

Over Cloud en Big Data

Uitdagingen en onontkoombaarheid

Dr. Geert-Jan van Bussel, Hogeschool van Amsterdam en Van Bussel Document Services

Vandaag de dag is er geen enkele leverancier van informatietechnologie die cloud computing niet heeft omarmd. Die term werd (voor zover ik kan nagaan) in 2006 voor de eerste keer publiekelijk gebruikt door Google's Eric Schmidt, die daaraan toevoegde dat hij het idee had dat niemand begreep waar het om ging, maar dat het wel enorme mogelijkheden bood. Cloud computing heeft een enorme ontwikkeling doorgemaakt, vooral omdat individuele computergebruikers het al vanaf het begin enthousiast omarmden en het gebruik ervan stimuleerden in hun dagelijkse werk.

De introductie van de iPhone door Apple in 2007 en de daarop volgende snelle opmars van de smartphones heeft 'de cloud' volledig in het dagelijks persoonlijke en zakelijke leven geïntegreerd. Niemand wilde de markt van smart devices, mobile apps en cloud services missen. Interactieve apps, diensten en sociale media (zoals Facebook, Gmail, Google Apps, Flickr, YouTube, iTunes, eBay, Evernote, TripAdvisor, Twitter, LinkedIn, Instagram, en honderden andere) zijn voorbeelden. Ze werden en worden breed geaccepteerd en gebruikt. Die acceptatie begon al toen die applicaties en diensten nog als Web 2.0. werden aangeduid, door Tom O'Reilly in 2005 uitgebreid beschreven. In diens beschrijving hoeft alleen de term 'web 2.0.' door 'cloud computing' te worden vervangen om het toepasselijk te maken. De enorme groei van het aantal data dat in de cloud wordt verwerkt en opgeslagen werd de basis voor de eveneens snelle ontwikkeling van 'Big Data' (en dan met name Big Data Analytics).

De cloud is in korte tijd uitgegroeid van iets 'uitzonderlijks' tot iets 'gewoons'. Uitdagingen en risico's zijn er voldoende, maar het lijkt erop dat de cloud onontkoombaar is geworden. De vraag om mijn licht eens te laten schijnen over de cloud (en daarbij Big Data niet te vergeten) bracht mij er toe om de meest enthousiaste beschrijvingen over deze verschijnselen nog eens ter hand te nemen: Nicholas Carr's *The Big Switch* en Victor Mayer-Schoenberger en Kenneth Cukier's *Big Data*.

VERSCHILLENDE CLOUD-MODELLEN

Cloud computing toont zich in verschillende vormen. De meest herkenbare vorm is die van de 'public cloud', die alle kenmerken vertoont die zijn genoemd. Een public cloud maakt gebruik van een publiek netwerk, zoals het internet, waarbij alle afnemers van de dienst tegelijkertijd gebruik maken van dezelfde, gedeelde infrastructuur. De public cloud wordt vooral gebruikt

door individuele personen, die voldoende hebben aan de beveiliging zoals die door de public cloud geboden wordt. Diensten in de public cloud zijn bijvoorbeeld de opslag van data en documenten, online software voor tekstverwerking, presentaties en spreadsheets en webmail. Een 'private cloud' biedt hetzelfde, maar dan voor een specifieke gebruiker. Dat biedt voor die specifieke gebruiker het voordeel van betere beveiliging en controle. Een private cloud maakt over het algemeen gebruik van veilige, versleutelde verbindingen. De kosten voor het operationaliseren van een private cloud zijn (vele malen) hoger dan die van een public cloud. Die kostenoverwegingen zijn de reden geweest voor het ontstaan van de 'hybrid cloud', een omgeving die gebruikt maakt van een samenstel van 'public' en 'private' cloudcomponenten.

