

*Tools pakken gebrekkige controle
over e-mail aan*

Vloedgolf van e-mail in het gareel

Bram Dons

Meer dan de helft van alle bedrijfskritische data bevindt zich in een messaging-systeem¹, in de vorm van attachments, messages of contactinformatie. Bijna de helft van de organisaties kan niet voorkomen dat gebruikers e-mail verwijderen. En bij driekwart ervan wordt informatie in lokale e-mailboxen opgeslagen, die lang niet altijd onder regulier uitgevoerde back-up-procedures vallen. Bram Dons bespreekt de (on)mogelijkheden van e-mailarchivering en typeert de gezamenlijke oplossing van KVS en EMC.

De gemiddelde grootte van een lokale opslagplaats bedraagt 175 Mb. Niet alleen blijft dus een enorme hoeveelheid informatie voor het beheersysteem verborgen, maar ook is deze informatie voor de IT-beheerafdeling niet eenvoudig boven water te krijgen. Het onvermogen om alle in potentie aanwezige data binnen een bepaald tijdsbestek te kunnen opvragen, zal met het oog op de toekomstige regelgeving (zie kader 'Wettelijke bepalingen') voor veel organisaties en ondernemingen een lastig probleem gaan vormen. Ondernemingen die niet aan de gestelde eisen kunnen voldoen, zullen in de toekomst worden geconfronteerd met hoge boetes.

Het e-mailquotaprobleem

Het beheerprobleem van de grote hoeveelheid opgeslagen e-mail van individuele gebruikers wordt weleens ondervangen met het instellen van *mailboxquota*. Maar in de praktijk wordt het quantum voortdurend overschreden. Het gebruik van quota introduceert ook andere ernstige opslag- en beheerproblemen. Ten eerste moeten gebruikers nu zelf tijd besteden aan het bewaken en beheren van hun e-mailbox, bijvoorbeeld door oude e-mail te verwijderen. Ten tweede kan voor de onderneming waardevolle e-mail door een gebruiker worden verwijderd. Het is duidelijk dat met de komende regelgeving de regels voor het wel of niet verwijderen van e-mail niet door individuele gebruikers mogen worden geïnterpreteerd.

Een alternatief voor het quotasysteem is om eenvoudigweg opslagcapaciteit toe te voegen. Dat is echter geen haalbare zaak; de praktijk laat zien dat organisaties nu al problemen hebben om hun reguliere dagelijkse back-ups te maken. De natuurlijke reactie daarop is om maar dure back-up-devices aan te blijven schaffen of om opslagcapaciteit op disk vrij te maken door data naar tape over te hevelen. Een alternatief is om disk-snapshot-technologie toe te passen, die meer dagelijkse back-ups mogelijk maakt.

Zoals we hierna zullen zien, maken moderne e-mailarchiveringssystemen dit soort dure maatregelen in feite overbodig. Een voorbeeld van een dergelijke implementatie is de productcombinatie van Enterprise Vault, een e-mailarchiveringsysteem van KVS, en het Centera-opslagsysteem van EMC, in combinatie met Exchange Server van Microsoft.

Wettelijke bepalingen

Dat de hoeveelheid geproduceerde e-mail de laatste jaren dramatisch is toegenomen is algemeen bekend. De belangrijkste oorzaken zijn het toegenomen gebruik van attachments, maar ook in het algemeen een frequenter gebruik van e-mail en de toename van spam.

Minder bekend zijn de wettelijke bepalingen op het gebied van e-mailarchivering, die ertoe zullen leiden dat de daarvoor benodigde opslagcapaciteit snel zal groeien. Met name de financiële wereld wordt geconfronteerd met een zeer strikte regelgeving op het gebied van e-mailarchivering. Ondernemingen zullen door wetgeving worden gedwongen om historische e-maildata, op verzoek van bijvoorbeeld overheidsinstanties, binnen korte tijd te kunnen tonen (de zogenaamde *retention*). Dat betekent dat e-maildata direct vanaf een direct te benaderen medium, bijvoorbeeld een vaste schijf, moeten kunnen worden opgevraagd.

