

'Big data' en de creatieve destructie van bestaande businessmodellen

'Big data' – kolossale hoeveelheden gegevens die met conventionele technologie bijna niet meer te verwerken zijn – kunnen de onderneming die er wél raad mee weet veel voordeel brengen.

Joseph Schumpeter, een van de beroemde economen van de twintigste eeuw, zei ooit: 'Een wezenlijk kenmerk van innovatie is dat er een grote stap wordt gezet en een grote verandering tot stand wordt gebracht ... en bijna geen enkele aanpak die voorheen optimaal was blijft dat ook nog daarna.' Zijn uitspraak lijkt zonder meer van toepassing op de creatieve destructie die de bestaande businessmodellen in het bedrijfsleven te wachten staat als gevolg van de stormachtige groei van het gegevensvolume die we tegenwoordig beleven. Iedere minuut van iedere dag komt er een – steeds grotere – stortvloed aan gegevens bij. Alleen al in 2012 is er naar schatting 2,5 zettabytes² aan gegevens gegenereerd, en de trends wijzen erop dat het volume van door bedrijven gegenereerde gegevens de komende jaren alsnog verder zal aanzwellen. Enkele cijfers ter illustratie:

- Iedere dag ontstaan er 2,5 quintiljoen bytes aan gegevens, en 90 procent van alle gegevens ter wereld is pas in de afgelopen twee jaar gegenereerd.³
- In 2020 zullen er naar verwachting 44 keer zoveel gegevens worden gegenereerd als in 2009.⁴
- Het wereldwijde volume aan bedrijfsgegevens zal naar verwachting iedere 1,2 jaar verdubbelen.⁵
- Wal-Mart verwerkt een miljoen klanttransacties per uur en slaat deze op in databases die naar schatting meer dan 2,5 petabytes aan gegevens bevatten.⁶
- De kolossale toestroom aan gegevens zet de infrastructuur voor IT onder druk. In een recente enquête zei 55 procent van de ondervraagde topmanagers dat hun IT-systemen ver-

traagd raken door het volume aan gegevens dat ze te verwerken krijgen.⁷

- Gebrekkig databeheer kan een onderneming tot wel 35 procent aan exploitatie-inkomsten kosten.⁸

'Big data' belooft fundamentele veranderingen te weeg te brengen. De rekenkracht van computers heeft zich alsnog verder ontwikkeld en er kunnen steeds grotere en complexere volumes aan gegevens mee worden verwerkt. Er liggen grote voordelen in het verschiet voor de onderneming die de bergen aan gegevens, waarover zij beschikt, weet te analyseren. 'Big data' staat dan ook hoog op de agenda voor directies en afdelingen voor IT. Tal van vooraanstaande ondernemingen laten het ook niet bij woorden. Uit een recent onderzoek van A.T. Kearney naar innovatie in IT, bleek dat 45 procent van de ondervraagde ondernemingen in de afgelopen twee jaar een initiatief had gelanceerd op het gebied van 'business intelligence' (BI) of 'big data'.⁹ Volgens ander onderzoek verwachtte ruim negentig procent van alle ondernemingen in de *Fortune 500* binnen een jaar minstens één 'big data'-initiatief te hebben opgestart.¹⁰ Een doeltreffend gebruik van deze stortvloed aan gegevens kan van grote invloed zijn op zowel omzet als winstgevendheid (zie figuur 1). Een onderneming die capaciteiten opbouwt op dit gebied zal niet alleen beter kunnen presteren in traditionele segmenten en functiegebieden, maar zal ook kansen kunnen creëren om het aanbod aan producten en diensten uit te breiden.

In sommige bedrijfstakken, zoals de financiële dienstverlening, zijn er door 'big data' volstrekt

De auteurs zijn verbonden aan adviesbureau A.T. Kearney: C. Hagen, K. Khan en J. Miller (Chicago) zijn resp. partner, vice president en principal, M. Ciobo en A. Yadav (Melbourne) principal resp. consultant, D. Wall (New York) is principal en H. Evans (San Francisco) director.¹

1. De auteurs zijn Matt Rahn, Bharath Thota, Anand Seshadri, Dushiant Kochar, Jim Lee, Mohit Mohal en Michael Wei erkentelijk voor hun bijdragen aan dit artikel. De auteurs zijn bereikbaar via [voornaam].[achternaam]@atkearney.com (bijv. christian.hagen@atkearney.com).
2. 1 zettabyte is ongeveer 1.000 exabytes ofwel een miljard terabytes. [Noot v. vert.: deze aanduidingen conform de decimale (i.t.t. binaire) naamgeving voor veelvouden van bytes van de IEC (<http://nl.wikipedia.org/wiki/Byte>).]

3. 'Bringing Big Data to the Enterprise', IBM, 2012. [Noot v. vert: 1 quintiljoen = 10^{18} (korte schaalverdeling) ofwel 1 Exabyte volgens de IEC-norm vermeld in noot 2.]

4. 'A Comprehensive List of Big Data Statistics', Wikibon Blog, 1 augustus 2012; Oracle schatte in 2012 dat het wereldwijde volume aan gegevens met 40 procent per jaar groeit (samengestelde groei) en tegen 2020 bijna 45 ZB zal bedragen.

