN

nodes te worden gerepliceerd. Op die manier kunnen duizenden camera's hun data continu en snel kwijt binnen de storage cloud, een optie die met een traditioneel NAS-clustersysteem niet goed mogelijk is. In dezelfde categorie toepassingen valt bijvoorbeeld ook video streaming en online video sharing van files, waarbij sprake is van intensieve read work loads. Bij grote sportevenementen, zoals bijvoorbeeld de Tour de France of de Olympische Spelen, willen miljoenen mensen via webservers kopieën van dezelfde video's opvragen. Door van honderden onafhankelijke storage nodes gebruik te maken, kan een enkele file parallel worden benaderd. Dit wordt mogelijk gemaakt doordat PCS automatisch een videofile naar meerdere storage nodes kopieert. Eenzelfde soort toepassing, zij het op kleinere schaal, is de parallele aanvraag van data binnen datawarehouses. Een weer geheel andere toepassing is het gebruik van cloud storage als archiverings- of back-up staging-platform. PCS biedt een ideale oplossing voor archivering en near-line storage van disk back-up door voor writes de beschikbare massale bandbreedte te combineren met parallele toegang tot meerdere storage nodes. Een algemeen toepasbare strategie is om back-up targets van bestaande NAS-systemen naar een PCS cloudtoepassing om te zetten, met als resultaat betere prestaties, een hogere kostenefficiëntie en een betere beheerbaarheid. Nadat de data in de storage cloud is opgeslagen, kan het in tweede instantie naar tape worden gemigreerd voor langetermijnarchivering en off-site opslag. Tenslotte kan een PCS storage cloud worden toegepast in het kader van

medische research en ruimtevaartonderzoek volgens het principe van 'write once, consume many' workloads.

