



RNA Networks maakt extra servers overbodig

# Pionieren met geheugenvirtualisatie

Om gelijke tred te kunnen houden met de uitbreiding van de server- en opslagcapaciteit, is het belangrijk dat ook andere kritische componenten binnen datacenters kunnen worden uitgebreid. Om de geheugencapaciteit van datacenters te kunnen vergroten, brengt RNA Networks een oplossing op de markt voor geheugenvirtualisatie. Bram Dons werpt een kritische blik op de oplossing van de startup.

Naast het opslag- en serversysteem, is het interne geheugen in toenemende mate een kritisch onderdeel van een datacenter. Een methode om de prestaties van dit geheugen te verbeteren, is om op elke server extra intern geheugen te plaatsen. Het op grote schaal uitbreiden van intern geheugen op lokale servers is echter nog steeds een dure aangelegenheid. Het zou economisch en technisch aantrekkelijk zijn, wanneer al het lokale geheugen via een netwerk in een gemeenschappelijke geheugenpool zou kunnen worden ondergebracht. Hierdoor kunnen meerdere servers toegang krijgen tot deze geheugenpool. Op aanvraag kunnen servers dan tijdelijk beschikken over meer geheugencapaciteit.

Vaak worden nieuwe servers alleen maar aangeschaft vanwege de mogelijkheid om de interne geheugencapaciteit met een factor  $x$  te kunnen vergroten, in plaats van dat er daadwerkelijk behoefte is aan meer computervermogen. Hoewel opslagcapaciteit en cpu-vermogen de afgelopen jaren exponentieel zijn toege-

nomen, hebben de capaciteit van het geheugen en daarmee de prestaties van opslagsystemen geen gelijke tred kunnen houden. Architecten van datacenters vervangen servers om de paar jaar, waarbij geheugen en opslagcapaciteit meestal worden over-provisioned. Tegenwoordig bieden steeds meer storageleveranciers over en/of thin provisioning van opslagcapaciteit.

In het algemeen leidt een gebrek aan het afstemmen van vraag en aanbod van systeemcapaciteit, tot extra kosten voor ruimte, stroom en beheer. Dit geldt ook voor het gebruik van het interne geheugen. Bij de toepassing van het lokale geheugen worden geheugenbronnen inefficiënt gebruikt. Uitbreiding van de interne geheugencapaciteit levert slechts lokaal een prestatieverbetering op, maar tegen aanmerkelijk hoge kosten.

Gevirtualiseerd geheugen kan het prijsprestatie-model binnen datacenters echter drastisch verbeteren. De startup RNA Networks heeft onlangs het product RNACache gepresenteerd dat geheugen-

virtualisatie voor High Performance Clustersystemen (HPC) binnen de datacenter-omgeving biedt.

## Prestatieverlies

Kritische zakelijke applicaties vragen hoge prestaties van alle netwerkbronnen. Geheugen is, naast cpu, netwerk en opslag, één van de vier sleutelcomputerbronnen die de algehele prestaties van een datacenter bepalen. Echter, de ontwikkeling van geheugen met betrekking tot capaciteit ligt ver achter op die van processors en opslag. Hoewel leveranciers van cpu's beweren dat datacenters *processor-bound* en leveranciers van opslagsystemen *storage-bound* zijn, vormt in veel gevallen het geheugen een barrière voor

## GEVIRTUALISEERD GEHEUGEN

### VERBETERT

### DATACENTERPRESTATIES

het behalen van maximale prestaties. Zo hebben videocontent, raw data voor simulaties op het gebied van olie- en gasreserves en texture maps voor door de computer gegenereerde filmscènes, vluchtreserveringsinformatie en zakelijke analyse grote hoeveelheden data in het terabyte-bereik nodig. Dergelijke datasets zijn veel te groot om zelfs in systemen met een grote hoeveelheid RAM te kunnen draaien. Tegenwoordig ligt het geheugen-

bereik van een enkele server tussen de één en 64 GB. De toenemende grootte van datasets veroorzaakt enorm prestatieverlies van applicaties, waarbij factoren als latency en doorvoer van grote invloed zijn. Doordat niet de hele dataset in het lokale geheugen kan worden geladen, moeten applicaties steeds wachten tot weer een gedeelte van de dataset vanaf disk kan worden opgehaald. Dit betekent natuurlijk een wachttijd van meerdere milliseconden.