5. 'eBay Study: How to Build Trust and Improve the Shopping Experience', KnowIT Information Systems, 8 mei 2012.

nieuwe businessmodellen ontstaan. Bij het algoritmisch handelen, bijvoorbeeld, worden tegenwoordig per minuut kolossale hoeveelheden aan marktgegevens geanalyseerd. Zo worden er vrijwel onmiddellijk kansen gesignaleerd om geld te verdienen – iets wat nog maar tien jaar geleden onvoorstelbaar was. In de detailhandel maakt 'big data' het mogelijk om het aankoopgedrag van de klant in de winkel vrijwel onmiddellijk te analyseren. Met dergelijke razendsnelle inzichten in veranderingen in de vraag van de consument kunnen producten, voorraden en prijzen in de winkels worden aangepast en verkoopresultaten geoptimaliseerd.

In allerlei bedrijfstakken kan 'big data' zorgen voor grote gegevensvolumes en grootschalige verwerkingscapaciteit om de groei aan te jagen en mogelijkheden te vinden om kosten te besparen. De gekozen benadering zal niet in iedere bedrijfstak hetzelfde zijn. Ook zullen her en der verschillende aspecten worden benadrukt – van marketing tot het beheer van de toeleveringsketen. Maar vrijwel steeds gaat het om een fundamentele verandering waarbij gebruik wordt gemaakt van gegevensanalyse ('analytics') en 'big data' (zie figuur 2). Ook bij overheden zien we deze ontwikkeling; daar worden grote volumes aan gegevens doorgeploegd om de dienstverlening aan de burger te verbeteren, de belastinginning te optimaliseren en

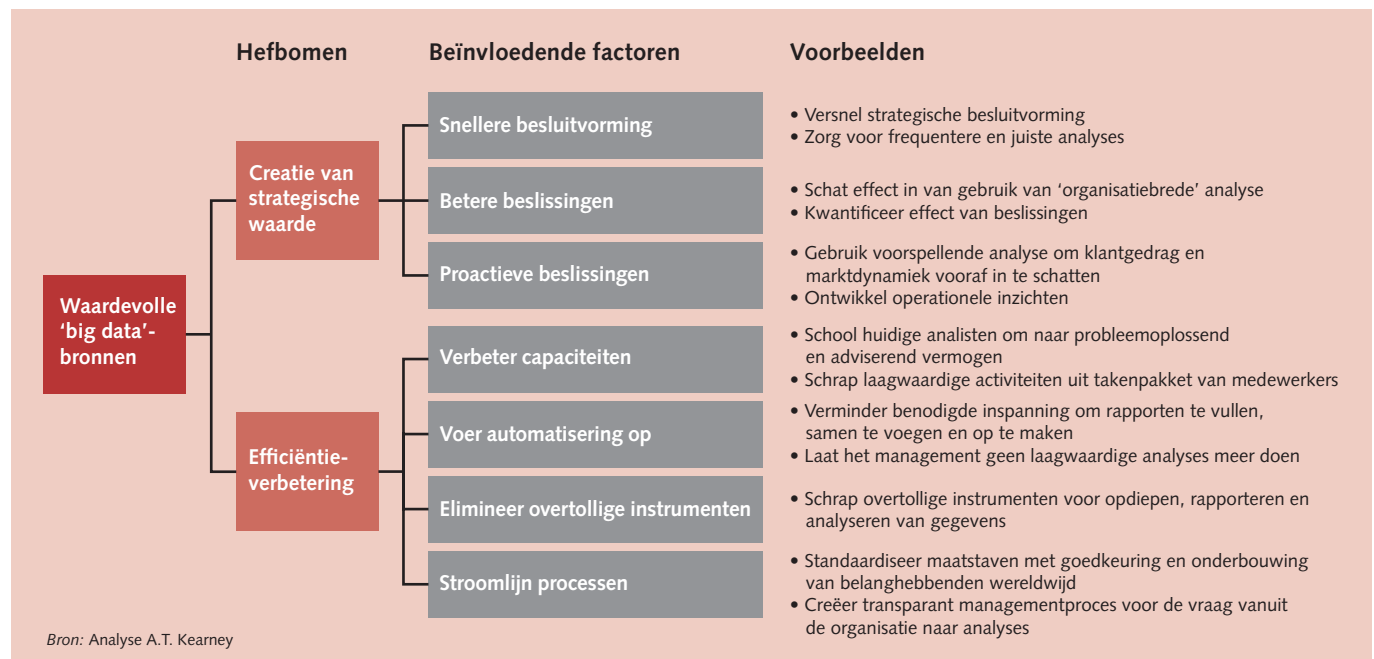
de landsbeveiliging te verbeteren met behulp van geavanceerde wapens, zoals onbemande voer- en vliegtuigen.

Ondanks alle aandacht die er voor blijkt te bestaan, begrijpen toch nog maar weinig ondernemingen wat 'big data' inhoudt en wat het voor de toekomst kan betekenen (zie kader 'Big data – Definiëring'). Bij veel 'big data'-initiatieven lijkt het vooralsnog te gaan om experimenten. Men 'ruikt' als het ware eens aan het onderwerp, maar bouwt nog niet echt aan de benodigde bedrijfs-capaciteit. Natuurlijk mag geen enkele organisatie de kansen, die 'big data' biedt, mislopen. Maar wie er werkelijk de vruchten van wil plukken, zal een visie moeten ontwikkelen op de verandering die 'big data' zal betekenen voor de eigen organisatie, capaciteiten en bedrijfstak.

