

Markt zet eerste grote stappen op weg naar volwassenheid

# Data in de Cloud

Rob Swinkels, Ronald van Aken en Michael Widjaja

**De Cloud biedt een nieuw scala van mogelijkheden aan als het gaat om het implementeren van IT-oplossingen. Het kan het meest eenvoudig omschreven worden als de 'publiekelijke' beschikbaarheid van IT resources en services capaciteit die zich kenmerken door virtualisatie en verrekening op gebruik. Er bestaan verschillende varianten van hoe organisaties deze cloud services kunnen gebruiken.**

Dit artikel focust voornamelijk op de database in de Cloud en beschrijft de mogelijkheden, uitdagingen en de verschillende toepassingen die momenteel beschikbaar zijn.

De Cloud is niet nieuw. Wel heeft het de laatste jaren extra aandacht gekregen door de technologische ontwikkelingen op het gebied van beheersbaarheid, flexibiliteit en integratiemogelijkheden. Enerzijds brengen nieuwe virtualisatie- en dynamische allocatietechnologieën speelruimte voor het flexibeler delen van applicatieplatformen en hardware. Anderzijds geven de ontwikkelingen op gebied van applicatie-integratie (zoals service georiënteerde architecturen (SOA), webservices, HTTP- en XML-protocollen), oplossingen die gemakkelijker zowel binnen als tegelijk buiten de bedrijfsmuren kunnen werken.

Traditioneel vergden complexe IT-implementaties grote initiële investeringen voor de aanschaf van de hardware- en software-infrastructuur. Daarbij kwam vaak ook nog een aanzienlijke doorlooptijd voor de installatie. De Cloud daarentegen biedt elastische capaciteit op basis van 'publieke' hardware en/of software infrastructuur en werkt op een aantal basisprincipes zoals geschetst in afbeelding 1:

1. *Pay-as-you-go*. Met de Cloud deelt men reeds geïnstalleerde en draaiende infrastructuur met andere gebruikers en het verbruik wordt op basis van bijvoorbeeld *subscription-base* verrekend. Grote initiële investeringen en de traditioneel lange bestel- en doorlooptijden zijn hierdoor niet meer vereist;
2. *Grow & Shrink*. De flexibiliteit van de Cloud geeft de mogelijkheid om makkelijk en snel extra capaciteit bij te schalen of af te bouwen als nodig is. Zo kan bij ingelaste marketing-campagnes snel extra IT resourcecapaciteit bijgeschakeld worden of in andere gevallen kunnen operationele kosten verlaagd worden als minder capaciteit nodig is dan verwacht;
3. *Illusion of infinite capacity*. De Cloud biedt in principe ook een ongelimiteerde pool van resources, processing, computing en/

