

# De Big Data revolutie

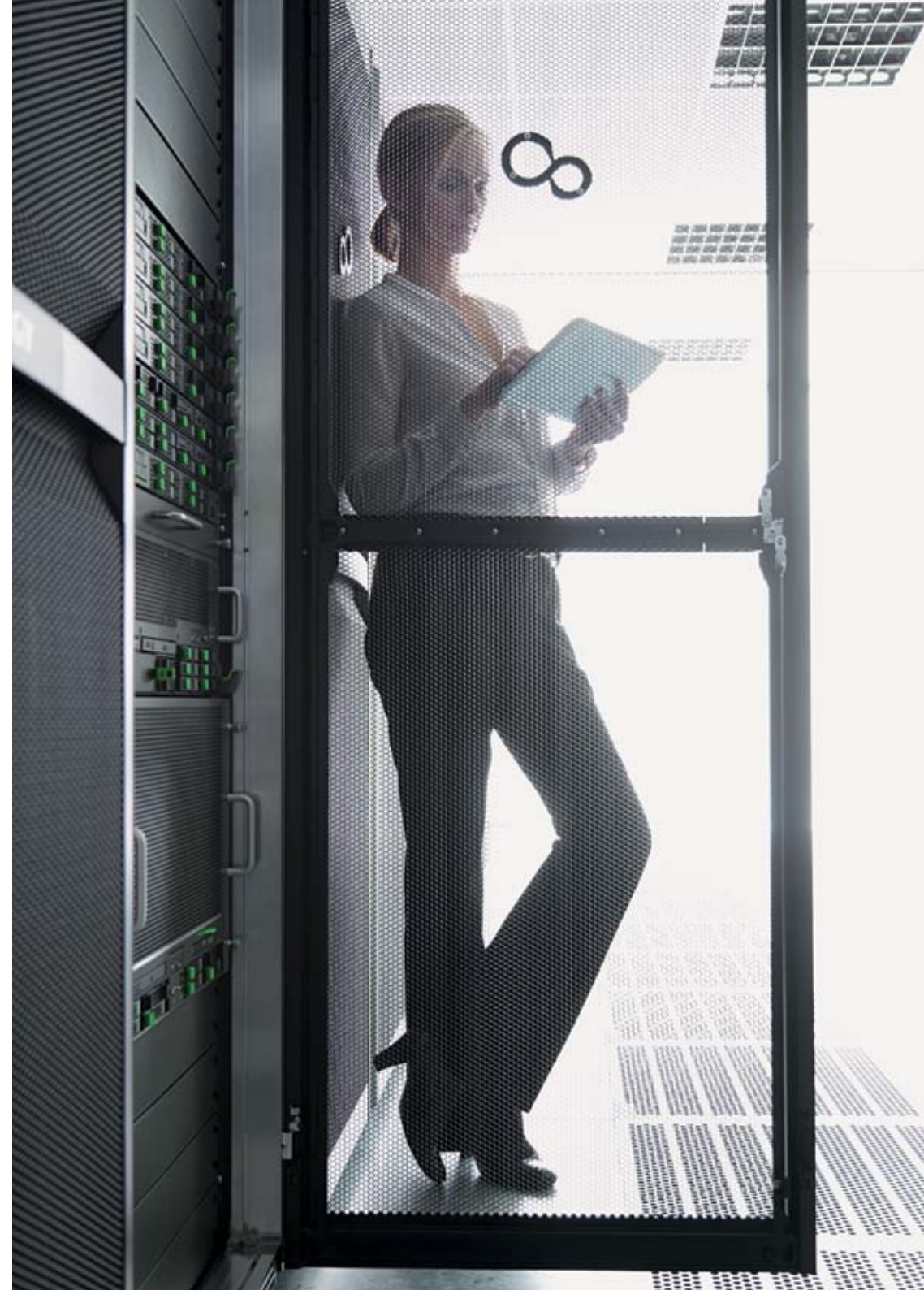
## Kansen, uitdagingen en visie



# Big Data als ultiem wapen in de concurrentiestrijd

De concurrentiekracht van een onderneming valt of staat met de kwaliteit en snelheid van de informatievoorziening. Wie beter, sneller en creatiever op veranderingen in de markt reageert, kan daarmee de concurrentie een belangrijke slag vóór zijn. CIO's zijn gefocust op 'agility'; op basis van verwachtingen, meten en analyseren, wil hij zijn strategie kunnen bijstellen. Daarom moet hij in staat zijn om uit steeds meer informatiebronnen almaar verfijnde informatie te halen. Het eindeloos opslaan van al die data is niet alleen kostbaar qua beheer en energieverbruik, ook de verwerking daarvan tot informatie wordt steeds moeilijker: er komt meer data binnen dan in korte tijd verwerkt kan worden, laat staan dat de juiste informatie tijdig daaruit gehaald kan worden.

De nieuwste technologie biedt nu eindelijk de middelen om beter met al die informatie te kunnen omgaan en daarmee de concurrentiekracht te verbeteren.



## »Zonder dynamische, flexibele bedrijfsprocessen zijn Big Data-toepassingen zinloos.«

Corporate databases op basis van onder meer Oracle, IBM of SAP – met hun ‘gestructureerde data’ – zijn binnen elke organisatie belangrijke gegevensbronnen. Maar, vooral bij complexe opvragingen, is het intensieve gegevensverkeer tussen opslag- en verwerkingssystemen steeds vaker een belemmering voor een snelle en gedegen opvraag van informatie (de I/O-bottleneck).

Daarnaast willen organisaties steeds meer ook ‘ongestructureerde data’ ontsluiten voor het maken van beslissingen. Het combineren van beide soorten data tot gegevenssets op basis waarvan men afgewogen beslissingen kan maken, is een uitdaging waarmee veel organisaties worstelen. Voor de raadpleging en analyse van die externe, ongestructureerde data is nog veel meer rekenkracht nodig. Deze gegevens, vaak afkomstig uit sensors of van internet, uit sociale media en van intelligente zoekmachines, zorgen voor een steeds grotere gegevens-poel. Bijvoorbeeld Facebook, Linked-In en Twitter voorzien in een vaak zeer open, weinig verhullende (dus interessante) en vaak actuele berichtgeving. Het zijn bronnen waaruit nog onvoldoende wordt geput. Bronnen die een verrassende toevoeging kunnen bieden op de zoekvraag. Want ondanks het informele karakter van deze netwerken komen er vaak zakelijke aspecten aan bod. Denk daarbij aan recruitment-toepassingen en marketinginformatie.

### Big Data

Alles bij elkaar gaat het om heel veel gegevens, waarvan een toenemend deel ongestructureerd; zodra hun gezamenlijke input groter is dan de verwerkingscapaciteit toelaat, spreken we van Big Data. Vijf jaar terug kenden we naar schatting wereldwijd zo’n 300 miljard gigabyte opslagcapaciteit; vier jaar later was dat al 988 miljard terabyte, ofwel ruim drieduizend keer zoveel en die groei zet zich exponentieel door.