Multi-core processoren blijven daardoor onderbezet en moeten telkens op data wachten die niet snel genoeg beschikbaar is, nog afgezien van het feit dat de meeste populaire besturingssystemen en applicaties nog lang niet zijn aangepast voor parallelle verwerking op basis van multi-core.

### Oplossingen

In een poging het prestatieprobleem op te lossen, zijn beheerders van datacenters gevangen in een vicieuze cirkel door steeds meer storage of servers aan te schaffen. Deze oplossingen zijn echter een inefficiënt lapmiddel en lossen niet het werkelijke probleem op.

of geheugen dan nodig is, en scaling-up, het toevoegen van geheugen of grotere servers. Ten slotte proberen beheerders zelfs het probleem op te lossen door speciale software te ontwerpen. Het toevoegen van meer servers verbetert de prestaties grotendeels niet, omdat elke server nog steeds met zijn eigen lokale geheugen moet werken. Door een verdeel-en-heersmethode toe te passen, raakt een dataset alleen maar meer gefragmenteerd en verhindert het een continue datastream, omdat het via het netwerk moet worden gerepliceerd. Hoewel opslagsystemen steeds sneller zijn geworden, is een opslagsysteem niet de plaats waar de actieve dataset zich bevindt. Een snel opslagsysteem leidt slechts tot een marginale verbetering van de applicatie, omdat het niet *application-aware* is. Met andere woorden: het begrijpt niets van de toestand waarin de applicatie zich bevindt.

In sommige gevallen hebben datacenter-architecten in hun zoektocht naar hogere prestaties data grids geïmplementeerd. In veel van dit soort systemen vormt data-replicatie een ongewenste overhead en zijn grid-oplossingen moeilijk te integre-

netwerk door de gelijktijdige requests nog verder vertraagd. De vraag is hoe meerdere systemen via het netwerk toch effectief dezelfde dataset kunnen benaderen. Een mogelijke oplossing hiervoor is geheugenvirtualisatie.

### Geheugenvirtualisatie

De computerindustrie heeft een grotere efficiency geboekt door virtualisatie van het besturingssysteem. Een voorbeeld hiervan is de hosting van meerdere besturingssystemen op een enkele server. Deze inspanningen hebben, via de opdeling van fysieke bronnen voor de ondersteuning van meerdere applicaties, tot een grotere efficiency geleid.

## EXTRA STORAGE

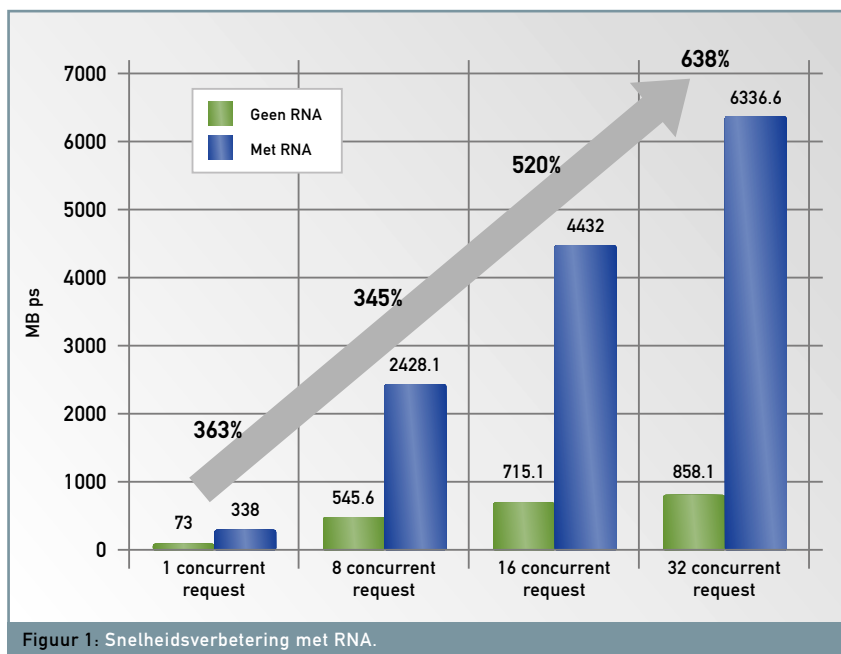
## EN SERVERS

## ZIJN GEEN OPLOSSING

Geheugenvirtualisatie gaat een stapje verder door al het geheugen in het datacenter te bundelen en dit als een gevirtualiseerde geheugenbron beschikbaar te stellen voor elke met het datacenter verbonden server. Deze methode reduceert de latency en verbetert de prestaties, in combinatie met een minimale toename van de complexiteit en tegen lagere kosten.