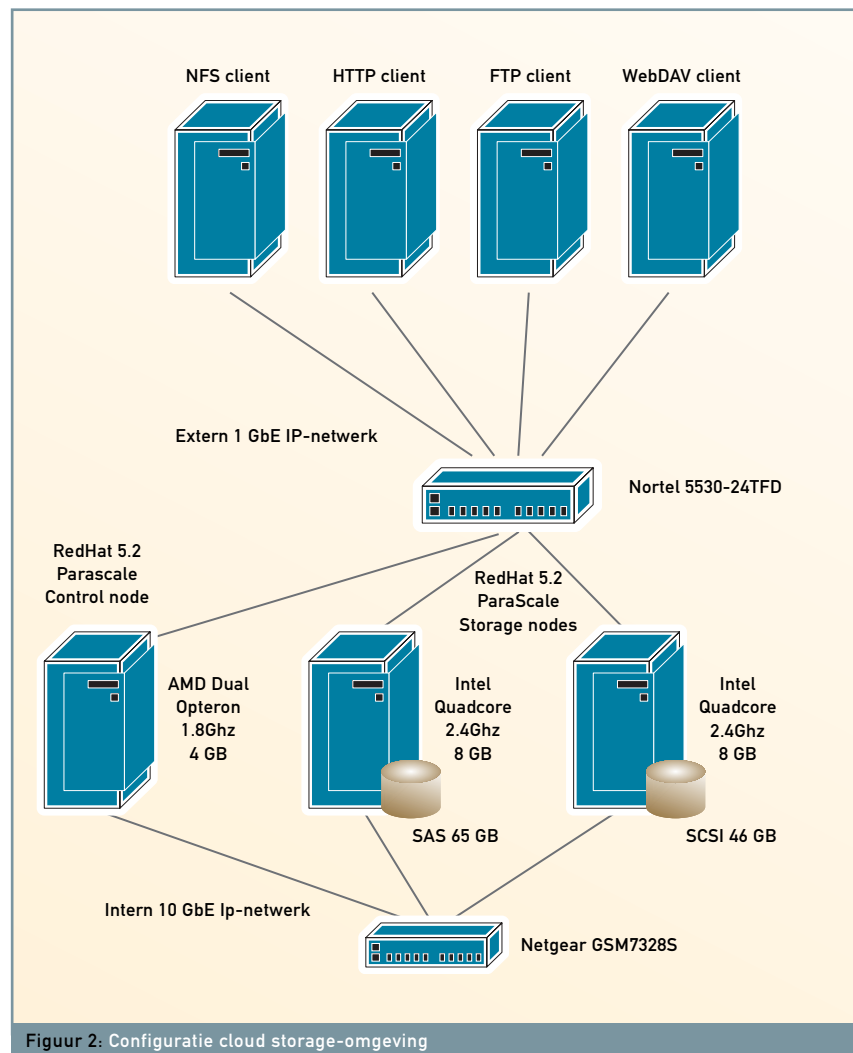
Storage nodes

De ParaScale cloud storage bundelt lokale storage-systemen van heterogene servers in een enkele storage pool. Meerdere thinly provisioned virtuele file systems (namespaces) worden bovenop deze storage pool gecreëerd en bieden diverse type clients toegang tot files op de storage nodes. Clients hebben toegang tot alle storage nodes via standaardprotocollen als NFS, HTTP en FTP. De control node regelt en bestuurt de replicatie, migratie en verwijdering van files tussen de storage nodes en beheert alle informatie, inclusief metadata, over de configuratie en namespaces binnen de storage cloud. Een PCS bestaat uit meerdere servers die met elkaar samenwerken: een of twee control nodes en meerdere storage nodes. De PCS-architectuur is gebaseerd op twee fysiek gescheiden netwerken: een exter-

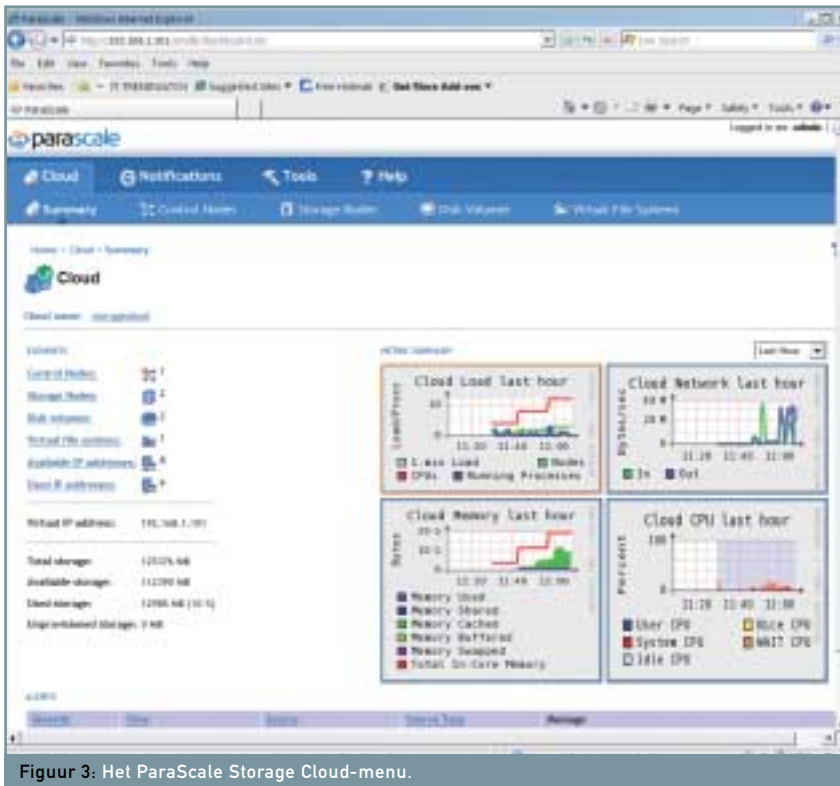
nal en internal IP-netwerk. Het externe netwerk dient voor file storage en communicatie tussen de storage cloud en de clients. De communicatie binnen de storage cloud zelf, tussen de control node en storage nodes, gebeurt via het interne netwerk. De fysieke opsplitsing van beide netwerken biedt belangrijke prestatievoordelen. De storage cloud koppelt geen externe IP-adressen statisch aan een storage node, maar doet dit vanuit een floating IP-adrespool. Bij uitval van een storage node wordt het daaraan gekoppelde IP-adres automatisch naar een andere storage node overgeheveld en wanneer de node weer beschikbaar komt, wordt het adres weer teruggezet. Als intern netwerk kan een 1GbE of 10GbE-netwerk of een InfiniBand-netwerk dienen.