Op zich is deze Big Data-problematiek niet nieuw, bij de vroeger gebruikte ponskaarten bestond ook al een verwerkingsprobleem. De input voor de steeds grotere mainframes werd voor de datatypistes zo omvangrijk dat de verwerking daarvan op grote problemen stuitte. Snellere systemen en betere databasetechnologie boden een oplossing, maar met de stortvloed ongestructureerde gegevens die bedrijven nu van buitenaf overspoelt, krijgt het begrip Big Data een nieuwe dimensie. En daarbij gaat het om ‘drie keer veel’, namelijk Veel data, Veel bronnen en Veel analyse. Bedrijven krijgen ook van hun klanten steeds vaker arbeidsintensieve opvragingen van informatie. Consumenten met hun smartphones en tablets worden veeleisender en willen betere, snellere informatie. Men wil nou eenmaal graag de status van een bestelling weten en kunnen natrekken wanneer die wordt afgeleverd. Het liefst wil men de dag daarvoor nog het tijdstip van levering wijzigen. En de luchtvaartpassagier die zijn vlucht wil omboeken, moet direct en perfect worden geholpen en daarna wil de maatschappij weten ‘wat twittert die man

over ons’, zeker als het een ‘frequent flyer’ is. Men wil immers een goed profiel van die klant opbouwen en ook dat is een enorme Big Data-uitdaging. Al die klantgerichte diensten vergen systemen die razendsnel de juiste informatie kunnen genereren. Hoe sneller en gedetailleerder die is, des te beter kan de service aan de klant zijn.

### Breed inzetbaar

Big Data-toepassingen zijn inzetbaar in allerlei organisaties van internationaal tot MKB-niveau. Zo werken banken met Big Data oplossingen om risicoanalyse en fraudedetectie te verbeteren. Grote retailers helpen hun franchisenemers om de juiste voorraadniveaus in hun winkels te bepalen. Ook voorzien we dat na verwerking van Big Data de resultaten beschikbaar worden gesteld of zelfs verkocht aan belanghebbenden in de financiële of logistieke keten. Een ander voorbeeld is de justitiemedewerker die bij een Internet Service

Provider onderzoek komt doen naar een zaak. Na opvraging van alle logfiles is het zinvol en snel bijeen brengen van alle data nog een hele klus. Op een heel ander terrein komen we ook Big Data tegen. Zoals onze elektriciteitsvoorziening die deels ook door zonne-energie wordt gevoed. Zogenaamde slimme meters geven door waar en hoeveel elektriciteit er op het transportnetwerk terugvloeit, wat van belang is voor een stabiele stroomvoorziening. Maar denk ook aan verbeteringen in forecasting & replenishment processen, een supermarkt die continu het juiste voorraadniveau bepaalt op basis van het advies van het hoofdkantoor in combinatie met lokale informatie zoals weersverwachting. Zo kan een regelmatig afwijkend weerbeeld aan onze kusten worden ingezet om de distributieketen flexibeler te maken en beter op die lokale verschillen af te stemmen. In de financiële markten gebruikt men voor financiële transacties zeer veel puntoplossingen, maar

## Big Data een ‘big issue’

‘De inzet van Big Data-toepassingen vergt meer dan alleen omvangrijk datamanagement’, stelt de Gartner Group. Het zal twee à drie jaar vergen om zinvol met de extra dimensies van Big Data te kunnen omgaan. Wereldwijd groeit die hoeveelheid data in bedrijven jaarlijks met minstens 59 procent. De opslag daarvan is geen probleem, maar het beheer en gebruik daarentegen wel. Waar het om gaat, is een effectief gebruik van de juiste data voor een betere bedrijfsvoering. Big Data is de belangrijkste ontwikkeling voor bedrijven gedurende de komende jaren.

Volgens McKinsey kan een bedrijf in de detailhandel bij een effectief gebruik van Big Data zijn operationele resultaten met zeker 60 procent verbeteren. De gezondheidszorg in de VS zou jaarlijks zo’n 300 miljard dollar kunnen besparen, ofwel grofweg 1000 dollar per inwoner. En bij de Europese overheden worden de besparingen geschat op 100 miljard euro, 300 à 400 euro voor elke burger. Kortom, een ‘big issue’.





met Big Data-technologie is het aantal systemen fors te reduceren. Ook op kleinere schaal, in het MKB, zijn talloze Big Data-toepassingen denkbaar.

### Aanpak

Veel bedrijven die Big Data-oplossingen zoeken, doen dat hand in hand met hun softwareleveranciers op basis van onder meer Oracle, de zogenaamde Exadata appliances, SAP Hana-technologie, Microsoft en de 'open source'/Hadoop-leveranciers, die allemaal hun eigen aanpak hebben. Veel bedrijven die aan de slag gaan met een open-source-oplossing, kijken pas later naar het noodzakelijke hardwareplatform. Dat is vaak een zoektocht, omdat bij veel vendors de geïntegreerde visie nog ontbreekt. En de ondernemers zelf weten ook vaak niet waar te beginnen, hoe zij uiteindelijk aan zo'n project uitvoering moeten geven. Immers, de uiteindelijke Big Data-toepassing moet uiterst betrouwbaar zijn door op het juiste moment over de juiste gegevens te beschikken en de juiste informatie te genereren. Een aanpak die bedrijven kunnen kiezen is samenwerking, bijvoorbeeld op brancheniveau,

zoals in de energiesector waar alle leveranciers hetzelfde transportnetwerk gebruiken. En kleinere bedrijven kunnen bij de ontwikkeling en inzet van Big Data-toepassingen mogelijk veel baat hebben bij samenwerking binnen hun brancheorganisatie. Verder zijn voor het MKB al tal van standaardtoepassingen beschikbaar. De implementatie van SAP® Business One Analytics, aangedreven door SAP HANA™ in zakelijke infrastructures is een voorbeeld dat zelfs met een relatief kleine database van 2,5 gigabyte en enkele SAP-gebruikers aanzienlijke voordelen te behalen zijn. Met slechts een muisklik kunnen gebruikers analyses genereren op basis van verschillende databases en informatie krijgen over de hele bedrijfsstructuur. Maar eerst moet de MKB-ondernemer zich bewust worden van de voordelen van Big data voor zijn organisatie. Als een foodketen met Big Data zijn voorraadkosten en die van zijn klanten wil minimaliseren en tegelijk zijn/diens omzet wil optimaliseren, moet hij weten hoe hoog die kosten zijn, op welke gegevens hij moet anticiperen en hoe dynamisch hij daarin moet zijn. Daarbij is een gedegen analyse van historische gegevens heel belangrijk.