Net als de virtualisatie van opslag en servers, biedt geheugenvirtualisatie een manier om de prestaties te verbeteren en de aanschaf- en operationele kosten te verlagen. Toepassing van geheugenvirtualisatie binnen een gevirtualiseerde server is al langere tijd bekend, een voorbeeld hiervan is de Memory Balloon-technologie binnen VMware's ESX Server. Daarbij is het interne geheugen van server als een globale geheugenpool voor de verschillende VMs beschikbaar en kunnen, afhankelijk van de belasting, de VMs hun geheugencapaciteit vergroten of verkleinen.

Intern geheugen kan ook worden gevirtualiseerd door dit buiten de server in een gevirtualiseerde geheugenpool op te nemen. Daarbij wordt het interne geheugen logisch ontkoppeld van de lokale fysieke server en als een gedeelde, globale netwerkbron voor elk daaraan verbonden systeem toegankelijk gemaakt. De firma RNA Networks is er in geslaagd een appliance te ontwikkelen die meerdere servers op een gedeeld geheugen kan aansluiten.



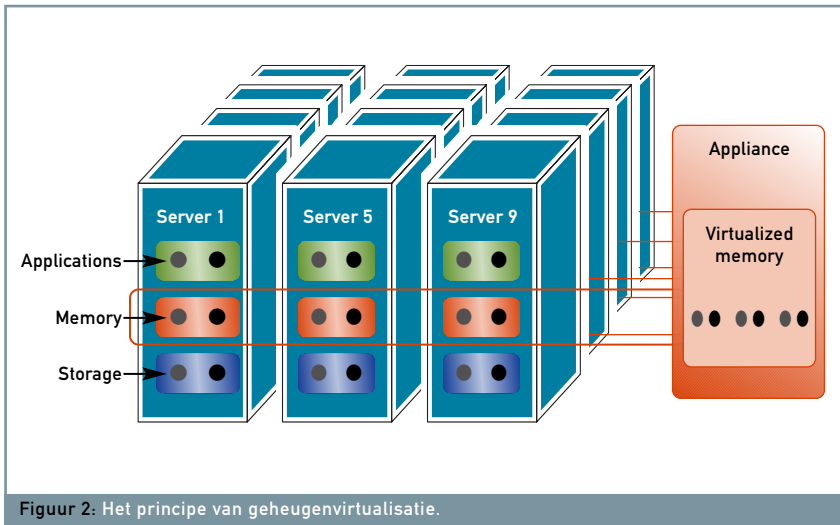
Figuur 1: Snelheidsverbetering met RNA.

Het resultaat is vertraging die kan oplopen tussen de tien en honderd milliseconde, wat een onacceptabele flessenhals vormt voor zakelijke en mission-critical applicaties.

Een andere veel toegepaste methode is scaling out, het toevoegen van servers. Daarnaast zijn er methoden als over-provisioning, het toevoegen van meer opslag

en optimaliseren en hebben een *tightly coupled* relatie met een applicatie nodig. Een hogesnelheidsnetwerk de data soms wel snel afleveren, maar de overhead en beheer van dergelijke oplossingen is aanzienlijk.

Het probleem wordt nog verergerd, wanneer meerdere systemen dezelfde data proberen te benaderen. Hierbij wordt het



Figuur 2: Het principe van geheugenvirtualisatie.

### RNA geheugenvirtualisatie

De oplossing voor geheugenvirtualisatie van RNA Networks creëert een grote pool met geheugen die voor alle computers in de cluster beschikbaar is. Servers hoeven dan niet langer de dataset te repliceren, omdat ze de data als een enkele gevirtualiseerde geheugenpool zien. Dat maakt een toegangssnelheid mogelijk die bijna gelijk is aan die van een lokaal geheugen. Bij deze methode kan geheugenvirtualisatie eenvoudig worden toegepast en zijn er geen aanpassingen aan bestaande applicaties of opslaginfrastructuur nodig. Geheugenvirtualisatie als gedeelde netwerkbron maakt het mogelijk om grote datasets sneller te verwerken. Dat betekent minder wachttijd voor zoekopdrachten en snellere runtimes voor applicaties.