Pst-bestanden

Sommige bedrijven kiezen ervoor om alle e-mail op te slaan en leggen gebruikers geen restricties op. Veel organisaties die van Exchange gebruik maken, hebben echter mailboxquota moeten introduceren om eindgebruikers te dwingen regelmatig hun e-mailbox op te schonen. Omdat de meeste gebruikers toch hun oude e-mail willen bewaren, zagen ze zich gedwongen om mail te archiveren in extra pst-bestanden (postoffice-bestanden) van Outlook. Hier kleeft een aantal nadelen en risico's aan.

Exchange maakt gebruik van een *single-instance*-opslagmodel waarbij, bijvoorbeeld, een naar honderd gebruikers verzonden bestand maar één keer op schijf wordt opgeslagen; daarbij wijzen pointers naar de honderd mailboxen, zodat slechts een minimale hoeveelheid schijfruimte hoeft te worden gebruikt. Wanneer gebruikers besluiten om hun e-maildata naar pst-bestanden te verhuizen, gaat het voordeel van dit opslagprincipe verloren. Het is duidelijk dat het gebruik van pst-bestanden op termijn leidt tot een explosieve gegevensaanwas van overbodige, dubbele e-maildata. Bovendien vallen de in pst-bestanden opgeslagen e-maildata nu onder de controle van de individuele gebruiker, waardoor het maken van een reguliere back-up een probleem wordt. Veel organisaties hebben dit probleem opgelost door eindgebruikers hun pst-bestanden op netwerkfileservers te laten zetten. Behalve dat dit al genoemde extra kosten met zich meebrengt, worden hiermee natuurlijk het back-up-systeem en het netwerk extra belast.

Een ander probleem van pst-bestanden vormen de wijzigingen die Outlook telkens bij het starten aanbrengt in de pst-bestanden die gelieerd zijn aan het

betreffende Outlook-profiel. Dat heeft tot gevolg dat de meeste pst-bestanden elke dag in de dagelijkse *differential* back-up-procedure worden meegenomen, ongeacht of ze zijn gewijzigd of niet.

Een laatste probleem is het gebrek aan controle van de beheerorganisatie over de pst-bestanden, die zich niet meer op de fileservers bevinden. Gegevens op laptops kunnen nu eenmaal in verkeerde handen vallen of verloren gaan.

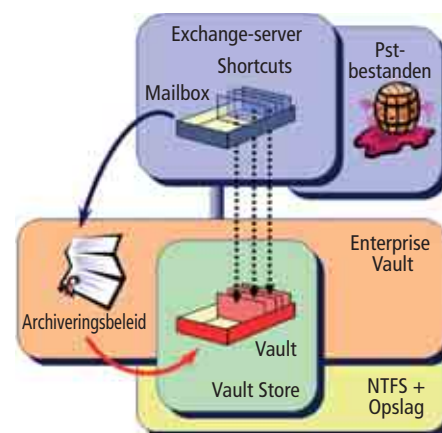
Goede oplossingen

De beste oplossing zou zijn om met een nieuwe architectuur te beginnen, maar gezien de grote *installed base* van Exchange-servers is dit geen praktische en haalbare oplossing. Er is dan ook een aantal leveranciers op het opslag- en beheerprobleem van Exchange ingesprongen, wat geresulteerd heeft in goede oplossingen. Een daarvan is het product Enterprise Vault van de firma KVS, dat onder meer een oplossing biedt voor het gebruik van pst-bestanden. Een direct hiermee verbonden e-mailarchiveringsoplossing is het Centera Content Address Storage-systeem (CAS) van EMC.