File storage en file systems

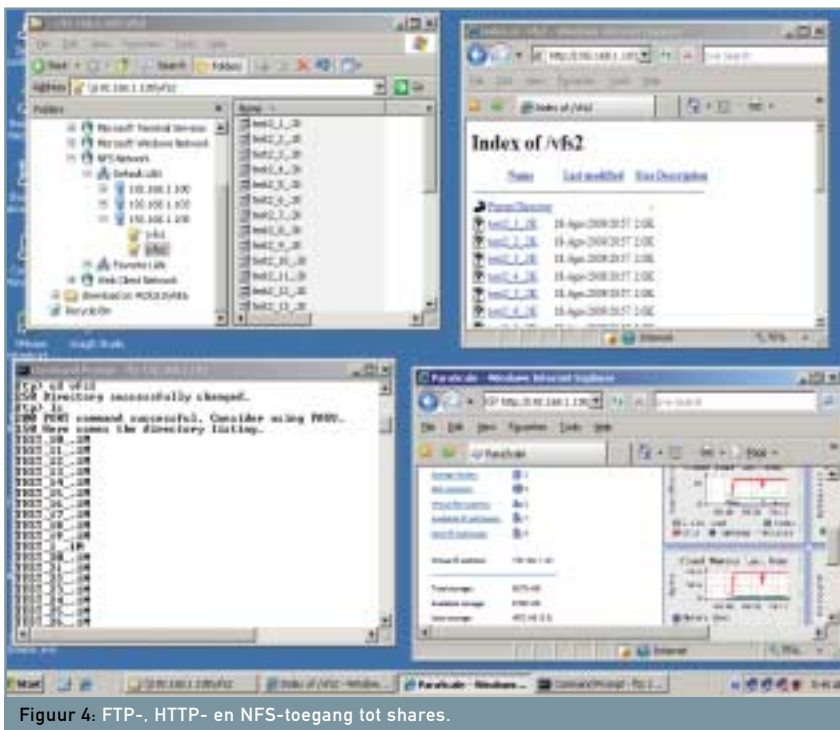
Een virtueel bestandssysteem is een namespace die toegang biedt tot de via storage nodes gerepliceerde files. Elke file komt maar eenmaal voor, maar files zelf



Figuur 2: Configuratie cloud storage-omgeving



Figuur 3: Het ParaScale Storage Cloud-menu.



Figuur 4: FTP-, HTTP- en NFS-toegang tot shares.

kunnen op meerdere volumes zijn opgeslagen en verspreid zijn over meerdere storage nodes. Het enige verschil met een traditioneel file system is dat er geen beperkingen worden opgelegd door het onderliggende storagestelsel dat op een enkele server aanwezig is. Via de storage nodes hebben gebruikers met standaard remote file access-protocollen toegang.

De control node houdt de mapping bij tussen file names binnen het virtuele file systeem en hun fysieke locatie binnen de storage cloud. De control node koppelt een file binnen een virtueel file systeem met een fysiek object-ID en houdt bij op welke disk volumes welke files zijn opgeslagen. Clients worden via DNS, automounter, virtuele server router of andere

standaardmechanismen weliswaar met een enkele storage node verbonden, maar elke node binnen de storage cloud is in staat om elke file binnen een virtueel NFS, HTTP-systeem of FPT share op te vragen. Wanneer een client een open request naar een bestaande file uitvaardigt, stuurt de storage node deze naar de control node door, die vervolgens zijn eigen namespace map raadpleegt om de file name naar een global object-ID om te zetten. Met behulp van zijn eigen replicatie-database vraagt deze een lijst op van volumes waarop de file is opgeslagen, stuurt deze informatie naar de storage node terug en retourneert tot slot de file.

Aanpassing disk volumes

De storagecapaciteit van de storage cloud is eenvoudig te vergroten door, naast de mogelijkheid van toevoeging van een extra storage node(s), een nieuwe disk of disk array aan een bestaande storage node toe te voegen en deze vervolgens via provisioning en enabling beschikbaar te maken. De nieuw verkregen ruimte is onmiddellijk beschikbaar voor elk in de storage cloud aanwezige virtual file system. De control node bepaalt vervolgens op basis van de storagecapaciteit van de toegevoegde disk en statistische gegevens welke files naar het nieuwe disk volume moeten worden gemigreerd. Dit migratieproces dient om de data evenwichtig over de storage nodes te verspreiden, een proces dat automatisch verloopt. De toevoeging van een disk volume gebeurt online en heeft geen invloed op de werking van de andere disk volumes. Een disk is vanwege bepaalde redenen weer offline te brengen, denk bijvoorbeeld aan het draaien van fsck bij een mogelijke verdenking van file system corruption, en daarna weer online te brengen. Wil men een disk compleet vervangen, dan moeten indien dit nog mogelijk is, eerst alle files worden geëvacueerd en pas daarna kan de disk worden uitgeschakeld. Ook evacuatie van files is een online operatie die geen invloed heeft op de andere disk volumes binnen een storage node. Verder kunnen disk volumes worden gemarkeerd met de term 'unavailable', waardoor ze niet in de storage cloud kunnen worden opgenomen, bijvoorbeeld een kritisch file system, of een disk volume die men wil reserveren als raw volume. Denk aan een database op een raw volume.