BUSINESS IN DE CLOUD

Alle IT-leveranciers bieden clouddiensten als een onontkoombaar component in hun producten en diensten. Zelfs een oor-

SaaS kunnen worden aangeboden. Een tweede model is IaaS, Infrastructure-as-a-Service. Hierbij specificeert een klant welke (hardware) systeembronnen nodig zijn. De leverancier bepaalt vervolgens zelf hoeveel gevirtualiseerde servers en andere hardware daarvoor nodig is. Een klant neemt een abonnement op het verbruik van virtuele systeembronnen. Een voorbeeld van een dergelijke leverancier is Terremark. En ten slotte kennen we PaaS: Platform-as-a-Service. Bij PaaS wordt naast de infrastructuur ook de software-infrastructuur als dienst aangeboden. Microsoft Azure is hiervan het beste voorbeeld. Bij veel clouddiensten overlappen de verschillende modellen elkaar. Bij IaaS en PaaS wordt de afhankelijkheid van de dienstverlener (erg) groot.

Nicholas Carr stelde terecht (zij het enigszins overdreven), dat de cloud het businessmodel van traditionele leveranciers hard raakte. Het bestaande model immers was gebaseerd op de verkoop van 'proprietary' hard- en software; dat model komt zwaar

“DE CLOUD IS IN KORTE TIJD UITGEGROEID VAN IETS ‘UITZONDERLIJKS’ TOT IETS ‘GEWOONS’. UITDAGINGEN EN RISICO’S ZIJN ER VOLDOENDE, MAAR HET LIJKT EROP DAT DE CLOUD ONONTKOOMBAAR IS GEWORDEN.”

spronkelijk verstokte tegenstander als Larry Ellison van Oracle kon er uiteindelijk niet omheen. Amazon, Microsoft en Google zijn de Big Three Cloud Computing Companies in 2015. Deze bedrijven operationaliseren de grootste cloudinfrastructuren, *“that provide resizable compute capacity in the cloud”*. Allerlei internetbedrijven maken (tegen betaling) van deze infrastructuren gebruik om hun bedrijfsactiviteiten uit te voeren en bieden via mobile apps (of full-fledged webapplicaties) (gratis of betaalde) diensten aan. Wereldwijd kunnen individuele computergebruikers of bedrijven vervolgens deze apps en webapplicaties gebruiken en hun data 'in de cloud' opslaan, verwerken, beveiligen en distribueren. De concurrentie tussen de aanbieders van deze cloudinfrastructuren is zo groot dat de gebruikers ervan (de bovengenoemde internetbedrijven die de leveranciers zijn van diensten) zulke lage prijzen doorberekend krijgen, dat het economisch onverantwoord lijkt te gaan worden.

Er zijn drie businessmodellen ontstaan, waarmee internetbedrijven proberen eindgebruikers aan zich te binden. Zo kennen we SaaS, Software-as-a-Service. SaaS-leveranciers bieden een clouddienst waarbij hun gebruikers specifieke softwaretoepassingen kunnen benaderen via het internet. Er zijn vele voorbeelden, Salesforce is er een van. Leveranciers bieden ook document- en recordmanagementtoepassingen via zo'n model, zoals bijvoorbeeld het Nederlandse Decos. Zelfs eDepots zouden via

onder druk te staan doordat het 'utility' model steeds meer ingang vindt, waarbij leveranciers diensten aanbieden via het internet. Hierdoor drogen de winsten van de traditionele verkopers op. Wat Carr niet in zijn overwegingen meenam is dat juist deze partijen (dankzij hun investeringskracht) in staat zijn de cloud te domineren. Hij voorzag niet dat ook het utility-denken uiteindelijk draait om het maken van winst en dat leveranciers klanten zullen opsluiten in proprietary-systemen, die het moeilijk maken om nieuwe clouddiensten te gaan gebruiken. Free Software Foundation-goeroe Richard Stallman zag dat terecht als een gevaar voor de klant.