»Wie zijn klant niet kent en adequaat op hem reageert, loopt de kans hem te verliezen.«

Als hij weet waar en waarom voorraadbeheer voorheen fout of juist heel goed verliep, is het gemakkelijker om dat beheer op basis van Big Data analyses bij te sturen. Het optimale resultaat is dan een dagelijkse bevoorrading van zijn klanten die voor hen net voldoende is om de dag door te komen, zonder ooit 'nee' te moeten verkopen.

### Stappenplan

De ondernemer die wil profiteren van de potentie van Big Data moet duidelijke stappen ondernemen, bepaald door de zogenaamde 'wisdom' piramide: uit data valt informatie te halen, informatie is de bron voor kennis, wijsheid is alleen mogelijk als je voldoende kennis hebt. Als we dit vertalen naar een aanpak voor de dagelijkse praktijk draaien we deze piramide om. Als we de strategie van een onderneming als wijsheid zien zal bepaald moeten worden welke kennis nodig is om deze strategie succesvol uit te voeren. De kennisbehoefte bepaalt vervolgens welke informatie nodig is voor de bedrijfsvoering. Na het bepalen van de informatiebehoefte is het zaak te onderzoeken welke informatie al beschikbaar is, zowel binnen als buiten de onderneming. Zodra men weet welke informatie ontbreekt is het ook duidelijk welke data nog nodig zijn en welke bronnen daarin kunnen voorzien. Deze bronnen kunnen zowel intern als extern gevonden worden, gestructureerd of ongestructureerd zijn etc. Vanaf dit moment wordt de route weer in de andere richting gelopen: al dan niet samen met externe consultants het proces van dataverzameling en dataontsluiting vastleggen en de tools implementeren om de informatievoorziening te faciliteren. Het succes van een dergelijk project wordt bepaald door de

mate waarin de onderneming daadwerkelijk kennis haalt uit de informatie en de manier waarop deze kennis wordt toegepast.

### Snelle hardware

De laatste belangrijke stap is de keuze van de juiste, snelle databasetechnologie en onderliggende hardware voordat de implementatie kan plaatsvinden. Cruciaal daarin is een onderliggend hardwareplatform met snelle servers en SSD-opslagsystemen, alsook snelle 'in memory computing' van online gegevens. Alleen daarmee is het mogelijk om de grote hoeveelheden data snel te verwerken tot de gewenste informatie. Nieuwe technologie speelt daar op in. Zo zijn bijvoorbeeld supersnelle SSD-geheugenchips veel sneller uit te lezen dan harde schijven met hun bewegende delen. En de 'in memory'-computers hebben alle noodzakelijke gegevens in één keer in hun werkgeheugen onder handbereik en hoeven niet steeds te wachten op tussentijdse aanvoer van nieuwe gegevens van buitenaf. Alleen dat maakt de verwerking van Big Data mogelijk binnen de vereiste korte responstijden.

### Conclusie

Pas na deze stappen zorgvuldig te hebben doorlopen, kan de onderneming volop profiteren van zijn Big Data-toepassing, resulterend in een betere service, een hechtere klantenbinding, een actievere marktbenadering, vergaande kostenreducties of wat dan ook. Allemaal zaken die de slagkracht en concurrentiepositie van de onderneming belangrijk kunnen verbeteren. Wie zich daarvan afzijdig houdt, is een dief van zijn eigen portemonnee!

# 'Geen organisatie kan Big Data ontlopen'

Als 's werelds grootste producent van netwerkoplossingen voor spraak-, data- en videocommunicatie, staat het Frans-Amerikaanse Alcatel-Lucent voor grote uitdagingen. De enorm groeiende markt van uiteenlopende mobiele apparatuur, de toenemende mobiliteit van de gebruikers daarvan, de expanderende behoefte aan realtime informatie en de brede beschikbaarheid van steeds meer (externe) informatiebronnen (waaronder ook social media) stellen hoge eisen aan de daarvoor ontwikkelde transportnetwerken.



»Houd de business centraal en haal met de juiste vragen de juiste informatie op tafel.«

Jeanine van der Vlist,  
countrymanager Alcatel-Lucent Nederland

# Veel organisaties onderschatten nog steeds de impact van Big Data

Volgens **Jeanine van der Vlist**, countrymanager Alcatel-Lucent Nederland, hebben met name consumenten bijgedragen aan enorme veranderingen in de ICT-markt: "Op een wereldbevolking van bijna zeven miljard mensen, tellen we zo'n zes miljard 'subscribers'. Onder deze zakelijke en individuele gebruikers van datanetwerken zijn er circa een miljard met een mobiele data-aansluiting. Zo kunnen we in het mobiele verkeer dus nog een enorme groei verwachten."

## Megagroei

Naast de autonome groei van het dataverkeer als gevolg van de steeds verder groeiende digitalisering van ons leven en het toenemend gebruik van vooral beeld- en video-overdracht, zijn er nog andere groeifactoren. "Het machine-to-machine (M2M) dataverkeer is ook een belangrijke ontwikkeling. Zo zijn er circa 1,2 miljard mobiele devices in gebruik in met name de productieomgeving. Volgens IDC groeit het aantal M2M-transacties in 2020 tot het dertigvoudige van de transacties tussen mensen onderling."

"Denk daarbij bijvoorbeeld aan de energiesector met de inzet van 'slimme meters'. Door via M2M-communicatie gemakkelijk meetgegevens door te geven, kan de frequentie daarvan ook enorm toenemen. Zo kunnen energieleveranciers relatief eenvoudig bijvoorbeeld maandelijks informatie vergaren in plaats van eens per jaar. En dat draagt

weer bij aan betere analyses van het energieverbruik van de klant, de herkenning van trends en mogelijk een betere dienstverlening aan die klant."

"Door al deze ontwikkelingen worden de datastromen", benadrukt Van der Vlist, "voortdurend groter. Dit fenomeen van 'Big Data' is met getallen alléén niet goed te bevatten. Het gaat daarbij – zo becijferen marktanalisten – rond 2020 om zo'n 35 zetabytes informatie; een 10 met 21 nullen. Om het toch in beeld te brengen, moeten we dat vergelijken met 2600 stapels boeken tussen aarde en zon; en negentig procent daarvan omvat ongestructureerde data! Daarmee is het fenomeen Big Data een uitdaging van ongekennde omvang voor alle betrokken partijen."

## Netwerkgroei

Om te voorkomen dat bestaande netwerken het dataverkeer straks niet aankunnen en dat de kwaliteit van de dienstverlening tekort schiet, moeten de netwerkmonitoring en het -beheer daar effectief op inspelen. "De traditionele tools zijn daarvoor ontoereikend", aldus Van der Vlist. "Voorheen ging het daarbij om de analyse van vaste datasets uit interne databases; maar tegenwoordig betreft het grotendeels ongestructureerde, dynamische data uit vele externe bronnen die constant wordt aangevuld."