Virtual file system

Een virtual file system is een namespace via welke clients gemeenschappelijke toegang hebben tot files. Elke file bestaat

weliswaar eenmalig binnen een virtual file system, maar kan op meerdere volumes zijn opgeslagen en verspreid zijn over meerdere storage nodes. Clients hebben niet direct toegang tot een specifieke file op een storage node. Dat komt omdat files zich niet statisch op een bepaalde storage node bevinden, maar als gevolg van load balancing, file replicatie en migratie van storage node kunnen wisselen. Let wel, de data wordt verplaatst buiten de control node om, zelfs als men van een redirector gebruik maakt, waardoor de bandbreedte nooit door de control node enige beperking ondervindt. Alle mount en file serving requests worden transparant naar een storage node omgeleid. Een kenmerkend verschil tussen een virtual- en een traditioneel file system is dat er geen beperking is door het onderliggende storage-systeem van een enkele server. Men kan bij de toewijzing van file system-capaciteit meer denken aan quota's die men kan definiëren voor een bepaalde gebruiksgroep. Een ander punt dat men in gedachten moet houden, is dat replicatie-politicijs, oftewel hoeveel

kopieën van elke file die men in het file system wil opslaan, op file system-niveau worden ingesteld.

Een client heeft niet eerder toegang tot een bepaald virtual file system voordat er een share met behulp van NFS v3, HTTP, en/of FTP-protocollen op het betreffende virtual file system is aangemaakt. NFS file locking (lockd) wordt daarbij niet

tems van clients (Windows of UNIX). Na de creatie is de grootte, via verkleinen en vergroten, van een virtual file system op elk moment aan te passen zonder interruptie van de verbonden client. De ParaScale software verdeelt na de creatie de file-objecten over de verschillende storage volumes, zodat de belasting gelijkmatig wordt verdeeld. Mochten volumes vol raken, dan is het eenvoudig om een nieuwe storage node of een disk volume online bij te plaatsen.

PCS IS EEN SOFTWARE-ONLY OPLOSSING

ondersteund. Afhankelijk van de toepassing kunnen UNIX en Windows clients van een van deze protocollen gebruikmaken. Factoren die daarbij een rol spelen zijn: gestelde eisen bij de toegang tot files (alleen lezen, of lezen en schrijven); type bestanden (grafisch, binair, documenten en/of of tekstuele bestandseisen); beveiligingseisen (bijvoorbeeld host- en/of toegang op gebruikersniveau); operating sys-

Replicatie en policies

De ParaScale software voert regelmatig automatisch replicaties uit om er voor te zorgen dat, op basis van gedefinieerde policies, meerdere kopieën van elke file op meerdere storage nodes worden opgeslagen. Als een storage volume, waarop een bepaalde file is opgeslagen, uitvalt dan vindt de storage node automatisch een andere kopie van deze file. Mocht de hele storage node uitvallen, dan wijst de control node een andere storage node aan om het uitgevallen externe adres van de falende node over te nemen. Komt een

(Advertentie)



**We nemen ICT-ers
in bewaring...**

Doit overwegen om jouw ICT-ervaring en kennis van data-opslag in te zetten voor het veiliger maken van Nederland? Achter de schermen van de politie wordt namelijk door een team ICT-experts hard en resultaatgericht gewerkt. Aan het beheren, opslaan én optimaal beveiligen van grote hoeveelheden gegevens bijvoorbeeld. En alles moet ook nog snel terug te vinden zijn door de juiste mensen van verschillende korpsen. Een stabiel digitaal archief is tenslotte essentieel voor alle politietaken.

Aan een motief om bij de vtsPN aan de slag te gaan zal het je niet ontbreken: werken in een spannende, complexe ICT-omgeving én uitstekende arbeidsvoorwaarden, zoals diverse mogelijkheden voor persoonlijke ontwikkeling en een goede carrière/privébalans.

Meer weten? Geef jezelf dan aan voor een van de beschikbare functies op www.politie-ict.nl We rekenen je graag in.

POLITIE
vts Politie Nederland

www.politie-ict.nl

uitgevallen volume of storage node weer beschikbaar, dan wordt deze vanzelf weer online en up-to-date gemaakt. File replication verbetert ook de prestaties, omdat het de kans vergroot dat een client request direct wordt afgehandeld door een beschikbare storage node. Let wel, hoe meer kopieën er worden bijgehouden, hoe meer storageruimte er wordt geconsumeerd. Er moet bij file replication door de gebruiker dan ook een afweging plaatsvinden tussen data-redundantie en de daarvoor benodigde extra storageruimte. De ParaScale-software kent standaard drie voorgeconfigureerde replication policies: 1 copy, 2 copy en recovery policy. Bij de 1 copy-policy wordt slechts een kopie van elke file in het file system bewaard. De 2 copy-policy dupliceert elke file naar een ander device of naar een heel andere storage node. De recovery policy wordt intern door het file system gebruikt voor de opslag van control node recovery snapshots. Mocht de control node totaal uitvallen, dan kan via de snapshot deze op een andere control node toch weer worden hersteld.

