

De voordelen van gedistribueerde object storage

Er is een trend in met name enterprise ondernemingen om data zoveel mogelijk te consolideren. De achterliggende gedachte is een 'single version of the truth', en het is de reden dat leveranciers van bijvoorbeeld ERP-systemen allerlei functionaliteit toevoegen, zodat het in één database kan worden opgeslagen. Maar dat is helemaal niet efficiënt, zegt Vincent van der Linden van Cloudian. "Gedistribueerd opslaan is de toekomst." *Tekst: Marcel Debets*

Cloudian is een leverancier van storage in de breedste zin van het woord: cloud, on-premise, appliances - het bedrijf levert het allemaal. Maar alle producten en diensten van Cloudian zijn gebaseerd op het principe van S3 object storage.

De basis van object storage is dat de metadata los staan van de daadwerkelijke data. Daarmee wordt het mogelijk om de individuele dataobjecten op een andere manier te benaderen dan data die in een conventioneel filesysteem worden opgeslagen. Zo kun je bijvoorbeeld van een object bepalen hoeveel kopieën je nodig hebt en hoe lang de levensduur van het object is. Het voorkomt daardoor onder andere dat je als bedrijf altijd maar weer meer opslag nodig hebt, want objecten worden uiteindelijk ook weer gewist.

"Dat lijkt in tegenstrijd met het bedrijfsmodel van een storageleverancier zoals wij", zegt Vincent van der Linden, directeur Sales voor de Benelux, het Midden-Oosten en Scandinavië. "Want hoe meer op-

slag je verkoopt, hoe meer omzet. Maar zijn we onze klanten veel beter van dienst door ze een oplossing aan te bieden waarbij data niet langer ongebreideld opgehoopt worden, zodat dat op den duur alleen wordt behouden wat je nodig hebt."

Intelligentie

Hiermee kunnen bedrijven ook gemakkelijker aan de AVG voldoen, want je kunt al bij het verzamelen van de data aangeven wat de levensduur is, of dat wanneer er aan een bepaalde voorwaarde wordt voldaan, de objecten gewist worden. De daadwerkelijke opslag heeft daarbij geen intelligentie nodig, maar de wijze waarop de opslag plaatsvindt uiteraard wel.

Een belangrijk onderdeel van object storage is versiebeheer. Dataobjecten krijgen een unieke 'identificatie'. "Stel dat een object wordt opgevraagd door een applicatie die er vervolgens iets mee doet. Het object

kan dan worden overschreven in de database, het is dan nog steeds hetzelfde object. Maar je kunt het ook als een nieuw object opslaan, en als er iets misgaat kun je teruggrijpen naar een eerdere versie."

Veilig & robuust

Dit heeft ook gevolgen voor het uitwisselen van gegevens tussen applicaties, want objecten kunnen zonder problemen worden gebruikt voor andere doeleinden zonder te kopiëren. Worden de gelezen data bewerkt, pas dan wordt een nieuwe versie opgeslagen. "Dat betekent dat wanneer een applicatie of andere storage uitvalt, alleen de processen die daarvan gebruik maken, niet verder

'We zijn onze klanten veel beter van dienst door ze een oplossing aan te bieden waarbij gegevens niet langer ongebreideld opgehoopt worden, zodat dat op den duur alleen wordt behouden wat je nodig hebt.'

Vincent van der Linden



kunnen. De processen die een van de kopieën gebruiken kunnen gewoon verder”, zegt Van der Linden.

Volgens hem maakt dat de systemen veel veiliger en robuuster. Hij zou dan ook nooit adviseren om alles zoveel mogelijk te consolideren in één systeem. “Hackers die ransomware gebruiken hoeven dan maar een keer binnen te komen en hebben toegang tot alle data om te versleutelen. Met gedistribueerde object storage moet ieder individueel object versleuteld worden en kan behalve de getroffen data de rest van het bedrijf gewoon door.”

Alleen gedeeltelijk toegang

Deze principes vormen ook de basis voor hoe de bedrijfswereld vandaag de dag werkt, zegt Van der Linden. “Bedrijven outsourcen grote delen van hun processen. Denk aan luchtvaartmaatschappijen die de catering uitbesteden. Uitwisseling van gegevens is essentieel, maar je hoeft ze daarbij niet direct toegang te geven tot jouw data. Alleen uitwisselen wat je moet delen is voldoende.”