De grote vraag is vooral waar je moet beginnen. Een onderneming die serieus aan de slag wil gaan met 'big data', zal zich vier vragen moeten stellen:

- Waar levert het analyseren van grote hoeveelheden gegevens de onderneming bij uitstek voordeel op?
- Hoe moeten wij ons organiseren om die voordelen ook daadwerkelijk te behalen?
- Met welke technologische investeringen kan de benodigde analytische capaciteit worden opgebouwd?
- Hoe kunnen wij dit hele traject het beste ingaan?

Figuur 1. Met 'big data' kunnen omzet en winst verbeteren



Waar levert het analyseren van grote hoeveelheden gegevens bij uitstek voordeel op?

Om er daadwerkelijk waarde uit te kunnen putten, moet een onderneming begrijpen hoe 'big data' concurrentievoordeel kan opleveren. Bij veel ondernemingen hebben inzichten ontleend aan 'big data' al gezorgd voor winstgevendende en aanhoudende groei op drie gebieden: klantcontact, productinnovatie en bedrijfsvoering.

Klantcontact

Met 'big data' komt de klant centraal te staan in de bedrijfsstrategie. Organisaties worden overspoeld door gegevens over de klant, uit allerlei bronnen: interactieve websites, fora op internet, databases van overheden en particuliere instellingen. Veelzeggend is vooral de informatie die ontstaat via sociale media als Facebook; de gebruikers daarvan genereren bijna dertig miljard delen 'content' per dag. Ook kunnen 'feeds' uit sociale media nu worden gecombineerd met allerlei andere soorten informatie; denk aan gegevens over het weer, culturele evenementen en bedrijfseigen gegevens, bijvoorbeeld uit het directe contact met klanten. Bovendien kan al deze informatie dankzij geavanceerde analytische instrumenten nu sneller, doeltreffender en goedkoper worden verwerkt. Dat maakt het mogelijk om snel nieuwe inzichten op te doen.

Maar dat wist u al. Laten we daarom eens aan de hand van enkele voorbeelden uit de praktijk verkennen hoe 'big data' zoal wordt ingezet. Bij een Amerikaanse bank worden aan de hand van berichten op sociale media de risico-klanten in kaart gebracht. Een bank in Azië analyseert audiobestanden van telefoongesprekken met klanten om hun ware gevoelens te bepalen. Met dit soort 'opiniepeilingen' kan feedback uit de eerste hand worden verkregen *zonder* de vertekening die veelal in enquêtes sluip. Bovendien kunnen er ook snel klantenprestatiedoelstellingen mee worden ontwikkeld. Er hoeft dan niet langer te worden gewacht op de maandelijkse rapportages om de strategische tactiek te kunnen bepalen.

Zo gebruikt de Amerikaanse warenhuisketen Macy's 'big data' om assortimenten af te stemmen op de klant. Voorheen analyseerde Macy's datapunten (voor winkelverkoop, voorraaduitputting of prijspromoties) binnen de merchandising hiërarchie. Met 'big data' kan Macy's deze datapunten echter per product of SKU¹¹ voor een specifiek tijdstip en een specifieke locatie analyseren. Daarmee kunnen duizenden scenario's worden gegenereerd om de kans te bepalen dat een bepaald product op een bepaald tijdstip en een bepaalde locatie zal worden verkocht. Assortimenten kunnen zo worden geoptimaliseerd naar locatie, tijd en winstgevendheid.

6. 'Big Data Meets Big Data Analytics', SAS, 2011. [Noot v. vert.: 1 petabyte (PB) = 1.000 terabytes (TB) ofwel 10¹⁵ bytes volgens de norm v.d. IEC (zie noot 2).]
7. 'Global Survey: The Business Impact of Big Data', *Avenade*, november 2010.
8. "Big Data" Facts and Statistics That Will Shock You', *Fathom Digital Marketing*, 8 mei 2012.
9. Zie 'IT Innovation Spurs Renewed Growth' op www.atkearney.com.
10. 'Big Data Market Set to Explode This Year, But What is Big Data?', *Smart Planet*, 21 februari 2012.
11. SKU: Stock-Keeping Unit, ofwel verpakking waarin een of meer eenheden consumentenproduct op voorraad worden gehouden (vert.).