»Big Data-toepassingen zijn slechts hulpmiddelen en vormen geen doel op zich.«





“Veel organisaties onderschatten nog steeds de impact van het fenomeen Big Data. Maar het is ook voor hen noodzakelijk daarin patronen te herkennen, bedreigingen inzichtelijk te maken en via goede analyses helder te krijgen hoe je succesvoller kunt opereren. Dat wil zeggen hoe je productiever, flexibeler of innovatiever kunt worden, of hoe je de ‘corporate position’ kunt verbeteren.”

“Voor de onderliggende netwerken betekent dit dat we die continu moet aanpassen aan het steeds veranderende dataverkeer. Enerzijds door

de groei van de hoeveelheid data en de datastromen in beeld te brengen zodat je de netwerkcapaciteit daarop structureel kunt aanpassen. Daarbij is een verdere simplificatie van die netwerken van belang om zodoende het gebruik daarvan eenvoudig en daarmee goedkoop te houden. Anderzijds is het nodig met de juiste tools – binnen de kaders die de Wet op de privacy stelt – te onderscheiden wat voor soort datastromen er lopen. Dat is nodig om in de operationele sfeer die data goed te kunnen prioriseren en te routeren.”

## Big Data is geen hype of revolutie maar gewoon een evolutie

### ‘Big Data’ een blijvertje

“Wat ik met name ook graag kwijt wil”, benadrukt Van der Vlist, “is dat geen bedrijf het fenomeen Big Data zomaar aan zich voorbij kan laten gaan. Men mag deze ontwikkeling niet onderschatten en er al helemaal niet voor weglopen. Om competitief te blijven is het essentieel om Big Data-toepassingen te ontwikkelen. Daarbij onderstreep ik ‘Stel altijd de business primair en bepaal goed wat je precies wilt bereiken’. Dat kan liggen op allerlei gebieden, variërend van productie tot innovatie en van winstmaximalisering tot optimalisatie van de klanttevredenheid. Begin daarmee klein en roep goede adviseurs in, zoals mensen van bijvoorbeeld toeleveranciers, collega’s, branche- of adviesorganisaties en raadpleeg internet. En zorg dat je aansluitend de juiste mensen inhuurt, aanneemt of opleidt en kies weloverwogen de tools die je voor de ontwikkeling wilt inzetten. Nogmaals, dit geldt voor nagenoeg elke organisatie. Van luchtvaart tot gezondheidszorg en van de food-sector tot toerisme, of wat je ook maar bedenkt.”

“En wie eenmaal aan de ontwikkeling van Big Data-toepassingen begint, moet zich goed realiseren nog helemaal aan het begin te staan van wat op dit gebied komen gaat. We praten dan ook niet over een eenmalige, kortstondige trend, maar om iets heel fundamenteels als de plaatsing van extra pijlers onder een hernieuwde, competitieve bedrijfsvoering binnen de bestaande organisatie. Big Data is dan ook”, naar de mening van Van der Vlist, “geen hype, geen revolutie, maar gewoon een evolutie!”

### Ondersteuning

In dit veld ziet Alcatel-Lucent voor zichzelf een belangrijke ondersteunende rol weggelegd ten behoeve van zijn afnemers. “We realiseren ons natuurlijk heel goed dat Big Data voor veel bedrijven een heel nieuw fenomeen is. En wanneer dat nodig is, kunnen wij hen bijstaan met enorm veel kennis en analysecapaciteit. Onze onderneming is immers zeer breed georiënteerd en wij hebben wereldwijde ervaring met de implementatie van nieuwe netwerken. Ook op het gebied van research kunnen wij het nodige bijdragen. Verder hebben wij voor onze klanten, naast onze producten, ook de juiste tools om daar het optimale uit te halen.”

Van der Vlist: “Ondanks de grote veranderingen die Big Data-toepassingen teweeg kunnen brengen, verandert één belangrijk gegeven niet: Centraal binnen elke organisatie blijven strategische vragen met betrekking tot de doelstellingen van de organisatie en hoe je die denkt te kunnen realiseren. Pas als dat helder is, kun je gaan nadenken hoe je uit alle beschikbare gegevens de juiste informatie naar boven haalt om die doelstellingen te kunnen realiseren. Pas dan komt de ontwikkeling van Big Data-toepassingen aan de orde. Deze vormen daarbij immers slechts een middel en geen doel op zich.”

# Deeltjesversneller als Big Data kanon

Het onderzoek dat het Nationaal instituut voor subatomaire fysica (Nikhef) richt zich op deeltjesfysica en astrodeeltjesfysica. In deze takken van de natuurkunde kijken wetenschappers naar elementaire deeltjes en zijn ze op zoek naar de interacties en structuur van die deeltjes, en hun onderlinge krachten. Dat alles op de kleinste schaal en met de hoogst haalbare energie. Binnen het Nikhef houdt Wim Heubers (Computer Technology) zich ondermeer bezig met het vraagstuk data opslag en Big Data.





»Voor ons is Big Data de gegevens die wij voor ons wetenschappelijke werk beschikbaar hebben vanuit het onderzoek naar de deeltjesfysica.«

## Nikhef filtert data om groei aan banden te leggen

Aan Wim Heubers leggen we de vraag voor wat Big Data voor het Nikhef is: "Voor ons zijn dat de gegevens die wij voor ons wetenschappelijke werk beschikbaar hebben vanuit het onderzoek naar de deeltjesfysica. Daarvoor hebben wij een deeltjesversneller in Geneve staan. Als die versneller productief is komt daar per seconde heel veel data vandaan. Die data noemen wij de ruwe onbewerkte data." Die data wordt gedistribueerd over een aantal computercentra in de wereld. Nikhef en Stichting Academisch reencentrum Amsterdam (Sara) zijn samen één van de tien computercentra in de wereld waar een kopie van de data van die deeltjesversneller wordt opgeslagen. "Dat zijn flinke hoeveelheden. Om een idee te geven: de opslagcapaciteit die wij hier voor hebben is 3 Petabytes, dus dat is wel 'big'. Daarvoor hebben we een paar dingen gedaan: de toegang tot die data moet niet alleen snel zijn, maar ook vanaf verschillende locaties te benaderen zijn. Dus niet alleen beschikbaar vanuit ons nikef.nl domein, maar ook via internet. Hier intern hebben we naast onze opslag de rekenclusters staan die snelle toegang moeten hebben tot die data. Die opslag zelf is niet super beveiligd in de zin van dat daar hele dure storage apparatuur staat. Het is relatief goedkope mass-storage want we gaan er van uit dat er altijd op een andere plek op de wereld een back-up van de data staat als er iets fout mocht gaan", aldus Wim Heubers en hij vervolgt: "Voor ons is Big Data dus vooral wetenschappelijke data."