High availability

Het zal duidelijk zijn dat de control node een single point of failure vormt binnen de ParaScale Storage Cloud. De control node High Availability (HA) biedt bescherming tegen uitval van de control node. Daarvoor is een identieke dedicated server nodig die fungeert als standby control node en een extern gedeeld storage system of SAN om actieve control node-informatie en metadata op te slaan. De toevoeging van een control node is een online operatie waarvoor geen enkele service hoeft te worden onderbroken. Wel is het zo dat elke uitstaande client request naar een uitgevallen control node wordt geblokkeerd totdat de services op de standby control node zijn opgestart. Wanneer de control node meer dan 30 seconden down is, vindt er geen interruptie plaats van de requests op de al geopende files.

Configuratie PCS

Met een beperkt aantal beschikbare nodes is de mate van schaalbaarheid en prestaties van een storage cloud nauwelijks te beoordelen, maar wel de functionaliteit, gebruiksgemak en betrouwbaarheid van de ParaScale cloud storage. Daarnaast willen we zien in hoeverre de belofte om commodity hardware, zowel servers, netwerken en opslagsystemen, in een cloud storage-omgeving te kunnen toepassen in de praktijk wordt waarge-

maakt. Voor onze test configureren we een cloud storage met twee storage nodes en een control node. De clients zijn via een 1GbE netwerk verbonden met de clients en de nodes onderling via een 10GbE netwerk op basis van de NetGear GSM7328 low cost switch. Als lokaal storage system voor de storage nodes dienen SAS, SATA II en via FC en iSCSI gekoppelde disksystemen.

De eerste stap bestaat uit de installatie van RedHat 5.3 en de ParaScale V1.3.0 software op de storage en control nodes. Vervolgens worden de lokale disk systemen (SCSI, SAS en FC disk array) aan de

CONTROL NODE HOUDT BIJ OP WELKE DISK VOLUMES WELKE FILES ZIJN OPGESLAGEN

storage pool toegewezen (provisioned). We creëren twee virtual file systems, voegen daar NFS shares aan toe en voor de clienttoegang worden de FTP, HTTP en NFS-protocollen geconfigureerd. Via een Windows Client, met daarop Windows Services for UNIX 3.5 geïnstalleerd, maken we via een FTP, HTTP en NFS login contact met de drietal shares waarop een aantal files worden gecreëerd. Vervolgens schakelen we beurtelings een van beide storage nodes aan en af en constateren dat, op de FTP-sessie na, de andere twee verbindingen intact blijven. De volgende

test bestaat uit de toevoeging van lokale storage systemen aan de storage nodes, waarna ze online aan de storage pool kunnen worden toegevoegd. Storage systemen zijn met unprovisioning trouwens weer af te koppelen. Tevens vergroten en verkleinen we online de opslagcapaciteit van beide virtual file systems.

Conclusie

Met behulp van de ParaScale-software is een hoogschaalbare, low cost, op basis van heterogene hardware, NAS-cluster te bouwen voor de implementatie van een storage cloud. Als cluster nodes kunnen standaard AMD/Intel hardware servers worden gebruikt, waarop Linux draait. Elk type server is te gebruiken als storage en control node, bijvoorbeeld op basis van SMP of dual/quad core. De storage nodes kunnen van elk type direct gekoppeld disk system gebruikmaken, waaronder (i)SCSI, SATA II, SAS of Fibre Channel Disk Array. De aanschafprijs van de software is afhankelijk van de fysiek aanwezige opslagcapaciteit binnen de cloud. De list price voor de Noord-Amerikaanse markt bedraagt \$1.05/GB. Voor ondernemingen die al vast kennis willen maken met cloud storage, een 4TB trial versie is te downloaden vanaf: <http://www.parascale.com/index.php/support-info>. ■

BRAM DONS IS ONAFHANKELIJK IT-ANALIST;
INFO@IT-TRENDWATCH.NL

(Advertentie)



File & E-mail Archiving

HSM is nog geen archiving! Kom kijken hoe het wel moet.



www.storageXperience.nl



Reserveer nú een 1-op-1 sessie met een ISIT consultant