Figuur 2. In allerlei sectoren wordt 'big data' gebruikt voor de transformatie van businessmodellen en het verbeteren van prestaties

VOORBEELDEN

Detailhandel	Productie
• CRM • Locatie en inrichting winkels • Opsporing/preventie van fraude • Optimalisering toeleveringsketen • Dynamische prijsstelling	• Productonderzoek • Technische analyse • Voorspellend onderhoud • Proces- en kwaliteitsanalyse • Distributie-optimalisatie
Financiële dienstverlening	Media- en telecommunicatie
• Algoritisch handelen • Risico-analyse • Opsporing fraude • Portfolio-analyse	• Netwerkoptimalisering • Scoren van klanten • Tegengaan klantverloop • Fraudepreventie
Reclame en public relations	Energie
• Vraagsignalering • Gerichte reclame • Sentiment-analyse • Klantenwerving	• Slimme netwerken • Exploratie • Modellen bouwen voor operationele processen • Sensoren voor elektriciteitsleidingen
Overheid	Gezondheidszorg en 'life sciences'
• Marktoezicht • Wapensystemen en terrorismebestrijding • Econometrie • Zorginformatica	• Farmacogenomie • Bio-informatica • Farmaceutisch onderzoek • Onderzoek klinische resultaten

Bron: Analyse A.T. Kearney

'Big data' – Definiëring

De aanduiding 'big data' verwijst in de regel naar de drie V's – 'volume', 'variety', 'velocity'* – van zowel gestructureerde als ongestructureerde gegevens die via netwerken in processoren en opslagapparaten belanden, en de verwerking van die gegevens tot bruikbare bedrijfsinformatie.

Volume – De hoeveelheden aan gegevens die tegenwoordig worden gegenereerd kunnen alleen met gespecialiseerde technologie worden verwerkt en beheerd. De technologische mogelijkheden zullen ten volle moeten worden benut om de kansen die 'big data' biedt te kunnen ontsluiten.

Velocity – De snelheid waarmee de grote gegevensvolumes worden verwerkt om bruikbare resultaten te verkrijgen.

Variety – De verscheidenheid aan soorten gegevens en aan gegevensbronnen waaruit kan worden geput om inzichten op te doen.

In essentie verwijst 'big data' naar de geïntegreerde combinatie van infrastructuur, gegevensbronnen, software en vaardigheden die aan de basis ligt van de drie V's. Met behulp van dit 'apparaat' kan een onderneming sneller relevantere analyses uitvoeren dan mogelijk is met traditionele methoden voor 'business intelligence'.

* **volume, variëteit, snelheid**

Voor online-aanbieders maakt 'big data' het mogelijk om het aanbod onmiddellijk af te stemmen op de bezoeker van de website. Amazon doet dit al jaren. Daar krijgt de consument die een bepaald product zoekt ook nog andere producten voorgeschoteld, waar andere consumenten met dezelfde productinteresse verder naar hebben gekeken. Bij reclame in meer traditionele media helpt 'big data' met de plaatsing van advertenties en spotjes zodat ermee kan worden bepaald welke tv-programma's (bijvoorbeeld) het meeste effect sorteren bij verschillende klantsegmenten. De technologie gaat tegenwoordig veel verder dan patroonvergelijking. Bluefin Labs gebruikt geavanceerde technologie om video's te 'bekijken' en naar audio te 'luisteren'; daarmee worden per klantsegment specifieke voorkeuren voor reclamespots per genre tv-programma in kaart gebracht. Door de mar-

keting af te stemmen op micro-segmenten kunnen investeringen in reclame beter laten renderen.

'Big data' was zelfs een factor bij de Amerikaanse presidentsverkiezingen van 2012. De Democratische Partij bracht enorme hoeveelheden gegevens over campagne-voering, fondsenwerving, vrijwilligers en sociale media bij elkaar in een centrale database. Zo konden de Democraten aan de hand van de activiteiten online van individuele kiezers bepalen in hoeverre hun campagnemethoden succes hadden.¹² Met behulp van een 'target sharing'-programma werd toegewerkt naar een buitengewoon hoge opkomst, en die bleek doorslaggevend te zijn voor de overwinning.

Productinnovatie

Niet alle 'big data' zijn nieuwe gegevens. Heel veel opgeslagen informatie die in voorbije jaren is gegenereerd wordt niet of ondoeltreffend gebruikt. Neem ZestCash, een kredietverschaffer aan mensen met een laag inkomen. Dit bedrijf verbeterde zijn winstmarge met twintig procent door het risico van klanten voortaan te bepalen aan de hand van gegevens over het gebruik van hun mobiele telefoon.

Over de hele wereld ontstaan er nieuwe mogelijkheden om met 'big data' gegevens uit allerlei bronnen bij elkaar te brengen en te organiseren, of om informatie te putten uit andere kanalen dan internet. Het Amerikaanse bedrijf Acxiom heeft vijftienhonderd datapunten over vijfhonderd miljoen klanten; deze informatie is verkregen uit openbare gegevensbestanden, enquêtes, en vijftig biljoen jaarlijkse data-transacties.¹³ Dit is slecht nieuws voor bedrijven die gespecialiseerd zijn in het leveren van gegevens die vroeger lastig te vinden waren. Informatie die ooit beschermd eigendom was – over onroerend goed, bijvoorbeeld – kan nu worden opgevraagd uit openbare en publieke bronnen. Een onderneming kan met bepaalde gegevens nieuwe omzet aanboren en producten verbeteren. GE werkt aan een nieuw soort 'verbonden apparaten' – denk aan straalmotoren, CT-scanners en aggregaten met sensoren – die terabytes aan gegevens terugsturen naar de technici van het concern. GE wil met deze informatie zijn producten efficiënter maken; klanten kunnen er miljarden dollars per jaar mee besparen, en GE zelf heeft er een nieuwe bron van omzet aan.¹⁴

Ook crowd-sourcing en andere vormen van productinnovatie op basis van sociale interactie zijn mogelijk geworden dankzij 'big data'. In-

12. 'Corporations Want Obama's Winning Formula', *Bloomberg Businessweek*, 21 november 2012.