KVS Enterprise Vault

Enterprise Vault is een softwareplatform voor het beheer van meerdere Microsoft-repository's, waaronder Exchange-mailboxen, public folders, journals, pst-bestanden en Sharepoint Portal Server. Met betrekking tot de opslag van e-mail kenmerkt het zich door een *policy driven* benadering, waardoor er sprake is van een automatische migratie van e-mail en/of journals naar de Enterprise Vault-database (zie figuur 1). Door *shortcuts* te plaatsen in de mailbox als vervanging van de originele e-mail werkt het voor de gebruiker geheel transparant.

Als de eenmaal opgeslagen e-mail wordt veranderd en daarna weer wordt opgeslagen, dan wordt het veranderde bericht als een nieuw item, naast de originele e-mail, in Vault opgeslagen. Alle e-mails en attachments worden eenmalig in de Enterprise Vault Store opgeslagen. De gebruiker ervaart onmiddellijk het voordeel van deze methode: hij wordt niet langer beperkt door de quotaregeling, want e-mail wordt centraal opgeslagen, en de aanmaak van pst-bestanden is niet langer nodig. Alle e-mail kan nu in een back-up-procedure worden meegenomen. Gebruikers hoeven bovendien geen tijd meer te spenderen aan e-mail-beheertaken, wat tot een enorme kostenbesparing leidt.



Figuur 1 E-mailarchivering met Enterprise Vault

Hiërarchische opslag op basis van regels

Naarmate e-mails langere tijd zijn opgeslagen, kunnen ze van een duur DAS- of SAN-opslagsysteem verhuisd worden naar een opslagsysteem met een lagere total cost of ownership en op een nog later tijdstip naar een tape- of optisch medium worden gemigreerd.

In tegenstelling tot wat een aantal opslagleveranciers ons wil doen geloven, is een dergelijk gelaagd archiverings-systeem allesbehalve nieuw. Het bestaat al vanaf de *mainframe*-tijd en is bekend

Verder lezen

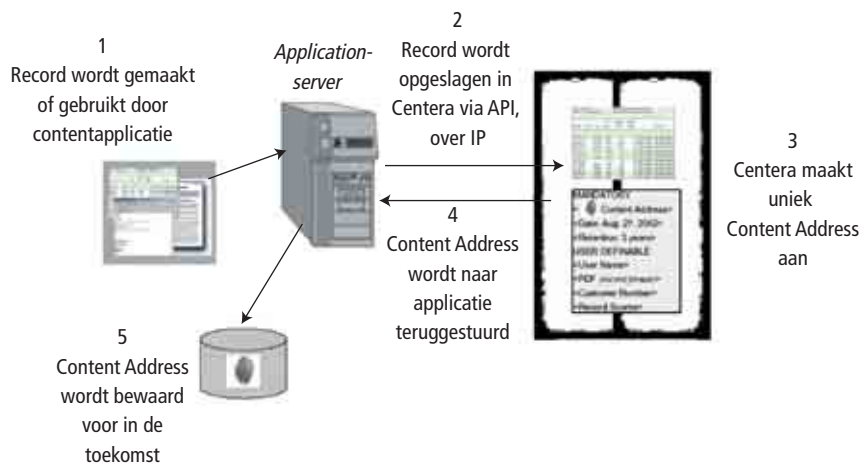
Lees ook het artikel over Opslag en regelgeving, op p. 19 van dit nummer.

onder de naam Hierarchical Storage Management (HSM). In aanvulling op Enterprise Vault biedt KVS de Compliance Accelerator, die specifiek voldoet aan de Amerikaanse NASD 3010- en NYSE 342-regelgeving. De Accelerator maakt het mogelijk dat de daartoe geautoriseerde gebruikers snel een bepaalde e-mail kunnen opzoeken, bijvoorbeeld voor strafzaken of (wettelijk) onderzoek.

Om te voldoen aan huidige en toekomstige regelgeving voor e-mailcontent moeten data in een veilig en niet-veranderbaar formaat worden opgeslagen. Alhoewel optische en tape-devices (*Worm*) aan deze eisen voldoen, staan de bijbehorende kosten de opslag van grote hoeveelheden data in de weg. Daarbij komt dat elk kostenvoordeel van een tape- of optische oplossing verloren kan gaan door productiviteitsverlies vanwege langere wachttijden van de gebruikers. Er is een opslagoplossing nodig die de total cost of ownership van tape combineert met de karakteristieken van optische devices én met behoud van de prestaties van de disktechnologie.