VOORDELEN

Maar het gevaar van te grote afhankelijkheid (dat ook in het proprietarymodel bestaat) doet niets af aan het feit dat juist de kenmerken van cloudomgevingen zeer aantrekkelijk gevonden worden door vele bedrijven. Het vooruitzicht van een structurele verlaging van kosten is voor die bedrijven de belangrijkste drijfveer om afscheid te nemen van eigen hard- en software infrastructures. Er zijn vooraf geen investeringen nodig op basis van inschattingen over de benodigde capaciteit die bijna altijd of veel te hoog of veel te laag is. Capaciteit wordt afgenomen op basis van behoefte. Dat sluit aan op inspanningen om kosten meer in overeenstemming te brengen met gebruik en om structurele kosten voor faciliteiten te verminderen. Daarnaast



veranderen markten sneller, waardoor organisaties zich steeds opnieuw moeten aanpassen aan veranderende omstandigheden. Eigen IT-infrastructuren zijn niet zo flexibel en het kost grote investeringen en veel tijd om voortdurend aanpassingen te doen. De kans dat de kwaliteit van de dienst in de cloud beter, sneller en betrouwbaarder is, speelt een belangrijke rol. Voor een dienstverlener in de cloud is de te leveren dienst immers de belangrijkste bedrijfsactiviteit, voor een gebruikende organisatie is dat niet zo: IT is een middel om het belangrijkste bedrijfsdoel te bereiken.

BIG DATA

Een van de gevolgen van de snelle acceptatie van de cloud is een enorme toename van het aantal data. De IT-mogelijkheden om betrouwbare en onbetrouwbare data van elkaar te kunnen scheiden, om data te beveiligen en om privacy te bewaren, komt zwaar onder druk te staan. De datamassa groeit vooral door de toenemende publicatiemogelijkheden die het internet biedt, de exploderende elektronische communicatie door de inmiddels revolutionaire impact van sociale media, de mobiele revolutie, de trend naar volledig digitaal werken (het 'Nieuwe Werken'), de enorm toenemende e-commerce en de grootschalige digitalisering van cultureel erfgoed zoals film, muziek, kunst, beeld en kaartmateriaal.

Big Data komen tot stand middels dataficatie: het omzetten van de aspecten van ons dagelijks leven naar data. Google slaat ons surf- en zoekgedrag op en zet het om in data. Mobiele telefoons met geolocatie registreren waar we geweest zijn en zetten dit om in data. Vloeren met druksensoren kunnen onze bewegingen dataficeren. Camera's registreren wat er gebeurt in de omgeving en dataficeren dit. Enzovoorts. Door het combineren van deze data kunnen ze voor totaal andere doeleinden gebruikt worden als waarvoor ze bedoeld zijn, waardoor problemen kunnen ontstaan met de compliance aangaande de bestaande wetgeving inzake privacy en persoonlijke gegevens.

Het feit bijvoorbeeld dat de kwaliteit en de context van Big Data nauwelijks belangstelling krijgen kan funest uitwerken. Juist de kwaliteit en de context van data zijn van groot belang om tot correcte en relevante interpretaties van de resultaten van data analyse te komen. Een data analyse die vooral in cloudomgevingen zal worden uitgevoerd, omdat Big Data de clustering van software- en hardwarebronnen in de cloud benodigd om te kunnen 'bestaan'.

DE UITDAGINGEN VAN DE CLOUD

Cloud computing is een andere leveringswijze van IT. Elektronische netwerken koppelen geografisch verspreide datacenters. Samen vormen zij in wisselende samenstelling één virtuele computerfabriek. Dat model heeft grote consequenties voor technologie- en informatiemanagement. Waar en wanneer wordt welke bedrijfsinformatie met welke software verwerkt? Welke partijen vervullen in het verwerkingsproces een rol? Welke regelgeving is van toepassing? Welke metadata

worden toegevoegd? Is de kwaliteit en de context van de data gewaarborgd? Hoe wordt met persoonsgegevens om gegaan? Op welke locaties vindt, al dan niet tijdelijk, gegevensopslag plaats? In welke formats? Wie is waarvoor verantwoordelijk en aansprakelijk? Transparantie, waarborgen en zekerheden zijn onvermijdelijk.