13. 'Mapping and Sharing the Consumer Genome', *The New York Times*, 16 juni 2012. [biljoen: duizend miljard, ofwel 10^{12} ; vert.]

14. 'GE Tries to Make Its Machines Cool and Connected', *Bloomberg Businessweek*, 6 december 2012.

formatie uit honderden miljoenen 'tweets', een kakafonie aan ongestructureerde gegevens, kan nu worden vertaald in inzichten over producten en diensten die weerklank vinden bij de consument. Aan de basis van dit alles ligt het vermogen om met geavanceerde geautomatiseerde taalkunde te analyseren hoe er op een bepaald moment wordt gedacht over de producten en klanten van een onderneming. De uitkomsten van dergelijke analyses kunnen worden gebruikt bij het ontwikkelen van de marketing- en innovatiestrategieën voor producten.

Gegevens en gegevensanalyse worden ook een product op zich. Er zijn technologische en analytische bedrijven ontstaan die voor anderen waardevolle inzichten uit gegevens halen. Zo'n bedrijf kan bijvoorbeeld voor een winkelketen de gegevens verzamelen en analyseren over transacties met leveranciers. Met de inzichten die dit oplevert, kan die winkelketen de bedrijfsvoering verbeteren, aanvullende diensten voor klanten ontwikkelen en zelfs andere dienstverleners verdringen. Op deze manier kunnen volstrekt nieuwe bronnen van omzet worden aangeboden.

Bedenk tot slot eens wat 'big data' kan betekenen voor het doen van experimenten. Je neemt een probleem waar de onderneming mee zit, of een hypothese. Daar laat je grote gegevensbestanden op los; je brengt gegevens onder in modellen, en integreert en analyseert ze. Zo bepaal je wat werkt en wat niet. Je verfijnt vervolgens het proces, en herhaalt het. Facebook voert dagelijks duizenden experimenten uit waarbij een bepaalde groep gebruikers andere aspecten voorgeschoteld krijgt dan de overige gebruikers. Amazon biedt verschillende producten en verschillende prijzen aan verschillende klanten en voert aan de hand van de bevindingen aanpassingen door.

Bedrijfsvoering

Gegevens over interacties in de toeleveringsketen kunnen veel waardevolle informatie opleveren. De verplaatsing van een fysiek product, bijvoorbeeld, kan worden gemeten met behulp van radiofrequentie-identificatie (RFID) en microsensoren. In het programma Value Chain Visibility van Airbus worden processen, materialen en verplaatsingen van activa bewaakt met behulp van RFID-lezers, bewegingssensoren en transporteurs. Dit programma omvat leveranciers, productievestigingen, klanten en dienstenpartners en verkleint voorraden, verhoogt de productiviteit en verlaagt de kosten.

Bij een winkelketen wordt aan de hand van gedetailleerde SKU-voorraadinformatie gecontroleerd of de ene winkel misschien te veel voorraad heeft, terwijl een andere winkel die zou kunnen verkopen. Het oude systeem voor snelle retour-logistiek liet alleen de honderd SKU's met de *grootste* voorraadoverschotten zien. Met 'big data' wordt echter

alle beschikbare informatie meegenomen – enkele terabytes aan operationele gegevens. Nu kan een alomvattend model worden gemaakt van *alle* SKU's die verspreid over

duizenden winkels op voorraad liggen. Bij deze winkelketen werden voorraden ter waarde van honderden miljoenen dollars razendsnel verplaatst van winkels met te veel voorraad naar andere winkels. Vervolgens werd een voorspellend distributiemodel ontwikkeld om in de hele winkelketen voorraadoverschotten zoveel mogelijk te voorkomen. Inmiddels worden er ook prijs-, promotie- en klantenkaartgegevens verzameld om nog diepgaandere inzichten te ontwikkelen over wat klanten kopen, op welke momenten, en waarom.

Bill Ruh, hoofd van GE's nieuwste centrum voor onderzoek en ontwikkeling in Silicon Valley, zegt in een interview met *Bloomberg Businessweek* dat hij 'big data' wil 'combineren met enkele van GE's grootste activiteiten'. Neem bijvoorbeeld de luchtvaartmaatschappijen die hun toestellen uitrusten met GE-motoren. Met behulp van 'big data' kan GE hen helpen om de prestaties van die motoren te bewaken. Zo kan tijdig worden ingeschat wanneer er onderhoud moet worden gepleegd. Hetzelfde geldt voor bedrijven die bedrijfsvoertuigen leasen via GE Capital. 'Je signaleert dat iets het minder goed doet, en repareert dat voordat het echt kapot gaat. Dat maakt een wezenlijk verschil,' aldus Ruh. 'Voorspelbaarheid, dat is wat iedereen wil.'¹⁵