Centera Content Address Storage

Het principe van de in EMC's Centera toegepaste *Content Address*-technologie (CA) is dat elk opgeslagen e-mailrecord wordt voorzien van een 'digitale vingerafdruk'. Dit is afgeleid van de inhoud van het record zelf (met behulp van een hash-algoritme). Elke verandering van een in het archief opgeslagen record



Figuur 3 Koppeling Enterprise Vault en Centera

resulteert in de aanmaak van een nieuw Content Address van dat betreffende record. Daarnaast slaat het Centera-systeem een *Content Descriptor File* (CDF) op, dat gegevens over het betreffende record bevat, waaronder de aanmaakdatum, format, periode dat het bewaard moet blijven en elke andere door de gebruiker gedefinieerde *business logic*. Het CA voor de CDF wordt naar de applicatie (in ons geval Exchange) teruggestuurd en is beschikbaar voor opvraging. Voor elke wijziging worden dus een nieuwe CA en C-Clip (die de bijbehorende bestanden identificeert) gecreëerd, waardoor de originele e-mail niet kan worden overschreven. Hiermee voldoet het systeem aan de regelgeving die vereist dat alle e-mail in zijn oorspronkelijke vorm worden bewaard.

De Centera CA-technologie biedt de garantie dat records niet kunnen worden overschreven, terwijl de *retention*-bescherming ervoor waakt dat een record niet verwijderd kan worden voordat de uiterste datum is verstreken. Centera garandeert de integriteit van

records tijdens de gehele levensperiode. Daarmee biedt het de integriteit van de Worm-media in combinatie met de efficiency en snelheid van een disk-gebaseerd opslagsysteem.

De samenvoeging van Enterprise Vault en Centera tot een e-mailarchiveringsoplossing is weergegeven in figuur 3.

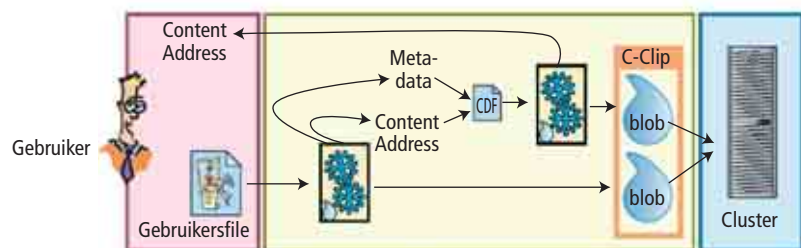
Conclusie

De spectaculair toenemende hoeveelheid e-mail dwingt organisaties ertoe te kiezen voor een bepaalde opslagstrategie. Niet alleen de hoeveelheid e-maildata maar ook nieuwe wettelijke regelingen op het gebied van gegevensopslag maakt het nodig dat het IT-management zich moet gaan bezinnen op een manier waarop content moet worden bewaard, specifiek die van e-mail. Ook aan de tijdsduur waarbinnen e-mail kan worden opgevraagd zullen eisen worden gesteld. Naast een doelmatige archiveringsmethode is een systeem nodig dat de authenticiteit en unieke opslag van e-mail kan waarborgen. In de eerste behoefte kan het softwarepakket Enterprise Vault van KVS voorzien. Voor het veilig opslaan van content in het algemeen, maar met name e-mail, biedt EMC een oplossing met het Centera-opslagsysteem. Gecombineerd bieden beide systemen een betrouwbare archivering van e-mail.

Bram Dons (b.dons@wanadoo.nl) is onafhankelijk IT-adviseur.

Noot

1 Zie white paper 'Protecting Your Business in a Changing Regulatory Climate', Osterman Research, Inc..



Figuur 2 Contentopslag met Centera Content Address Storage (NB. BLOB: binary large object, ofwel bestand).