Cloud computing kent dus vele uitdagingen. In het oog springende uitdagingen zijn de afhankelijkheid van netwerkverbindingen en cloudleveranciers. Is de netwerkverbinding snel genoeg qua bandbreedte en reactiesnelheid? Is de verbinding continue? Hoe groot is de kans op wegvallen van de verbinding? Welke cloudleverancier biedt de gewenste betrouwbaarheid qua bedrijfscontinuïteit en gebruikte technische infrastructuur? Als grote cloudpartijen als Google, Amazon, Dropbox, Facebook, Microsoft en Twitter met storingen te maken hebben, hoe houden kleinere partijen zich dan over-eind? Systeembeheerders hebben zorgen om de beveiliging: een grote cloud is een aantrekkelijk doelwit voor hackers en crackers. DVO's (Dienstverleningsovereenkomsten) zijn uitgevonden om deze problemen aan te pakken. Ze leveren schadevergoedingen op, maar ze lossen problemen zelf niet op.

Een andere uitdaging is het feit dat er nauwelijks standaarden zijn voor deze nieuwe vorm van IT-outsourcing. Zo is overstappen van de ene naar de andere cloud-aanbieder nog onbekend terrein. Migreren, duurzaamheid van data en 'eigenaarschap' komen in het vocabulaire van de cloudleveranciers nauwelijks voor. Dit wordt nog versterkt door de ethische en privacyproblematieken van Big Data, die veelal juist ergens in dat cloudplatform worden opgeslagen, verwerkt, geanalyseerd en gedistribueerd. En verkocht of gebruikt 'out of context', waardoor er persoonlijk zeer ingrijpende gevolgen kunnen zijn voor individuele personen. Er kan bijvoorbeeld worden herkend dat een polishouder een risico vormt, waardoor een hogere premie wordt doorgerekend. En wat als die analyse onjuist is? Bewijzen dat dat zo is door een individu is zo goed als onmogelijk.

DE JURIDISCHE COMPLEXITEIT

De grootste uitdaging vormt de juridische complexiteit van cloud computing. Het rechtskader van cloud computing bestaat zowel uit wetgeving als uit contracten, waarop vooral privacywetgeving (zoals de databeschermingsrichtlijn of Richtlijn 95/46 EG) haar stempel drukt. Deze wetgeving schrijft voor het gehele verwerkingstraject allerlei organisatorische en technische maatregelen voor. Bovendien mogen persoonsgegevens zonder toestemming van het ministerie van Justitie niet buiten de Europese Economische Ruimte (EER) worden verwerkt. Dit laatste punt is problematisch in de cloud. De aard van cloud computing brengt met zich mee dat er sprake is van grensoverschrijdende gegevensverwerking. Daarmee krijgen de rechtsverhoudingen internationale dimensies en raken zij verschillende jurisdicties, zowel privaatrechtelijk als (mogelijk) strafrechtelijk. Welk recht is op de rechtsverhouding van toepassing en welke rechter is in geval van welk type conflict bevoegd hierover te oordelen?

Naar verwachting zal in 2016 de Europese Privacyverordening intreden. Deze verordening zal één regime voor de privacyregulering voor heel Europa brengen. Een aantal van de nieuwe regels zullen grote invloed hebben op Big Data. De verordening streeft naar dataminimalisatie. Enkel de gegevens die nodig zijn om het vastgestelde doel te bereiken mogen worden verwerkt. Daar gaat de Big Data-beweging lijnrecht tegenin: het gaat om het verzamelen van zo veel mogelijk data, om zo volledig en betrouwbaar mogelijke analyses te realiseren. Daartegenover worden in deze verordening marketingdoeleinden wel als een

penibel de rechtspositie van de afnemers van clouddiensten is. In een DVO met een cloudleverancier dienen deze aspecten te worden ondervangen om alle juridische complicaties zo goed als mogelijk tegemoet te treden. Absoluut verplichtend in contracten zijn: de vastlegging van de instructiebevoegdheid van de klant, de verplichting van de leverancier iedere analyse op de data van de klant voor te leggen voor toestemming aan de klant, de verplichting van de leverancier tegemoet te komen aan de technische en organisatorische maatregelen die de wet de klant oplegt; en — zeer belangrijk — het recht van de klant