Dit soort kansen pak je alleen als je beseft dat gegevens een bezit zijn waar je geld mee kunt verdienen. Informatie over klanten die ongebruikt ligt opgeslagen is een verspilde kans. De waarde van gegevens is bovendien niet zozeer de informatie die je eruit haalt, maar de kijk op die informatie vanuit het perspectief van de klant. Het gaat dus niet alleen om het effect op de winstgevendheid, maar ook om de vraag hoe de markt wordt beïnvloed door veranderingen in consumentenvoorkeuren. In hoeverre zal, bijvoorbeeld, een promotie-actie

**Niet alle 'big data' is
nieuw – veel informatie uit
voorbij jaren wordt niet,
of niet effectief, gebruikt**

15. 'GE's Billion-Dollar Bet on Big Data', *Bloomberg Businessweek*, 26 april 2012.

de verkoop van een bepaald product kunnen bevorderen? Om dat in te schatten, kun je gegevens verzamelen over promotie-acties bij concurrenten, en dan vooral voor producten die een alternatief vormen voor het eigen product. Zo kun je in kaart brengen hoe de voorkeuren van de consument zich hebben ontwikkeld.

Hoe moeten wij ons organiseren om die voordelen te behalen?

‘Big data’ staat dus voor het verwerken van grote en uiteenlopende gegevensbestanden om complexe verbanden bloot te leggen. Uiteindelijk is het echter aan de mens om deze gegevens en verbanden te interpreteren en te vertalen in inzichten.

Deze interpretaties zijn in toenemende mate de taak van een nieuwe generatie analisten – de rekenmeesters van de 21e eeuw in het bedrijfsleven. Vaak worden zij ‘data scientists’ (letterlijk: gegevenswetenschappers) genoemd. Deze data scientists zijn een nieuw slag mensen: ze hebben geavanceerde statistische en wiskundige kennis, maar zijn ook bedrijfseconomisch onderlegd. Ze werken nauw samen met het management om inzichten te ontwikkelen op basis waarvan nieuwe strategische beslissingen kunnen worden genomen. Voor een manager is ‘business intelligence’ slechts

een onderdeel van het dagelijkse takenpakket. Een data scientist, daarentegen, doet vrijwel niets anders dan zakelijke inzichten distilleren uit gegevens.

Data scientists lopen voorop bij de grote veranderingen die zich voltrekken in de manier waarop ondernemingen de markt bewerken. De organisatiestructuur en analytische processen van de onderneming zullen dus ook met het oog op de rol van de data scientists moeten worden ingericht.

Ontwikkelen van businessmodellen

Naarmate organisaties zich verder ontwikkelen, zullen hun analytische capaciteiten zich eveneens moeten ontwikkelen: van elementaire respectievelijk vooruitblikkende naar een meer volwassen voorspellende analyse (zie figuur 3). Elementaire analyse biedt een historische kijk op de bedrijfsprestaties: wat er gebeurde, waar het gebeurde, en hoe vaak het gebeurde. Met vooruitblikkende analyse worden unieke bevorderlijke factoren (‘drivers’), fundamentele oorzaken en gevoeligheden in kaart gebracht. Met voorspellende analyse wordt de bedrijfsvoering ondergebracht in een model en wordt geprobeerd te voorspellen wat er zal gebeuren.

De uitdaging is om als organisatie niet langer te

Figuur 3. Optimale realisatie ‘big data’-potentieel vereist ontwikkeling van analytische capaciteiten en processen

Ontwikkelingsstadium analytisch proces	Methoden	
Elementair - Biedt statische, historische kijk op bedrijfsprestatie - Gebaseerd op elementaire scorecards en statische rapportages	Vragen en uitzoeken	Waar zit het probleem?
	Ad-hoc-rapportages	Hoe veel? Hoe vaak? Waar?
	Standaardrapportages	Wat is er gebeurd?
Vooruitblikkend - Biedt inzicht in voorbijgaande en mogelijke toekomstige prestatie-beïnvloedende factoren - Benut systemen en processen om uiteenlopende beschrijvende analyses uit te voeren	Segmenteringsanalyse	Wat zijn de unieke beïnvloedende factoren?
	Statistische analyse	Waarom gebeurt dit?
	Gevoelighedsanalyse	Wat als de omstandigheden veranderen?
Voorspellend - Biedt dynamische, toekomstgerichte inzichten met gekwantificeerde afwegingen - Vereist kwalitatief hoogwaardige en geïntegreerde capaciteiten m.b.t. gegevens en complexe wiskunde	Optimalisering	Wat is het beste dat kan gebeuren?
	Simulatie	Wat zou er gebeuren indien...?
	Voorspellende modellen	Wat zou er vervolgens kunnen gebeuren?