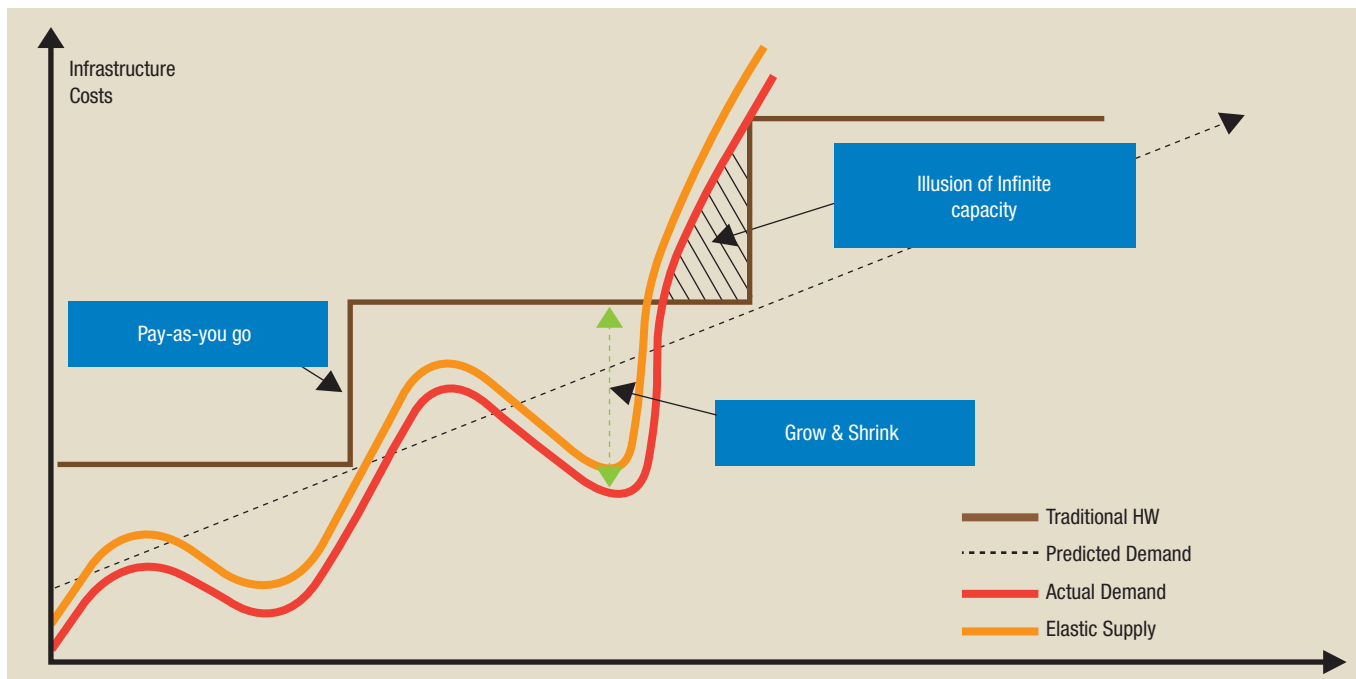
of storage aan, die extra flexibiliteit geeft bij bijvoorbeeld het verwerken van piekbelasting.

## Nieuwe mogelijkheden

De Cloud biedt bedrijven dus meer keuzes in het gebruik van IT-infrastructuur over de grenzen van de organisatie heen. De mogelijkheden lopen uiteen van het gebruik van basale server hardware cloudcapaciteit tot en met het volledig uitbesteden van de oplossing in de Cloud. Verschillende mogelijkheden van de Cloud zijn te zien in afbeelding 2:

- Process/Industry Clouds zijn delen van processen die zo gestandaardiseerd zijn voor een specifieke industrie dat ze vaak als de facto beschouwd worden. Voorbeelden zijn betalingen via iDeal, PayPal of ticketboekingen;
- Software-as-a-Service (SaaS) staat ook bekend als applications-as-a-service en is een vorm waarbij de gehele applicatie op de Cloud draait;
- Platform-as-a-service Clouds bieden een ontwikkelplatform aan waarmee applicaties in een Cloud kunnen worden gebouwd. Hiermee ontstaat dus de mogelijkheid om maatwerk aan te bieden en/of te integreren met bestaande oplossingen;
- Een recente ontwikkeling in de Cloud is de Database-as-a-Service (Daas) waar de database in de Cloud draait. De meeste grote databasleveranciers hebben een databaseversie die geschikt is om op een cloudinfrastructuur te draaien. Daarnaast zijn er DaaS proposities waarbij de database en ook het beheer daarvan als dienst worden aangeboden en afgerekend;
- Infrastructure-as-a-service Clouds leveren een serverinfrastructuur die op basis van actuele behoefte CPU en/of andere resources beschikbaar stelt.

Combinaties van de cloudlagen (afbeelding 2) zijn ook mogelijk en zijn weergegeven in afbeelding 3. In principe kan een cloud-



**Afbeelding 1:** Elastische cloudcapaciteit.

oplossing zowel met een infrastructuur bij dezelfde leverancier worden aangeboden alsook door gebruik te maken van de Cloud van een andere leverancier. Deze laatste optie komt in de praktijk echter weinig voor, gezien de extra risico's die het oplevert op het gebied van performance, beveiliging en totale beschikbaarheid naar de gebruiker toe.