### Meer mogelijk

Als er gekeken wordt naar de hoeveelheid data die nu en tien jaar geleden beschikbaar was, dan is er sprake van een flinke groei. Volgens Wim Heubers was de hoeveelheid data een factor tien tot honderd minder. "Maar dat is een beetje kip en ei kwestie. Tien jaar geleden kon je de opslagcapaciteit die we nu hebben gewoon niet betalen. Nu wel en dat wordt benut door de wetenschappers die door hoe meer data ze kunnen analyseren, hoe beter ze hun werk kunnen doen", zegt hij. De kwaliteit van het onderzoek gaat daarmee vooruit. "Zo'n peperdure deeltjesversneller in Geneve kun je optimaler benutten door per tijdseenheid veel meer data op te slaan. Big Data is eigenlijk onze corebusiness. Als wij niet in ons wereldje in staat zijn om dergelijke hoeveelheden data op te slaan, heeft het ook geen zin om zo'n deeltjesversneller te bouwen. De opslag en netwerktechnologie maken het mogelijk dat ding optimaal te benutten."

### Tegendraads

Je zou verwachten dat de groei van het data gebruik de stijgende lijn van de afgelopen jaren gaat voortzetten, maar daar is volgens hem geen sprake van. "We denken dat we qua hoeveelheden data redelijk op de top zitten. Dat heeft te maken met dat we de deeltjesversneller steeds beter leren kennen. Hoe langer je dat apparaat gebruik hoe beter je aan de bron kunt gaan filteren. Dus als we aan de bron al vrij veel data kunnen

# Voor Nikhef is Big Data hoofdzakelijk wetenschappelijke data

uitsluiten waarvan je weet dat het toch niet belangrijk is voor het onderzoek, dan hoef je minder op te slaan. Dus voor ons geldt dat de hoeveelheid data die we gaan verwachten niet veel groter is dan dat het nu is. We hebben echt het idee dat we wat dat betreft nu aan de top zitten. Voor veel bedrijven zal het zo zijn dat ze nu pas beginnen te denken over Big Data maar wij lopen daar dus op vooruit. Er geldt natuurlijk ook dat je wel heel veel kunt opslaan, maar het analyseren kost enorm veel tijd en computerkracht. Dat heeft voor ons niet zo veel zin. Het moet handelbaar blijven en dat is vaak het probleem met Big Data. Al is het alleen maar omdat je het niet zo eenvoudig van de ene naar de andere plek kunt verplaatsen. Als het op een gegeven moment drie weken duurt voordat alles verplaatst is, moet je je afvragen waar je mee bezig bent."

## Toch groei van Big Data

Buiten de wereld van het Nikhef ziet Heubers wel een enorme groei in Big Data. Die opslag van Big Data wordt naar zijn mening vooral gedreven door consumentengedrag. "Vooral via internet zie je dat er steeds meer data in allerlei vormen wordt opgeslagen en die toegankelijk moet worden gemaakt. Eigenlijk het zelfde zoals bij ons een beetje aan de orde was. De technologie maakt het mogelijk om het op te slaan. Elke foto, film en tv-uitzending kun je opslaan en dat gebeurt dan ook. Het verkeer op het internet wordt voor het grootste deel bepaald door het consumentengedrag en dat

geldt voor Big Data ook. Veel bedrijven zijn geïnteresseerd in die consumentendata om het gedrag te kunnen voorspellen en daarmee beter hun producten op het publiek te kunnen afstemmen."

## Filteren moet

Bij het Nikhef geldt zoals gezegd dat de hoeveelheid data per tijdseenheid minder wordt, maar de behoefte aan opslag wordt niet minder wordt. "Dat komt omdat we niks weggooien en alles moeten bewaren." Op de vraag of de filtering aan de bron die het Nikhef toepast ook iets is waar andere bedrijven die met Big Data te maken krijgen naar moeten gaan kijken antwoordt Heubers: "Dat denk ik wel. Je kunt wel alles opslaan en 80% niet analyseren omdat dat niet interessant is, maar dan ben je verkeerd bezig. Daar kun je alleen maar achter komen door er mee aan de slag te gaan. Weggooien kan altijd nog, maar als je te vroeg dingen weggooit dan kun je essentiële zaken missen als je je analyseproces niet goed kent. Je kunt alleen maar goed filteren aan de bron als je weet hoe die filters er uit moeten zien. En dat geldt misschien ook voor niet wetenschappelijke Big Data toepassingen." Hij geeft aan dat het niet makkelijk zal zijn om mensen zelf hun YouTube filmpjes te laten filteren. "Misschien dat er een soort automatische correctie zal komen die aangeeft dat het geen zin heeft om een bepaald filmpje op te slaan. Ik zie het bij me zelf ook. Ik heb bijna 30 duizend foto's op m'n laptop staan. Da's eigenlijk niet meer te vatten."




»Je kunt wel alles opslaan en 80% niet analyseren omdat dat niet interessant is, maar dan ben je verkeerd bezig.«

Ooit is er in een onderzoek gesteld dat Big Data een bron van nieuwe inkomsten zou kunnen genereren zoals ook olie die voor de economie was en is. Zijn daar parallellen richting Big Data als inkomstenbron mee te trekken volgens Heubers? "Als je kijkt naar wat er allemaal op

sociale netwerksites als Facebook, Google en LinkedIn wordt opgeslagen en als je dan vervolgens ziet dat dergelijke bedrijven veel geld verdienen en veel waard zijn op de beurs dan zit daar wel iets in."

# Grote effecten met Big Data en slimme zetten

Als een van 's werelds grootste spelers in software voor 'business intelligence', ontwikkelt SAP oplossingen voor 'real time' toegang tot en verwerking van zeer grote hoeveelheden data, ofwel 'Big Data'. Vroeger bestond daarvoor onvoldoende processorcapaciteit, maar vandaag de dag echter wel en draait het meer om de inzet van de juiste applicaties. Wie zijn data-infrastructuur op orde heeft, kan daarmee de concurrentie een slag voor zijn door het optimaliseren en innoveren van zijn business.

A close-up portrait of Mark Raben, a bald man with a slight smile, looking upwards and to the right. He is wearing a white shirt and a dark jacket. The background is blurred, showing what appears to be a modern office or conference setting.

»Wij ordenen bij onze klanten die groeiende bergen data naar informatie.«

**Mark Raben,**  
director Innovation  
& Product Strategy,  
bij SAP in Nederland



# Door fusies en afsplitsingen is de kwaliteit van de databestanden in het geding gekomen

“Bij het nemen van beslissingen richten bedrijven zich niet meer alleen op de interne bedrijfsgegevens, maar steeds meer ook op externe gegevens uit het speelveld waarop zij zich bewegen. En die hoeveelheid informatie groeit wereldwijd exponentieel; van circa 300 miljard gigabyte medio 2007 tot bijna 1000 terabyte (1 zettabyte) vier jaar later; een jaarlijkse verachtvoudiging”, aldus Mark Raben, Director Innovation & Product Strategy bij SAP in Nederland.