### RNA-OPLOSSING

#### CREËERT GEVIRTUALISEERDE GEHEUGENPOOL

Geheugen kan dynamisch zonder downtime op elke server in de datacenter worden gealloceerd, waardoor er minder swapping naar storage nodig is en het over-provisioning van lokaal geheugen voorkomt. Het gebruik van virtueel geheugen als een actieve opslagplaats voor zakelijke kritische data, verdeeld over meerdere applicaties, verbetert, in vergelijking met de huidige opslagoplossingen die op disk spindles en bewegende delen die zich ver van de applicatieprocessor bevinden zijn gebaseerd, de prestaties met een factor tien of meer. Door het opslagsysteem uit het kritische

pad van applicatieprestaties te verwijderen, kunnen datacenterbeheerders de aanschaf van dure proprietary storage-systemen, die uiteindelijk maar een marginaal effect zouden hebben op de algehele prestaties van applicaties, voorkomen. Met behulp van geheugenvirtualisatie kan, op basis van bestaande commodity servers, een hoogpresterende datacenter-omgeving worden geïmplementeerd en kan de ROI van de bestaande datacenterbronnen worden verbeterd.

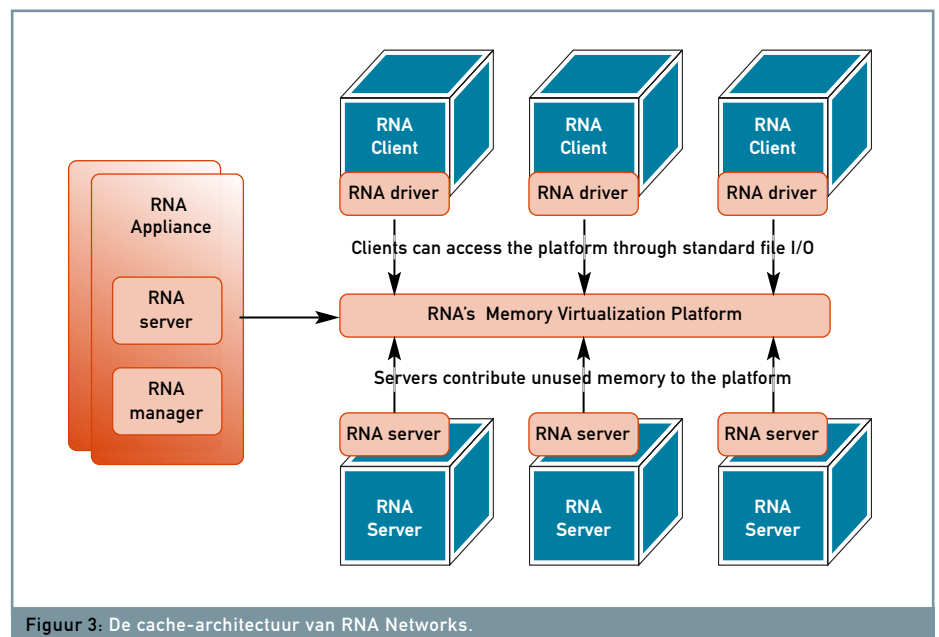
#### Architectuur

De RNACache-geheugenvirtualisatieoplossing creëert een gevirtualiseerde, gedeelde geheugenpool als een *collaborative* cache. De grootte van deze cache wordt bepaald door de totale hoeveelheid aanwezige lokale RAM van alle aangesloten systemen. Systemen hebben dezelfde toe-