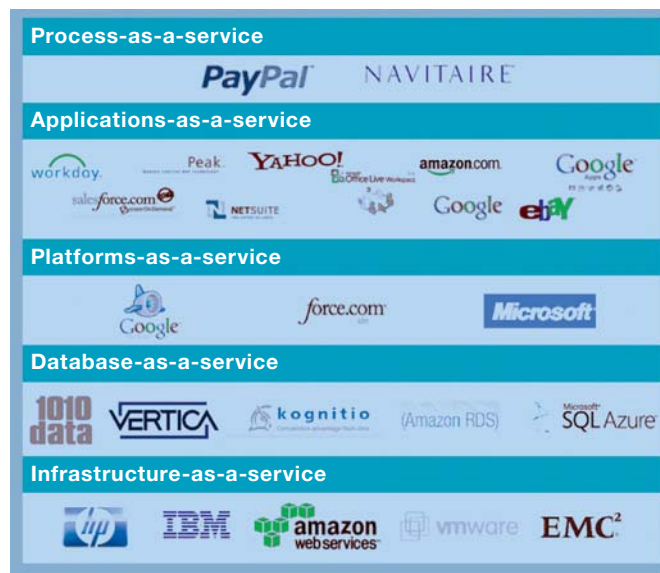
Het vervolg van dit artikel beschrijft met name het concept Database-as-a-Service aan de hand van de volgende vragen: wat zijn de voordelen van het gebruik van DaaS; welke uitdagingen zijn er op dit gebied; hoe werkt het uitwisselen van data binnen de Cloud en tussen Clouds; welke typen DaaS zijn er welke vorm past het beste in welke situatie; welke leveranciers zijn er?

## Voordelen

Voordat een organisatie een DaaS oplossing zou willen gebruiken, ligt er echter de vraag waarom een bedrijf *überhaupt* de data laag buiten de eigen organisatie zou willen onderbrengen? Vele organisaties beschouwen hun databases en vooral de gegevens hierin als cruciaal voor hun bedrijfsvoering en zullen hierdoor terughoudend zijn met het buiten de deur plaatsen ervan, maar er zijn toch verschillende redenen om dit te doen:

- De eigen IT-afdeling kan (of wil) de database niet zelf beheren. Dit kan bijvoorbeeld omdat er gebruik gemaakt wordt van een database van een andere leverancier dan de gebruikelijke standaard. Bij gebruik van DaaS is het dan niet noodzakelijk om die specifieke expertise en skills te trainen of aan te trekken;
- De eigen IT-afdeling heeft niet voldoende IT-capaciteit om de database te beheren. In het geval dat de database grote volumes gegevens bevat en de eigen IT-afdeling onvoldoende opslagcapaciteit of processorcapaciteit heeft om de bewerkingen uit te voeren kan DaaS een oplossing bieden;

- Doorlooptijd is cruciaal. Voor projecten waar korte doorlooptijden cruciaal zijn, brengt DaaS een oplossing omdat het onmiddellijk ingezet kan worden;
- Delen van data. In steeds meer organisaties wordt het delen van gegevens met derden steeds belangrijker en DaaS geeft de mogelijkheid om bij een derde partij de data te beheren. Bovendien is toegang van buiten geregeld aangezien dit inherent is aan het model;
- Total-Cost-of-Ownership (TCO). In bepaalde gevallen kan het gebruik van DaaS of een cloud database de TCO verlagen. Een voorbeeld is waar een database slechts gedurende een korte periode nodig is.



**Afbeelding 2:** Lagen in de Cloud.

## Karakteristieke toepassingen

In een aantal situaties is een DaaS oplossing een uitstekend alternatief ten opzichte van de traditionele database op eigen *premise*. Kleine organisaties die niet zelf een database kunnen of willen beheren kunnen met DaaS hun voordeel doen. Dit kan voordelen opleveren in termen van kosten, doorlooptijd en vergemakkelijkt ook het samenwerken met derde partijen.