## Focus op bruikbare informatie

Raben: “Wij ordenen bij onze klanten die groeiende bergen data naar informatie. Daarbij laten we ons niet te veel leiden door de omvang van Big Data, maar focussen we op de daaruit te destilleren bruikbare informatie. Al velen jaren generen organisaties data, waarvan de informatiewaarde niet (juist) werd ingeschat. Voorzover die data is bewaard, is er niets mee gedaan. Bij optimalisering van de informatievoorziening, bijvoorbeeld bij de invoering van datamanagementsystemen, komen die data en de betekenis daarvan vanzelf bovendrijven.”

“Ondernemingen realiseren zich dan dat zij aan hun per bedrijfsonderdeel keurig bijgehouden en opgeslagen data weinig hebben omdat die afzonderlijke silo's geen eenduidige stuurinformatie naar het management opleveren. Rationaliseren van dat versnipperde IT-landschap naar een bedrijfsbrede opzet is heel goed mogelijk zonder de onderscheidende kenmerken van de diverse

business units op te geven. Zo'n optimalisatieslag zorgt voor een harmonisering van de datadefinities. Wat is van waarde, waarvoor en wanneer? Bedrijven die aan 'data life cycle'-management doen, kunnen hun business intelligence applicaties veel eenvoudiger snel en effectief door de gehele organisatie uitrollen. En of je die slimme vragen nu op veel of op extreem veel data afvuurt, is meer een kwestie van processorcapaciteit”, aldus Raben.

## Vervuilde bestanden opschonen

“Daarvoor moet je wel de reeds aanwezige eerst opschonen”, benadrukt hij. “Door vele fusies en afsplitsingen in bepaalde bedrijfstakken is daar de kwaliteit van de databestanden in het geding gekomen. Zo koos Energienetwerkbedrijf Alliander er voor om op dat vlak eerst de organisatie op orde te brengen alvorens nieuwe IT-oplossingen in te voeren. Want wie zich met klantvriendelijkheid staande wil houden in deze sterk competitieve energiewereld, moet eerst zijn data ontdoen van vervuiling.”

“Na selectie van de data-eigenaren vormde het bedrijf datakwaliteitsteams die zich richtten op de volledigheid, actualiteit, juistheid en nauwkeurigheid van de data. Dat gaf direct ook een beter inzicht in de situatie achter alle abstracte cijfers. Zo kregen 6000 klanten van het bedrijf een naheffing van vijf- tot veertienhonderd euro. Want ze betaalden ondanks hun aansluiting voor grootverbruikers jarenlang het gewone tarief. Enkel van



hen hadden die aansluiting niet nodig, waarop die werd verkleind en hun naheffing verviel.”

#### Waarde voor de business

“De technologie om vervolgens snel en doeltreffend de groeiende databrij te rubriceren en te analyseren, is voorhanden. Denk daarbij aan ‘in memory’-databases en algoritmen voor het indexeren van gestructureerde én ongestructureerde data. Zowel in- als externe databronnen laten zich combineren en verschaffen het management de mogelijkheid te zoeken naar trends. Dat proces is heel goed te structureren, mits men zich op alle niveaus in de organisatie bewust is van de waarde van hun lokale databronnen voor de totale onderneming.”

“Wanneer het fundament voor ‘business master data management’ eenmaal is gelegd, vergt een volgende verfijninglaag om informatie te leveren voor bijvoorbeeld de aansturing van de ‘supply chain’ of het beheer van de kasstroom, minder inspanningen. Vergeleken met de investering in de uitrol van een bedrijfsbreed erp-systeem, is een additioneel systeem voor operationele ‘business intelligence’ maar een kleine stap”, benadrukt Raben. Dat vindt ook het Rotterdamse veevoederconcern Provimi. Door een direct en beter inzicht in de voorraden betaalt hun investering zich snel terug. En minder voorraden betekenen een geringer beslag op het werkkapitaal. “Na een optimalisatieslag kunnen organisaties nadenken over een vernieuwing van hun bedrijfs-

## Dankzij ‘in memory computing’ zijn die meetgegevens zeer snel in te lezen en real-time te analyseren

processen, innovatie dus. Velen koppelen dat aan Big Data, nu de samenleving steeds sneller digitaliseert. Een voorbeeld is de slimme elektriciteitsmeter die energiebedrijven bij hun afnemers plaatsen. De daaruit komende meetwaarden worden centraal door de elektriciteitsleverancier uitgelezen en opgeslagen. Met miljoenen aangesloten huishoudens en jaarlijks meerdere uitlezingen gaat het om écht Big Data. Dankzij ‘in memory computing’ zijn die meetgegevens zeer snel in te lezen en real-time te analyseren, anders dan met conventionele databasetoepassingen.”

#### Panklare analyse-apps

“Velen koppelen Big Data aan nieuwe vormen van klantinteractie. De manier waarop consumenten via social media heel direct oordelen over producten, trekt een zware wissel op de analysecapaciteit van marketingorganisaties. Panklare ‘in memory’-oplossingen, direct beschikbaar als appliance of uit de cloud, ontrafelen na een korte inleertijd de stroom data die een marketingorganisatie 24 uur per dag opvangt ten behoeve van een gerichte en accurate databasemarketing. Je moet alleen weten waar en wanneer je de voelsprietten moet plaatsen”, geeft Raben aan. “In de industrie zijn sensoren al langer gemeengoed en met ‘in memory’-technologie worden eerdere barrières in reactietijd en effectiviteit doorbroken.”

“Van grote maatschappelijke waarde is Fujitsu’s concept voor de vernieuwing van landbouwprocessen. Dankzij Big Data-analysemodellen zijn nu

betere oogsten te realiseren. Sensoren monitoren nauwlettend atmosferische veranderingen (het weer), de bodemgesteldheid en het effect van eerdere bemesting. Een analytisch systeem berekent voorspellingen en werkt, met de kennis van vorige oogstjaren, scenario’s uit voor een verantwoorde agricultuur.”

“Zowel bij maatschappelijke als bedrijfsbelangen leidt de verwerking van Big Data al snel tot rendement op de investering. Gebaseerd op ‘in memory’-technologie levert SAP kant-en-klare appliances voor onder andere ‘controlling and profitability analysis’, ‘smart metering’ met slimme energiemeters, ‘trade promotion management’ en ‘strategic work force management’. Kortom”, resumeert Raben, “allemaal toepassingen die met geringe investeringen al snel kunnen bijdragen aan innovatieve real-time Big Data-toepassingen, resulterend in meer efficiency, flexibiliteit en kostenreducties.”

# Versnelling van DNA analyse helpt kankerpatiënten

Voor Bert Reijmerink, CEO van Genallice, is Big Data niet heel veel data van één type, maar veel data van verschillende soort. "Daarom ligt voor ons de uitdaging in het uit die berg data het antwoord te halen dat we zoeken. Er zit heel veel kennis versleuteld in die data en voor ons is het de uitdaging om die kennis er uit te halen. Het heeft voor ons dus niet te maken met de omvang van de data maar de diversiteit en de complexiteit van de data."