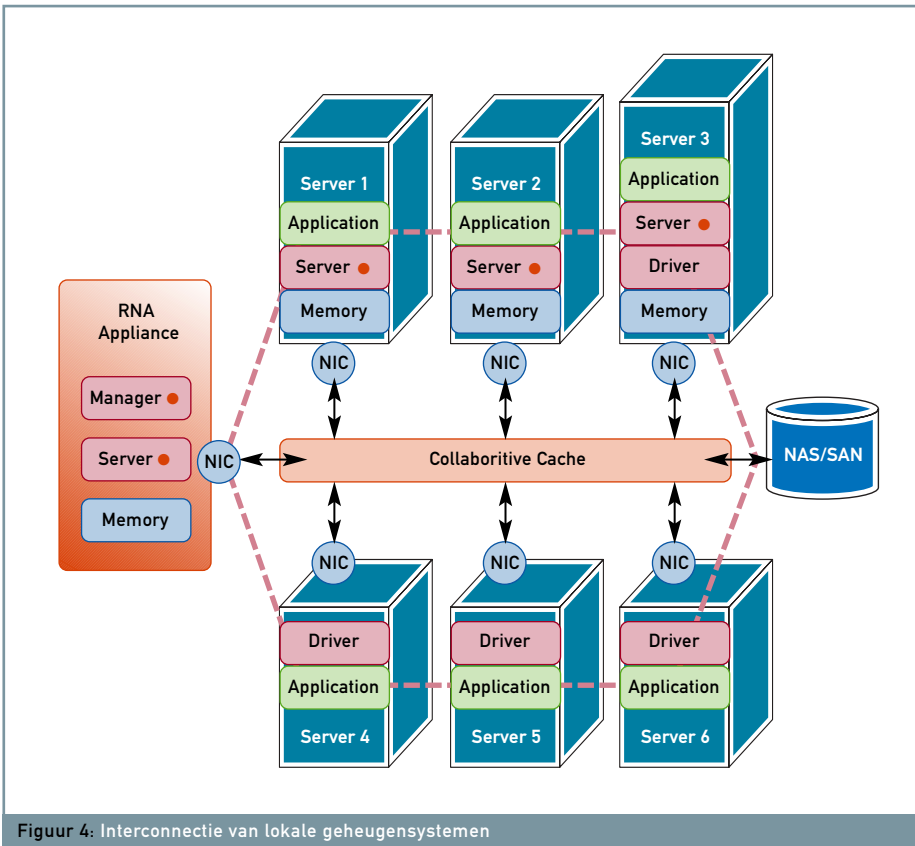
gang tot het RNACache alsof het een lokaal geheugen betreft. Ze zijn via een hogesnelheid fabric aangesloten, waardoor de toegangssnelheid bijna gelijk is aan het lokale geheugen. Dit cache-systeem maakt deel uit van wat RNA het *Memory Virtualization Platform (MVP)* noemt. Het MVP vormt de kern de RNACache-producten. De basis RNACache-architectuur bestaat uit een RNA Appliance en twee softwarecomponenten, de RNAdriver en RNAserver (zie schema 3). De beheersoftware voor de RNACache-omgeving is vooraf op de appliance geïnstalleerd. Het op de appliance intern aanwezige geheugen kan aan de totale hoeveelheid cachegeheugen, samengesteld uit het lokale geheugen op de servers, worden toegevoegd. Met de RNAserver-software kunnen nodes hun lokale geheugen via het netwerk aan de geheugenpool toevoegen.

### PRESTATIES VERBETEREN MET FACTOR TIEN

De beheersoftware op de appliance heeft de volledige controle en beheer over de bijdrage van de hoeveelheid lokaal geheugen van elke node aan de totale cachecapaciteit. Nodes krijgen toegang tot de cache door het draaien van een system service via de RNAdriver. Nodes kunnen bij toepassing van RNAserver en RNAdriver zowel als server of als client voor het cache-systeem fungeren. Dat wil zeggen, een systeem kan als client het cache-



Figuur 3: De cache-architectuur van RNA Networks.



Figuur 4: Interconnectie van lokale geheugensystemen

systeem gebruiken, als server levert het via het lokale geheugen ook een bijdrage aan de capaciteit van het cache-systeem. Voor de toegang tot de geheugenpool moeten beheerders op elke node een RNACache directory mounten, zodat alle toegang tot deze directory automatisch op het cache wordt uitgevoerd. De integratie van RNA MVP kan op file system-, block- of applicatieniveau plaatsvinden. Op file system-niveau wordt RNA's virtuele geheugen als een drive gemount, zodat applicaties rechtstreeks, zonder enige aanpassing toegang hebben tot de cache. Op applicatieniveau ondersteunt RNA een Java API, zodat applicaties elke soort data binnen enkele milliseconden kunnen delen en opslaan. Op blockniveau ondersteunt RNA's MVP-geheugen virtualisatie bij de toegang tot gestructureerde data. Zoals gezegd, elk serversysteem kan als server intern RAM-geheugen aan het cache toevoegen via de CLI of via een configuratie en beheerdashboard. Dit dashboard regelt de instellingen en het beheer via een webgebaseerde GUI.