Daarnaast kan DaaS een voordeel bieden bij grotere volumes van ontwikkelomgevingen met vaak lage datagebruiksvolumes gecombineerd met weinig CPU-verbruik. Vaak dienen deze ontwikkelomgevingen snel ter beschikking gesteld te worden en zijn ze vaak tijdelijk van aard. Een DaaS biedt uitkomst aangezien de omgeving snel kan worden opgeleverd en even zo snel weer kan worden stopgezet.

Ook voor korte en/of tijdelijke projecten kunnen kosten voor het opzetten en beheren van een database binnen organisaties snel oplopen, zodat het gunstiger is om deze als een databasedienst te betrekken. DaaS is bijzonder geschikt voor Proof-of-Concept (PoC), innovatieve projecten of bijvoorbeeld conversieprojecten. Hierbij kan snel iets uitgetoetst worden, waarna een eventuele volgende versie binnen de organisatie verder ontwikkeld en beheerd wordt.

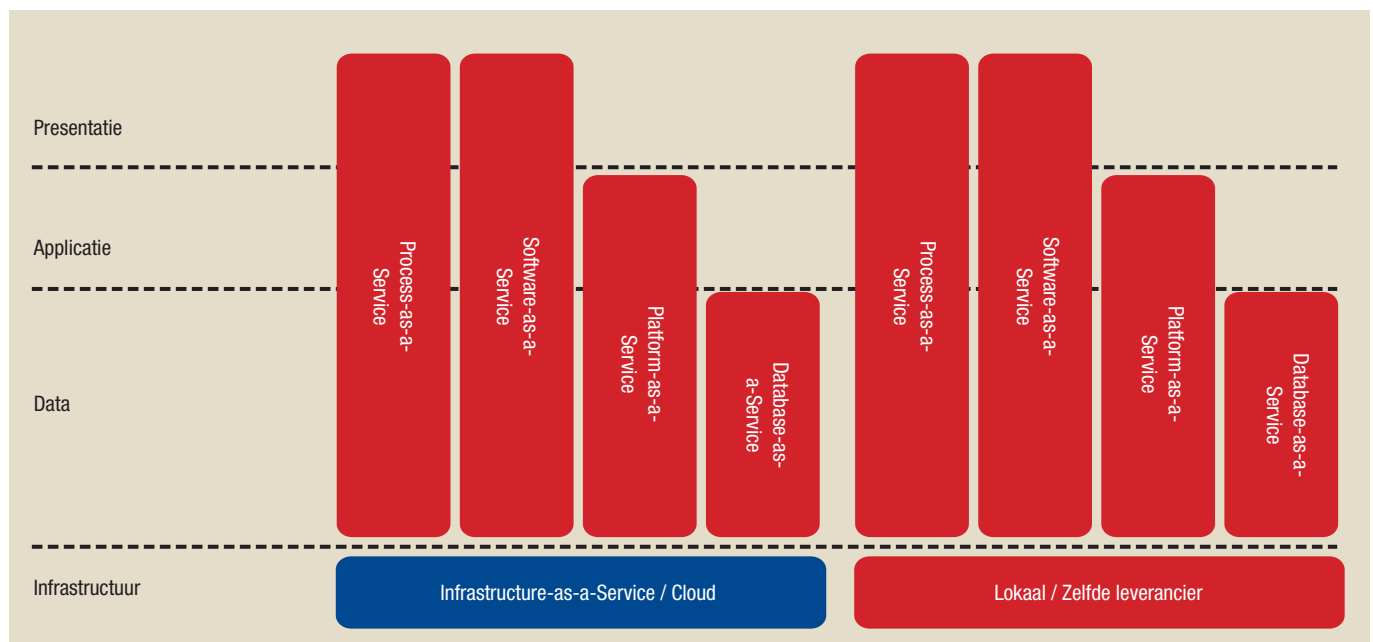
## Uitdagingen

Het is duidelijk dat DaaS vele kansen biedt voor bedrijven. Gebruikers moeten echter ook rekening houden met diverse uitdagingen, zowel op technologisch als wel op organisatorisch vlak:

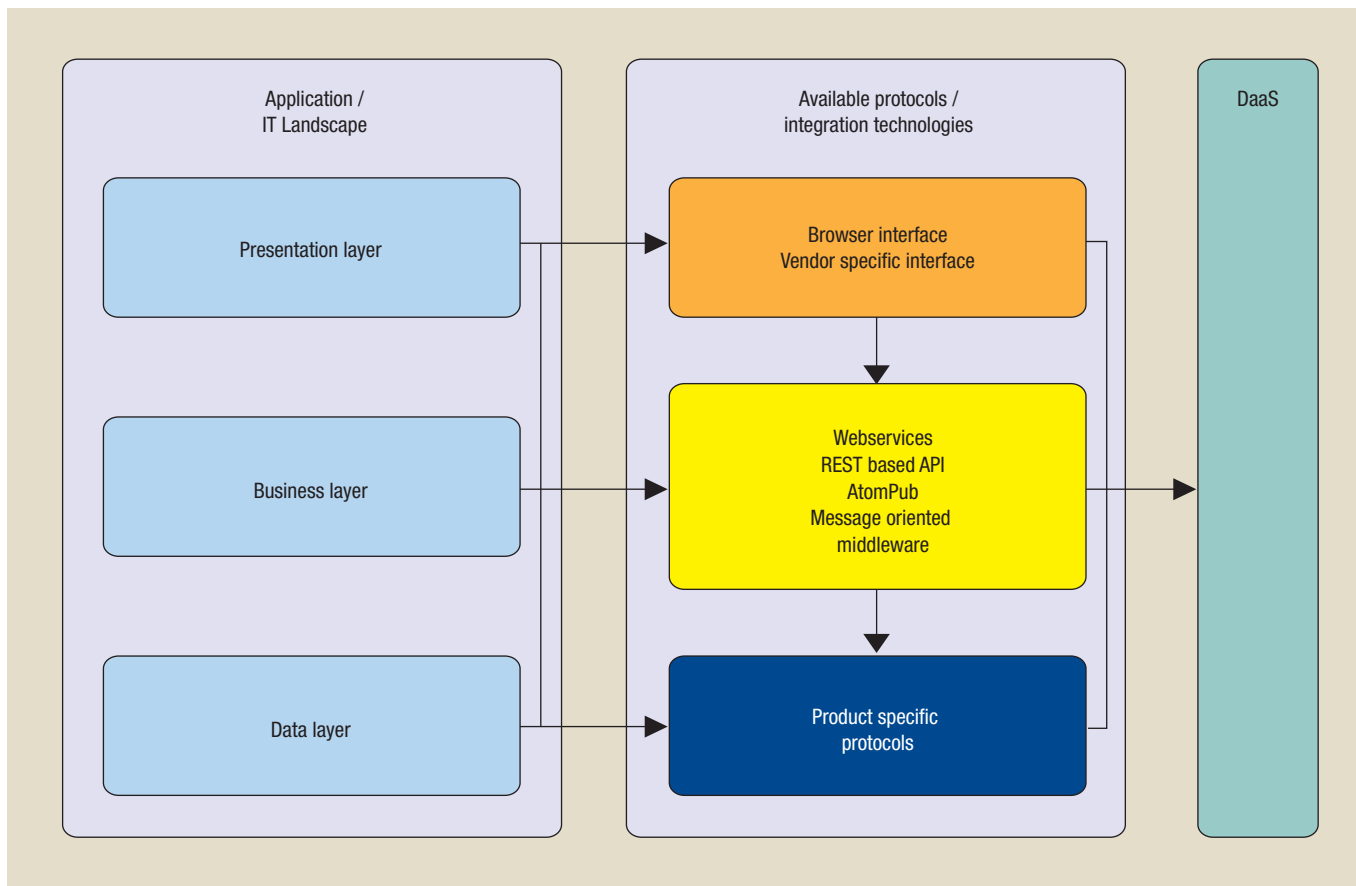
- **Latency.** Een technische uitdaging voor DaaS is latency. Het gebruik van DaaS voor databases met een hoog volume aan datatransfer tussen applicaties en database dient goed overwogen te worden. Afhankelijk of de leverancier van een lokale of

cloudinfrastructuur gebruik maakt, moet goede afstemming plaatsvinden;