»De snelheid waarmee die medische data op ons af komt gaat met zo'n razend tempo dat je bijna kunt spreken van een data-tsunami.«

**Bert Reijmerink,**  
CEO Genallice



# Genalice ontwikkelt unieke methode om medische data te 'cruchen'

Hij geeft aan dat zijn bedrijf een andere visie heeft op data analyse dan andere partijen. "Als je kijkt naar high volume data-analyse dan wordt er gezegd dat je zo veel mogelijk data moet analyseren. Wij zeggen dan nee, je moet de 'right volume' of data analyseren. Wij beginnen met Big Data om te zetten naar smart data om vandaar uit verder te gaan. Uit ervaring weten we dat relevante informatie verstopt is als een speld in een hooiberg. Om die speld te vinden moet je wel eerst alle data bekijken. Je begint altijd met een basis die een bepaalde omvang heeft. Als we kijken naar de medische data dan is DNA data verreweg de grootste en onhandigste dataset die je kunt hebben. Bijna 99,9% van die data is identiek tussen verschillende personen. Dat wil niet zeggen dat dat grote deel niet relevant is, maar die tiende procent bepaalt hoe je je ontwikkelt en hoe bijvoorbeeld je organen er uit gaan zien. Afwijkingen in in het DNA bepalen of bijvoorbeeld cellen zich gaan ontwikkelen en er tumoren gaan ontstaan. Dus wij zijn op zoek naar die afwijkingen. Op die manier zijn we dus in staat om die terabytes aan DNA-data met een factor duizend te verkleinen naar de essentie. Maar je bent hoe dan ook gebonden om eerst die grote hoeveelheid data vast te leggen voordat je het kunt schiften."

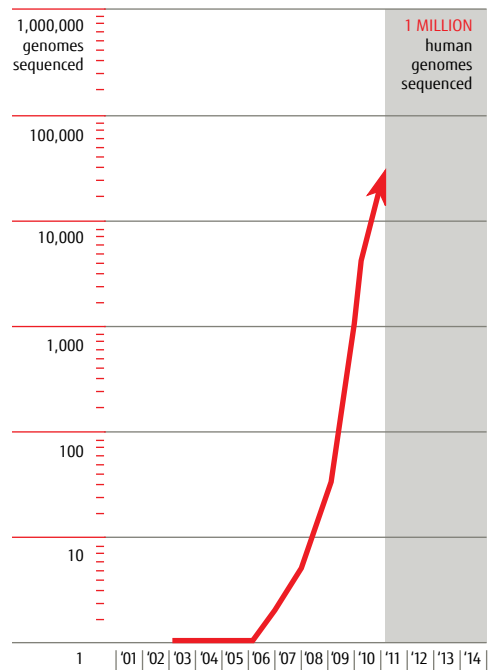
## Tsunami

"De snelheid waarmee die medische data op ons af komt gaat met zo'n razend tempo, en dat kost zo veel om op te slaan, dat je bijna kunt spreken van een data-tsunami. En dat komt hoofdzakelijk omdat de techniek van de productie van data hand over hand toeneemt. Nemen we DNA-data als voorbeeld dan zien we dat we nog geen tien jaar geleden voor het eerst in staat waren om een menselijk genome in kaart te brengen. Het heeft uiteindelijk 13 jaar geduurd en 13 miljard gekost. Nu zijn we 9 jaar verder, kunnen we het in een paar dagen en kost het een paar duizend dollar. Voor volgend jaar verwachten we dat het nog maar een kwartier kost en ruim onder de duizend dollar zal kosten." Bert Reijmerink verwacht dat het produceren van DNA data commodity wordt en dat bij een ziekenhuis bezoek dat tot de standaard onderzoeken gaan behoren, want het is snel, betrouwbaar en het bevat volgens hem de sleutel voor het oplossen van de problemen van de patiënt.

"Op jaarbasis komen er 20 miljoen kankerpatiënten bij waarvan 30% toegang heeft tot moderne gezondheidszorg, dus van 5 miljoen patiënten wordt het DNA in kaart gebracht. Daarvoor is al 5 miljoen keer één terabyte aan opslag nodig. Dat moet ergens opgeslagen en verwerkt worden", zegt hij. De snelheid en de hoeveelheid waarmee die data op ons afkomt kan volgens de huidige technologie niet verwerkt worden. "Die datagolf gaat ons overspoelen. Er wordt in dit verband dan

# Wij brengen terabytes aan DNA data terug tot de essentie

## DNA data explosie



Bron: Technology Review, mei/juni 2012, 'Bases to Bytes' door Mike Orcutt

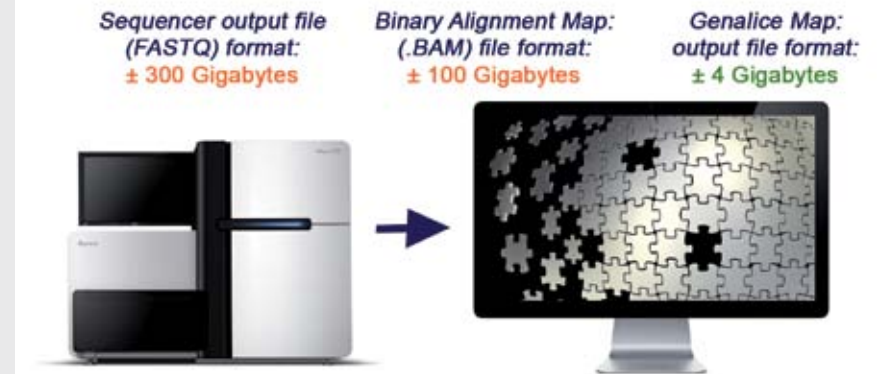
ook in onze wereld gesproken over de mogelijke problemen die er zullen zijn om voldoende opslagcapaciteit ter beschikking te hebben. Dus wij hebben gezocht naar methoden om die data vele malen kleiner te maken en gelukkig hebben we daar ook manieren voor gevonden. Dat is niet in het nadeel van de storage fabrikanten, want die krijgen de data toch wel voor hun kiezen. Maar bedrijven die data verwerken moet echt gaan nadenken hoe ze hun data efficiënter kunnen maken. Bedrijven die dat kunnen zullen zich gaan onderscheiden in de markt", voorspelt Reijmerink.