Bron: Analyse A.T. Kearney

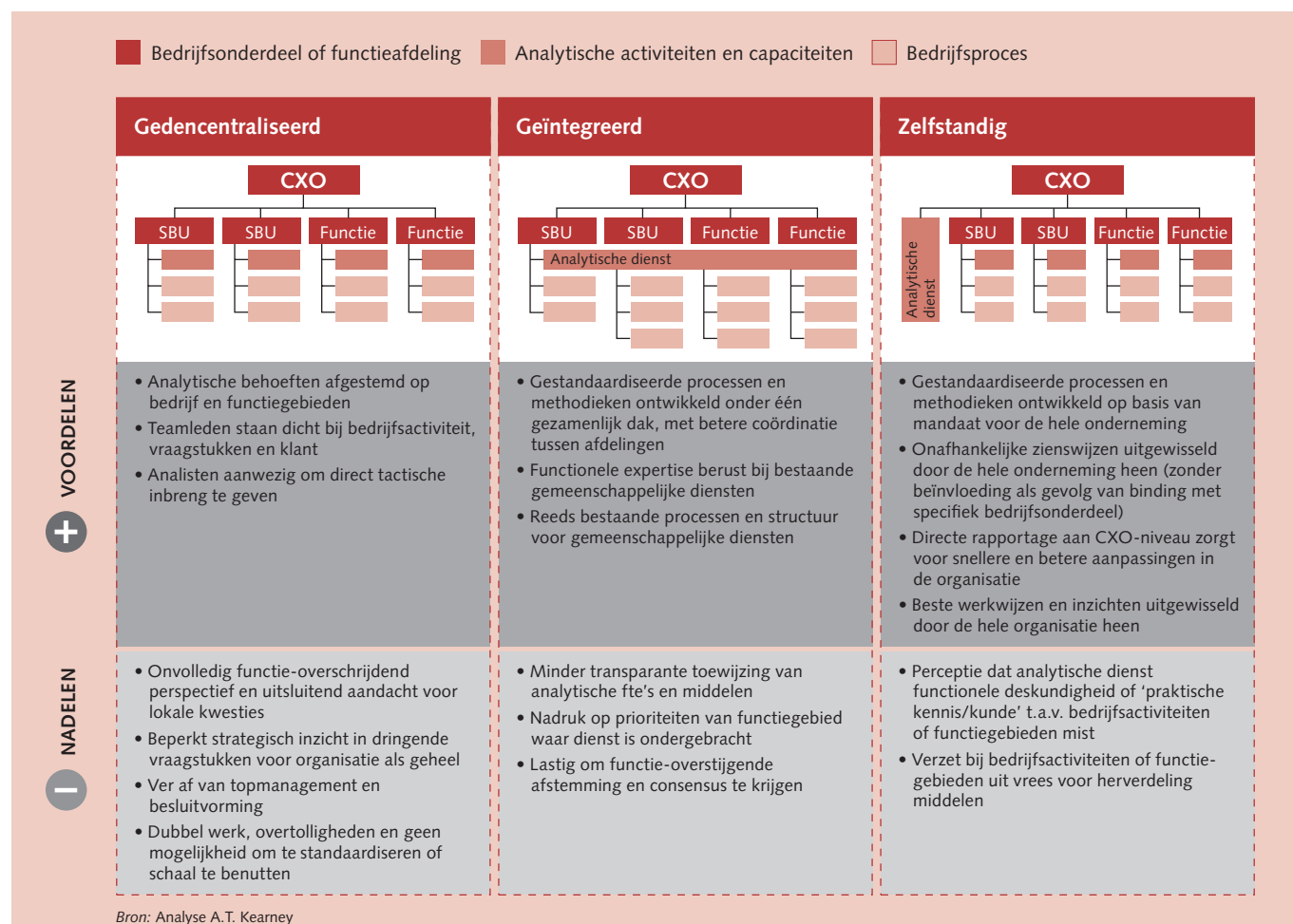
denken in termen van gegevens en informatie, maar om gericht bedrijfsmatig relevante inzichten met voorspellende waarde te gaan ontwikkelen. Om die omslag te maken, kunnen drie potentieel waardevolle operationele modellen worden overwogen: decentrale diensten, een geïntegreerde gemeenschappelijke dienst, of een zelfstandige gemeenschappelijke dienst. Elk van deze drie modellen heeft bepaalde voor- en nadelen (zie figuur 4). In het model van *decentrale diensten* heeft iedere bedrijfsactiviteit of functie een eigen analytische afdeling. Dit faciliteert en bevordert een snelle besluitvorming en uitvoering. Meestal is er dan echter geen aparte strategische planning en worden er ook geen ‘best practices’ uitgewisseld, met als mogelijk gevolg dubbel werk en verspilling van middelen en infrastructuur. In dit model is de focus weliswaar groter, maar kunnen kansen worden misgelopen doordat het perspectief van de onderneming als geheel ontbreekt.

Bij een *geïntegreerde gemeenschappelijke dienst* is er sprake van een gecentraliseerd model dat voor de hele onderneming wordt uitgevoerd vanuit een bestaande afdeling of business-unit. Besluitvorming en uitvoering kunnen ermee worden versneld. Ook leiden de structuur, ondersteunende processen en normen van zo’n centrale aanpak tot een grotere efficiëntie en IT-expertise. Maar de rol van deze dienst wordt eerder uitgevoerd als klant dan als partner.

Het model van een *zelfstandige gemeenschappelijke dienst* lijkt op dat van de geïntegreerde gemeenschappelijke dienst, maar nu is die dienst buiten de bedrijfsonderdelen en -functies geplaatst. Deze externe dienst rapporteert direct aan het topmanagement en maakt van de analyse eerder een cruciale kerncompetentie dan een uitvoeringsgerichte capaciteit.