“CLOUD COMPUTING WORDT IN TIJDEN VAN BEZUINIGING VOORAL GEZIEN ALS EEN MOGELIJKE EN INTERESSANTE AUTOMATISERINGSOPTIE, OOK IN BIBLIOTHEKEN, MUSEA EN ARCHIEVEN.”

rechtsgeldige grondslag voor dataverzameling erkend. Dat was voorheen niet het geval. Er zal vaker sprake zijn van de verwerking van persoonsgegevens, zonder dat een individu ondubbelzinnig toestemming moet geven voor de verwerking van zijn persoonsgegevens. Wel staat daar een strenge informatieplicht tegenover, waardoor de transparantie over gegevensverwerking wordt verhoogd.

Naast de privacy problematiek kent cloud computing nog een aantal andere juridische risico's, zoals: een grotere kans op datalekken, een grotere kans op te snel vernietigen of te lang bewaren van data, de opslag en verwerking van data in landen buiten de EER, een verminderde controle van de klant op de verwerking van data door de leverancier in overeenstemming met de toepasselijke regelgeving en de mogelijke beslaglegging op de hardware (bijvoorbeeld in het kader van de Amerikaanse Patriot Act), waarop de data van een niet betrokken partij kan zijn opgeslagen.

Een dergelijke risicovolle en bedrijfskritische situatie maakt een standaardcontract met een cloudleverancier onacceptabel. Die contracten stellen afnemers in zeer afhankelijke en onmogelijke posities. De verdeling van verantwoordelijkheden is daarin meestal zeer onduidelijk. De leverancier is op basis van zo'n contract veelal in staat de data van de klant te gebruiken en te analyseren, terwijl er nauwelijks waarborgen zijn getroffen voor de bescherming van de privacy en de geheimhouding van de gegevens van de klant. Dat een leverancier zich meestal onttrekt aan aansprakelijkheid, het recht heeft de dienst aan te passen of te beëindigen zonder reden, niet de plicht heeft de data van een klant te bewaren bij het eenzijdig staken van de dienstverlening en een ongelimiteerd en ongecontroleerd aantal onderaannemers mag gebruiken, geeft aan hoe juridisch

om dit alles te controleren en audits daarop uit te voeren. En uiteraard dienen er zeer duidelijke afspraken te worden gemaakt over de verwerking van persoonsgegevens, exit-plannen, het bewaren van data bij het staken van dienstverlening, beveiliging, aansprakelijkheid en onbereikbaarheid. Hierdoor worden de meeste juridische implicaties van cloud computing (en van Big Data Analytics) afgedekt.

Gezien de (nog onbekende) ontwikkelingen binnen de cloud is volledige uitsluiting van juridische risico's echter een utopie. De huidige regelgeving zal toepassing krijgen in rechtspraak. Waar nodig zal ze in de toekomst verder worden aangescherpt. Tezamen met best practices zal dat ervoor zorgen dat veel van de juridische aandachtspunten worden getackeld.

BESLUIT

Cloud computing wordt in tijden van bezuiniging vooral gezien als een mogelijke en interessante automatiseringsoptie, ook in bibliotheken, musea en archieven. Het concept biedt vele mogelijkheden om kosten te besparen en tegelijkertijd kwaliteit en performance te verhogen. Uiteraard kunnen die laatste twee alleen indien de hiervoor aangegeven uitdagingen en juridische complicaties kunnen worden ondervangen. Veel organisaties zijn echter vooral gecharmeerd door de kostenverminderingen die kunnen worden gerealiseerd en de mogelijkheden om de eigen, complexe informatie infrastructuur af te bouwen. Ze zijn zich niet echt bewust van de potentiële problematiek, die cloud computing en Big Data met zich meebrengen. Juist die uitdagingen en complicaties echter oefenen rechtstreeks invloed uit op de performance van bedrijfsprocessen en zijn niet zomaar te ondervangen. Dat vergt nogal wat, waardoor het van belang is de organisatorische risico's goed in kaart te brengen en af te wegen. ■■