- **Kosten.** Leveranciers berekenen kosten voor de hoeveelheid data die in of uit de Cloud gaat en dit dient zeker in ogenschouw genomen te worden. Het gebruik van DaaS kan bijvoorbeeld voordelig zijn in een ontwikkelomgeving, waarna de uiteindelijke productieomgeving in de eigen organisatie beheerd wordt;
- **Standaarden.** Afbeelding 4 laat zien dat er verschillende standaarden zijn voor databasetoegang in de Cloud en moet duidelijk afgestemd zijn met de gebruikers. Voor 'eenvoudige' implementaties zoals SimpleDB en BigTable wordt steeds meer gebruikt gemaakt van REST. Een ander initiatief is AtomPub, wat een specificatie is voor het transparant benaderen van relationele databases. Het AtomPub protocol is ontwikkeld door IETF en wordt ondersteund door onder andere Microsoft en Google;
- **Vertrouwen.** Ook op het organisatorische gebied liggen er diverse uitdagingen. Verreweg de belangrijkste is die van het vertrouwen van klanten in de producten en diensten. Een goed voorbeeld hiervan is Salesforce die een lange tijd heeft moeten aantonen dat zijn dienstverlening veilig, schaalbaar en betrouwbaar is, alvorens het aantal klanten fors toenam. Dit patroon zal ook voor andere modellen gelden. Daarnaast moeten organisaties ervaring opdoen met zaken als Service Level Agreements, licentiebeheer en intellectueel eigendom in samenhang met DaaS en/of cloud database-oplossingen;
- **Juridische aspecten.** Bij de Cloud blijven de juridische aspecten een heet hangijzer en dat geldt zeker voor data. Welke jurisdictie is van toepassing op de data in de Cloud? Moet men zich zorgen maken om bijvoorbeeld de USA Patriot Act, waarbij gegevens die zich op Amerikaans grondgebied bevinden



**Afbeelding 3:** Cloud biedt verschillende mogelijkheden.



**Afbeelding 4:** Gebruikerstoegang tot DaaS.

door de overheid mogen worden ingezien? Zonder duidelijke afspraken is bij het gebruik van de Cloud voor een klant niet transparant waar de servers en data zich bevinden en dus welke wetgeving van toepassing is.