### Innovatie award

Genalice is ontstaan van uit een bedrijf InverselT dat zich bezig hield met het optimaliseren van IT-landschappen binnen bedrijven om daarmee besparingen te realiseren. Op een gegeven moment kwam er het verzoek van het Erasmus ziekenhuis om te kijken of er niet efficiënter met DNA data om te kunnen worden gegaan. "Dat was voor ons een eye-opener. Er lag en ligt een enorme markt en bovendien konden we technologie iets bijdragen aan mens en maatschappij. Uiteindelijk zijn we via de organisatie van Alpe d'HuZes die geld inzamelt ten behoeve van kankeronderzoek door het KWF, in contact gekomen met het UMC Utrecht. Daar hebben we verdiepingsonderzoek gedaan om te zien hoe de onderzoeksweld werkte, welke data verwerkt werd, en wat we daaraan konden bijdragen. Het rapport wat daaruit voortkwam hebben we gepresenteerd

## Reductie van genome DNA data

Een DNA-sequencer produceert ongeveer 300 GB aan ruwe data (puzzelstukjes) voor een enkel menselijk DNA. Het leggen van de puzzel duurt op dit moment minimaal 24 uur en resulteert in ongeveer 100 GB aan data. De Genalice technologie lost de puzzel in een paar minuten op en er blijft slechts 4 GB aan data over.

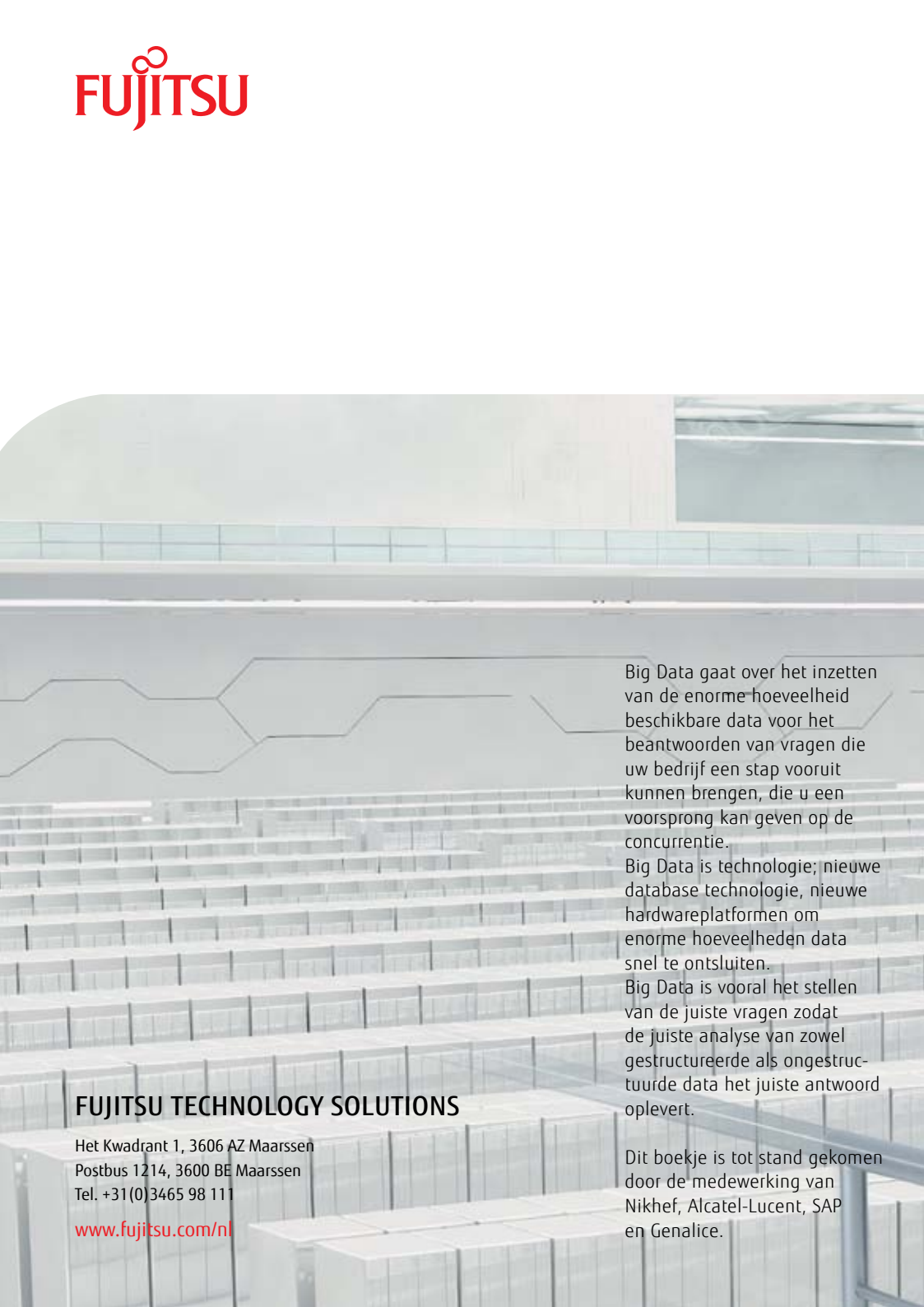


en men was unaniem van mening dat we door moesten met het ontwikkelen van onze technologie", aldus Reijmerink. Dat resulteerde in de oprichting van het huidige bedrijf en dat daarmee de goede weg is ingeslagen blijkt ook wel uit dat dit jaar de nationale ICT-award voor innovatie is binnengesleept.

### Menselijke maat

Big Data is voor Genalice niet alleen veel data, maar ook data van verschillende bronnen. "Neem ik als voorbeeld de casus van een schildklierkanker patiënt, dan spreek je over zeven verschillende bronnen die gezamenlijk 45 TB aan data genereren gedurende het onderzoeks- en genezingsproces. Om daar doorheen te werken kost waanzinnig veel tijd aan processorcracht. Inmiddels onderkent

ook de onderzoeksweld dit en ziet men in dat snelheid nu echt van belang gaat worden. Er is maar één antwoord om die uitdaging te slechten en dat is technologie. Als we u eens in staat zouden zijn om die enorme hoeveelheid data klein en efficiënt te maken, dat op een hele snelle en elegante manier te processen en dat ook via de cloud een platform biedt aan wetenschappers en onderzoekers, dan kunnen we stappen maken in de voortgang van big medical data", aldus Bert Reijmerink.

The background of the entire page is a photograph of a data center. It shows rows of server racks stretching into the distance, with a curved ceiling and a grid of lights. The image is slightly blurred, giving a sense of depth and scale.

## FUJITSU TECHNOLOGY SOLUTIONS

Het Kwadrant 1, 3606 AZ Maarsse  
Postbus 1214, 3600 BE Maarsse  
Tel. +31(0)3465 98 111

[www.fujitsu.com/nl](http://www.fujitsu.com/nl)

Big Data gaat over het inzetten van de enorme hoeveelheid beschikbare data voor het beantwoorden van vragen die uw bedrijf een stap vooruit kunnen brengen, die u een voorsprong kan geven op de concurrentie.

Big Data is technologie; nieuwe database technologie, nieuwe hardwareplatformen om enorme hoeveelheden data snel te ontsluiten.

Big Data is vooral het stellen van de juiste vragen zodat de juiste analyse van zowel gestructureerde als ongestructureerde data het juiste antwoord oplevert.

Dit boekje is tot stand gekomen door de medewerking van Nikhef, Alcatel-Lucent, SAP en Genalice.