#### Eenvoud en betrouwbaarheid

De RNA Appliance ondersteunt hogesnelheidsnetwerk fabrics, InfiniBand en 10GbE en Remote Direct Memory Access (RDMA) waardoor een lage latency en hoge doorvoer worden gegarandeerd, terwijl de com-

plexiteit van de RDMA API voor de applicatie verborgen blijft. InfiniBand en 10GbE zijn non-proprietary op standaarden gebaseerde netwerken die de prestaties van een omgeving voor High Performance Computing (HPC) tegen een redelijke prijs voor het datacenter beschikbaar maken. Mellanox, de leverancier van InfiniBand ConnectX InfiniBand-gebaseerde adapters en InfiniScale-gebaseerde switches, ondersteunt de op geheugenvirtualisatie gebaseerde clusteromgevingen. De firma claimt prestaties, bij 40 Gbps, in het

bereik van 10 tot 40 Gbps met een 1 microseconde latency!

Applicatieontwikkelaars hoeven zo niet langer rekening te houden met de beperkingen van een lokaal fysiek geheugensysteem en standaardapplicaties kunnen ongewijzigd blijven functioneren. Voorzietingen die een hoge beschikbaarheid bieden, voorkomen, door meerdere kopieën van de data in cache op te slaan en met persistent writes die aan gecertificeerde messaging-standaarden voldoen, het verlies van data wanneer servers of net-

## BEPERKINGEN LOKAAL FYSIEK GEHEUGENSYSTEEM ZIJN VERLEDEN TIJD

werken uitvallen. De oplossing van RNA sluit ook naadloos aan op de bestaande integriteitpoliticies. RNA reduceert honderden of duizenden op opslagsystemen en databases uitgevoerde leesopdrachten door frequent geraadpleegde data in de geheugenpool op te nemen. Dit vermindert de noodzaak voor een dure load balancer en maakt het mogelijk dat servers optimaal presteren, zelfs wanneer een eenvoudige en niet dure round-robin load balancer wordt toegepast. RNA is gekoppeld aan de algemene file system calls of via API's op applicatieniveau.

#### RNA Appliance

De appliance van RNA is een 1U hoog systeem met 32 GB RAM on-board, twee poorten, DDR InfiniBand HCA naar PCIe, GbE NIC en twee RJ45-poorten. DDR Infini-

(Advertentie)

**Backup/Restore**  
Backup to disk net zo goedkoop als tape? Met deduplicatie is dit realistisch.

[www.storageXperience.nl](http://www.storageXperience.nl)