### Kiezen van DaaS model

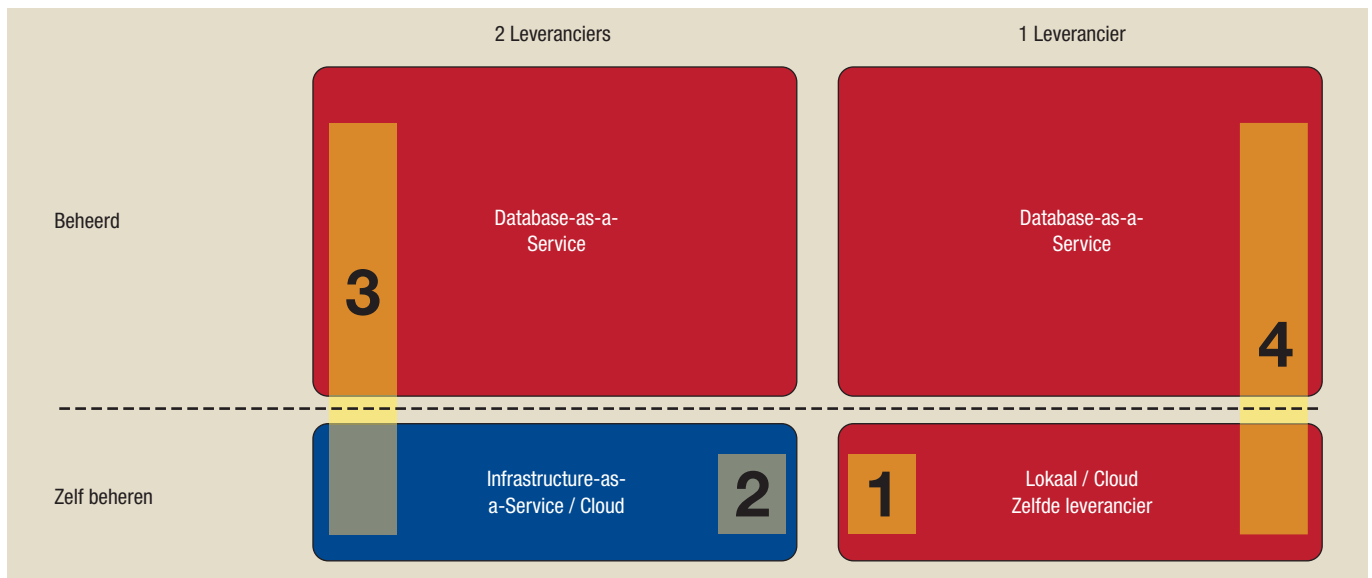
De volgende stap is om te bepalen welke DaaS variant het beste past bij een organisatie. Hierbij zijn de volgende mogelijkheden te onderscheiden, weergegeven in afbeelding 5:

## Portability en delen van data in de Cloud

Wanneer een bedrijf gebruik gaat maken van verschillende clouds om gegevens op te slaan, ontstaan ook nieuwe uitdagingen. Een belangrijk aspect is om een architectuur te realiseren waarmee data op een eenduidige manier opgehaald kunnen worden en tegelijkertijd ervoor zorgen dat de architectuur aansluit op de eisen vanuit de business voor de desbetreffende data. Om de toegang op een efficiënte manier te realiseren moet de cloudleverancier een 'single point of entry' aanbieden waarmee informatie kan worden opgehaald. Dit toegangspunt kan het beste gezien worden als een generieke data access service die toegang biedt tot alle mogelijke informatiestromen die in de Cloud zijn opgeslagen. Het delen van data over verschillende clouds is ook een aspect dat hier meespeelt, aangezien bedrijven vanwege beperkingen en specialisaties van cloudleveranciers verschillende clouds gebruiken voor het opslaan van de informatie. Het is daarom van belang dat informatie over de verschillende clouds op éénduidige wijze opgehaald en uitgewisseld kan worden. Bij het delen en uitwisselen van data over verschillende clouds verdienen de

volgende zaken zeker de aandacht:

- Biedt de leverancier voldoende genoeg fysieke mogelijkheden om te integreren met andere clouds?
- Welke technologische standaarden worden ondersteund door de leverancier zoals bijvoorbeeld webservices, message oriented middleware, transportprotocollen zoals FTP en e-mail? Daarnaast is het ook belangrijk welke security standaarden ondersteund en verwacht worden vanuit de organisatie.
- Qua performance moeten gegevens snel genoeg uitgewisseld kunnen worden over de verschillende clouds. Netwerk- als ook processorwachtijden spelen een belangrijke rol hierbij. Typische valkuil hierbij is dat een bepaalde cloudleverancier zijn servers verhuist naar een ander geografisch gebied waardoor de latency van data-uitwisseling omhoog gaat!
- Qua beschikbaarheid moet goed afgestemd worden dat de aaneenschakeling van verschillende clouds nog steeds voldoet aan de verwachte beschikbaarheid van de complete end-to-end keten.