Cloudian heeft ook een eigen kijk op business continuity & disaster recovery. “Als je alleen naar de data kijkt, is het niet heel erg waarschijnlijk dat de continuïteit in gevaar komt. Je kunt zelf aangeven welke data voor jouw bedrijf essentieel is, en dan het systeem vertellen dat je daar vier of meer replica’s van wilt hebben op evenzoveel plekken. Mocht een applicatie waar een van die kopieën staat uitvallen, dan creëert het systeem automatisch op een andere plek een extra replica. Want de opdracht was immers dat je vier replica’s wilde houden.”

Amazon S3 – het dataopslagprotocol voor de toekomst

Het S3-protocol van Amazon is een protocol voor de opslag van data in de cloud. Het staat voor Simple Storage Service en is inmiddels de facto de standaard geworden. Conventionele opslag – bijvoorbeeld de harddisk in de computer – maakt gebruik van blocks en files, waarbij de kleinste eenheid al begint bij 512 bytes. Grotere bestanden moeten dan eerst opgebroken worden in kleinere stukjes, om vervolgens verdeeld te worden over de vrije plekken in het opslagsysteem. Voor grotere bestanden is dit niet efficiënt. Vandaar dat, met name voor databases, object storage is bedacht. De data wordt gescheiden van de metadata (het deel dat beschrijft wat er in de data staat), en een enkel dataobject kan daarbij wel 5TB omvang hebben. S3 maakt gebruik van dit principe, met als extra dat het protocol gebaseerd is op standaard HTTP-opdrachten voor het opslaan, terughalen en verwijderen. Amazon gebruikt S3 zelf al sinds 2006 voor zijn cloudoplossingen. Het grootste deel van het protocol is open-source, en inmiddels door vrijwel alle leveranciers van storage omarmd.

Amazon S3

Cloudian was er al vroeg bij en ondersteunt het Amazon S3-protocol sinds eind 2011 (zie kader). Het bedrijf ondersteunt alle functionaliteiten in de API is daarmee de meest uitgebreide partner van Amazon. Crux daarbij is dus dat het niet uitmaakt waar de data wordt opgeslagen, geheel in lijn met het S3-protocol van Amazon. Dat kan in elke willekeurige database, in de cloud-variaties zoals de Cloud, Hybrid-Cloud en Private-Cloud of lokaal (eigen ICT omgeving/on-premise), op een Storage-server “appliance” enzovoort, om de klant maximale flexibiliteit te bieden. Die flexibiliteit gaat heel ver. Wie voor een on-premise systeem gebruik wil maken van de intelligentie van de Cloudian software-defined storage (SDS) kan kiezen voor zijn eigen hardwareleverancier voor de opslag, waarbij Cloudian zorgt voor de SDS. “We zien soms ook wel dat klanten liever alles van een leverancier hebben, daarom leveren wij ook storage appliances, waarbij we zowel de hardware als de software verzorgen.”

Data verrijken

Van der Linden ziet ook nog een

ander probleem met de enorme groei van data. “Tachtig procent van alle data is ongestructureerd. Dat zijn objecten waarbij geen metadata beschikbaar zijn. Met andere woorden, zonder nadere analyse kun je er niks mee.” Daarom werkt Cloudian nauw samen met veel leveranciers van softwareoplossingen voor back-up, voor datamanagement, maar ook van reporting tools. Sommige van deze software kan ervoor zorgen dat ongestructureerde data gestructureerd wordt opgeslagen, door metadata en de daadwerkelijke data op te splitsen. Zo kunnen datamanagementtools ongestructureerde data analyseren en er structuur in aanbrengen. “Door die nauwe samenwerking krijgt het systeem nog meer intelligentie en de data meer waarde.” Ook wordt daardoor de hoeveelheid ongestructureerde data verminderd en dat is weer van belang voor het herstellen na dataverlies. “Wanneer data gestructureerd zijn, kun je per object bekijken of het van belang is. Data die al langer inactief zijn, kun je dan in een goedkopere ‘tier’ opslaan. En uiteindelijk hoef je geen petabytes aan data te herstellen. Dat is een van de mooiste dingen aan het leveren van object storage: je bent altijd bezig om data voor je klanten te verrijken.” ■