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

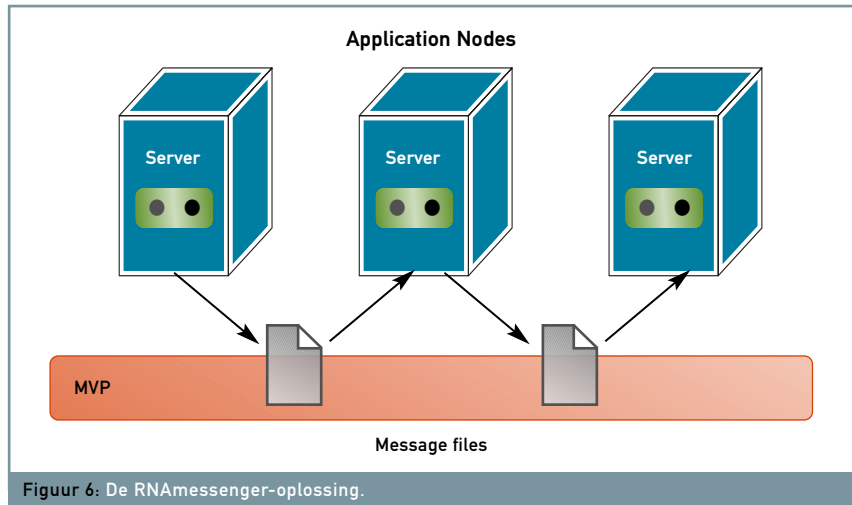
**ISIT**  
THE STORAGE COMPANY

Band biedt verder 1.8 GBs bandbreedte. De RNACache ondersteunt maximaal 10 TB en gelijktijdige toegang tot duizend nodes. Een doorsneeomgeving heeft vijftig RNA-servers en 250 RNA-drivers. De toegangstijd voor een 4 KB datatransfer bedraagt vijf microseconde. De doorvoer van een enkele node bedraagt 12 Gbps voor een random read en 8,8 Gbps voor write. De ondersteunde besturingssystemen voor de clusternodes zijn RHEL 4/5, CentOS 4/5, Fedora Core 7/8, en SUSE LES 10/11.

### RNAMessenger

RNAmessenger is het andere product van RNA Networks en biedt een alternatieve oplossing voor de bestaande message bus-architecturen. Het is een oplossing die vooral voor elektronisch handelsverkeer is bedoeld, waarin een groot aantal transacties per seconde, meer dan 53.000 transacties per seconde, moeten worden afgehandeld. De combinatie van een lage latency en een hoge doorvoer maken RNAmessenger tot een goede oplossing voor deze markt.

RNAmessenger is gebaseerd op het MVP-platform en gebruikt de bestaande gedeelde geheugenpool. In plaats de op sockets gebaseerde communicatie te gebruiken, voegt RNAmessenger nieuwe messages toe aan een subscription-bestand die in de pool is opgeslagen. Vanwege de unieke architectuur van RNAmessenger is er nauwelijks sprake van enig prestatieverlies voor het gebruik van een certified message delivery. RNAmessenger wordt met de appliance meegeleverd, waardoor applicaties de Messaging API en RNAdriver kunnen gebruiken bij de toegang tot de geheugenpool om naar het gespecificeerde message subscription-bestand te kunnen lezen of schrijven.



Figuur 6: De RNAmessenger-oplossing.

### Conclusies

De door de firma RNA Networks gepresenteerde appliance voor geheugenvirtualisatie is een nieuwe vorm van virtualisatie binnen de wereld van open systemen. De toepassing beperkt zich nog tot de HPC-omgeving voor bepaalde applicaties en ondersteunt alleen op Linux gebaseerde clusternodes. Alleen wanneer deze technologie op basis van standaarden door meer leveranciers wordt toegepast, kan het in de open omgeving een succes worden. Aangezien het merendeel van de systemen in de open omgeving op Windows is gebaseerd, moet voor de grootschalige toepassing van geheugenvirtualisatie ook deze omgeving worden ondersteund. De

leverancier zou er inmiddels over nadenken om drivers te gaan ontwikkelen voor Solaris en op de wat langere termijn voor Windows.

Verder introduceert de enkelvoudige toepassing van de appliance een single point of failure. Voor een kritische omgeving is dan ook een nu nog niet ondersteunde voorziening voor failover gewenst. Ten slotte claimt de leverancier dat de appliance een non-proprietary oplossing is, maar in werkelijkheid is dat natuurlijk niet zo. Voor de ondersteuning van geheugenvirtualisatie blijven gebruikers afhankelijk van de softwaredrivers en de appliance van RNA Networks.

In de toekomst zal geheugenvirtualisatie misschien een sleutelrol kunnen gaan spelen bij de toepassing van servervirtualisatie doordat er meer VMs op een fysieke server kunnen draaien en uiteindelijk de toepassing van VMs bij kritische bedrijfsapplicaties in een virtuele omgeving mogelijk maken. Toepassing van geheugenvirtualisatie kan ook een alternatief zijn voor de in populariteit toeneemende Solid State Disks (SSD), vooral voor database-omgevingen.

Een 8-node RNACache-pakket kost minder dan 60.000 dollar en een pakket voor RNAmessenger 70.000 dollar. Qua prijsstelling kan dit zeker een aantrekkelijk alternatief zijn, omdat de lokale geheugencapaciteit op die manier voor alle systemen betrekkelijk goedkoop kan worden uitgebreid. Startup RNA Networks had op moment van dit schrijven nog geen units ter beschikking om te evalueren. De verwachting is dat de tweede helft dit jaar wel het geval zal zijn. ■

## TOEPASSING MOET DOOR MEERDERE LEVERANCIERS WORDEN ONDERSTEUND



Figuur 5: De RNA Appliance.

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