**Afbeelding 5:** DaaS implementatiemodellen.

1. Infrastructuur cloudleverancier met eigen cloud database. Zowel Google als Amazon leveren een simpele database in hun cloud (BigTable respectievelijk SimpleDB). Deze volstaan voor een behoefte aan niet-relatieve opslag van gegevens (lijsten), maar er zijn ook relationele alternatieven van andere leveranciers. Beheer van de gegevens en lijsten dient door de gebruiker zelf uitgevoerd te worden;
2. Database met cloudversie op infrastructuurcloud. Grote databaseleveranciers zoals Oracle en MySQL kunnen op een cloudinfrastructuur gebruikt worden of hebben hiervoor een speciale versie. In deze gevallen heeft de klant de keuze over welke cloudinfrastructuur gebruikt wordt en is dus sprake van twee leveranciers: één van de database en één voor de (cloud) infrastructuur. Bij deze oplossing dient nog steeds de gebruiker het databasebeheer zelf te verzorgen;
3. DaaS op infrastructuurcloud. Eén stap verder is om een cloudversie van een database als DaaS aan te bieden, zoals Amazon die met hun Relational Database Service (RDS) een versie van MySQL aanbieden die in hun cloud draait. Vertica biedt een dienst waarbij ook gebruik gemaakt wordt van de Amazon cloud. De service wordt verzorgd door de database- en infrastructuurcloudleveranciers;
4. DaaS op eigen infrastructuur. De laatste mogelijkheid is waarbij de leverancier zelf de nodige infrastructuur- en database cloud exploiteert en daarop voor meerdere klanten databases draait. De klant kan dan vaak kiezen of deze gebruik maakt van eigen of gedeelde hardware. Voorbeelden van deze categorie zijn Kognitio en 1010data. Microsoft biedt met SQL Azure een speciale cloudversie aan van SQL Server in hun eigen cloud (Azure Platform).

## Leveranciers

Spelers binnen data cloudoplossingen zijn bedrijven als Oracle, Microsoft, RainStor, Greenplum, Postgres EnterpriseDB, Vertica,

Amazon, Google en Yahoo. Zij bieden oplossingen die variëren van het gebruik van hun producten in een cloud tot meer gespecialiseerde oplossingen (zoals specifieke analyse- en datamining-functies bij Microsoft SDDS). De leveranciers werken over het algemeen samen met bedrijven die al de infrastructuur hebben om de Cloud aan te bieden zoals Microsoft en Oracle, die beide samenwerken met Amazon in het aanbod van hun databases. Naast de grote bekende softwaregiganten zijn er daarnaast enkele niche-aanbieders met specifieke oplossingen voor het opslaan, manipuleren en of bewerken van informatie. Voorbeelden van deze laatste groep zijn onder andere QuickBase en TrackVia. Ook een aantal niet-relatieve databases heeft momenteel een duidelijke positie in de Cloud, zoals onder andere HyperTable (open source variant van Google BigTable), HBase welke wordt gebruikt door Hadoop en tot slot Project Cassandra, dat initieel door Facebook is ontwikkeld maar momenteel beschikbaar is als open source. De niet-relatieve databases zijn vaak goed te gebruiken om documenten in op te slaan en in webomgevingen te gebruiken. Daarnaast zijn de meeste niet-relatieve databases beschikbaar als open source.

## Tot Slot

Database-as-a-Service (DaaS) clouds zijn nog erg in beweging of zelfs alleen nog als bètaversie beschikbaar. Specifieke toepassingen zullen zich de komende tijd meer uitkristalliseren, maar eerst moeten verschillende uitdagingen binnen organisaties op zowel technisch als op organisatorisch vlak overwonnen worden. DaaS biedt verschillende implementatiemogelijkheden aan en met het groeiende aantal DaaS leveranciers heeft de DaaS markt eigenlijk al haar eerste grote stappen op weg naar volwassenheid gezet.

## Rob Swinkels, Ronald van Aken en Michael Widjaja

Rob Swinkels is Manager, Ronald van Aken Principal Consultant en Michael Widjaja Partner, allen werkzaam bij Accenture.