Het doeltreffendste businessmodel voor ‘big data’ bestaat vaak uit decentrale diensten voor de ont-

Figuur 4. Drie organisatiemodellen voor de analytische afdeling



wikkeling van bedrijfskennis ('business intelligence'), in combinatie met een zelfstandige gemeenschappelijke analytische dienst. Met bedrijfskennis verankerd in de business-units ontwikkelt de hele organisatie snel prestatie-inzichten. In de bedrijfscultuur bevordert dit bovendien een be-

Decentrale diensten voor business intelligence in combinatie met een centrale analytische dienst werkt vaak het beste

sluitvorming op basis van hypothesen getoetst aan harde feiten, in plaats van besluitvorming op basis van een 'onderbuikgevoel'. Een centrale analytische dienst zorgt in deze benadering voor een perspectief op de onderneming als geheel, en voor een voorspellende in plaats van een reactieve beschouwing van de kansen waarop kan worden ingespeeld.

Wie werkt het beste met 'big data'?

Veel rollen die bij 'big data' horen, bestaan al in menige organisatie. De reikwijdte, zichtbaarheid voor het topmanagement en benodigde technische vaardigheden van die rollen zouden echter kunnen worden aangescherpt. Bovendien zullen degenen die deze rollen vervullen over bepaalde eigenschappen moeten beschikken, zoals een bereidheid om te experimenteren en het vermogen om de toekomst 'in een model te vangen' op basis van feiten.

Volgens de CEO van Procter & Gamble, Bob McDonald, is 'de manier waarop we innoveren (...) wezenlijk aan het veranderen door het gebruik van modellering, simulaties en ander digitaal gereedschap'. Dit stelt andere eisen aan de vaardigheden van medewerkers. P&G heeft voor elk niveau in de organisatie de 'minimale digitale vaardigheden' in kaart gebracht waarover iemand moet beschikken, aldus McDonald. Managers voor 'business intelligence' (BI) weten hoe zij informatie moeten vertalen in strategie voor een business-unit of de onderneming als geheel. Bedrijfsanalisten ('business analysts') weten hoe zij vraagstukken kunnen oplossen door hypothesen te toetsen, en zoeken vanuit een afdelingsoverstijgend perspectief in gegevens naar kansen voor de onderneming als geheel. Gegevensanalisten ('data analysts') bouwen met ongestructureerde gegevens complexe statistische modellen. En informatiemanagers ('data managers') richten de IT-infrastructuur opnieuw in om een 'bedrijfsbreed' perspectief te krijgen; daarbij worden grotere volumes aan (ongestructureerde) gegevens verwerkt en

activiteiten gemodelleerd ten behoeve van de besluitvorming.¹⁶

Zoals gezegd zijn data scientists erg gewild in het bedrijfsleven. Zij kunnen grote bestanden aan gegevens analyseren om er trends uit te distilleren, bouwen modellen met voorspellende waarde en signaleren kansen op prestatieverbetering voor de onderneming. Data scientists zijn hoog opgeleid (en bij voorkeur zelfs gepromoveerd) in vakken als automatisering, statistiek, toegepaste wiskunde of andere relevante terreinen; hun kwalificaties liggen vaak veel hoger dan die van informatiemanagers en gegevensanalisten. Idealiter kunnen zij ook als programmeur grote gegevensbestanden verwerken en analyseren. Verder kunnen ze goed communiceren en beschikken ze veelal over een natuurlijk vermogen om bedrijfsmatige vraagstukken te vertalen in zinvolle analyse.

Welke technologische investeringen zijn nodig?

Een van de uitdagingen bij het inzetten van 'big data'-capaciteiten is om de IT-architectuur tegen aanvaardbare kosten om te bouwen. Met andere woorden: de aanvullingen moeten het mogelijk maken om het bestaande optimaal te benutten. De mogelijkheden zijn legio, maar het gaat erom, te investeren in technologie die de analytische inspanning optimaal laat renderen. Een juiste investeringskeuze vereist een volledig beeld van de 'big data'-architectuur waarmee de onderneming haar analytische visie in praktijk kan brengen. Het rendement op analyse stijgt naarmate de complexiteit van gegevens verandert – van gestructureerd in ongestructureerd, van 'schone' interne naar 'vervuild' externe gegevens, en van gegevens die slechts één kant uit stromen naar een uitwisseling van gegevens tussen de onderneming en haar klanten en partners. Al deze verschillende soorten gegevens moeten efficiënt kunnen worden verwerkt door de analytische en 'big data'-architectuur.

Een technologische architectuur voor 'big data' bestaat uit vijf niveaus, vergelijkbaar met een gebruikelijke inrichting: presentatie, toepassing, verwerking, opslag en integratie (zie figuur 5). Het presentatieniveau biedt de primaire bediening voor proces en beheer, de verspreiding van gegevens via 'dashboards' voor de rapportage, en de visualisering van gegevens. Op het toepassingsniveau, waar de bedrijfslogica ligt, worden analyses uitgevoerd en modellen gemaakt en wordt bedrijfskennis ('business intelligence') opgebouwd. Op de verwerkings- en opslagniveaus

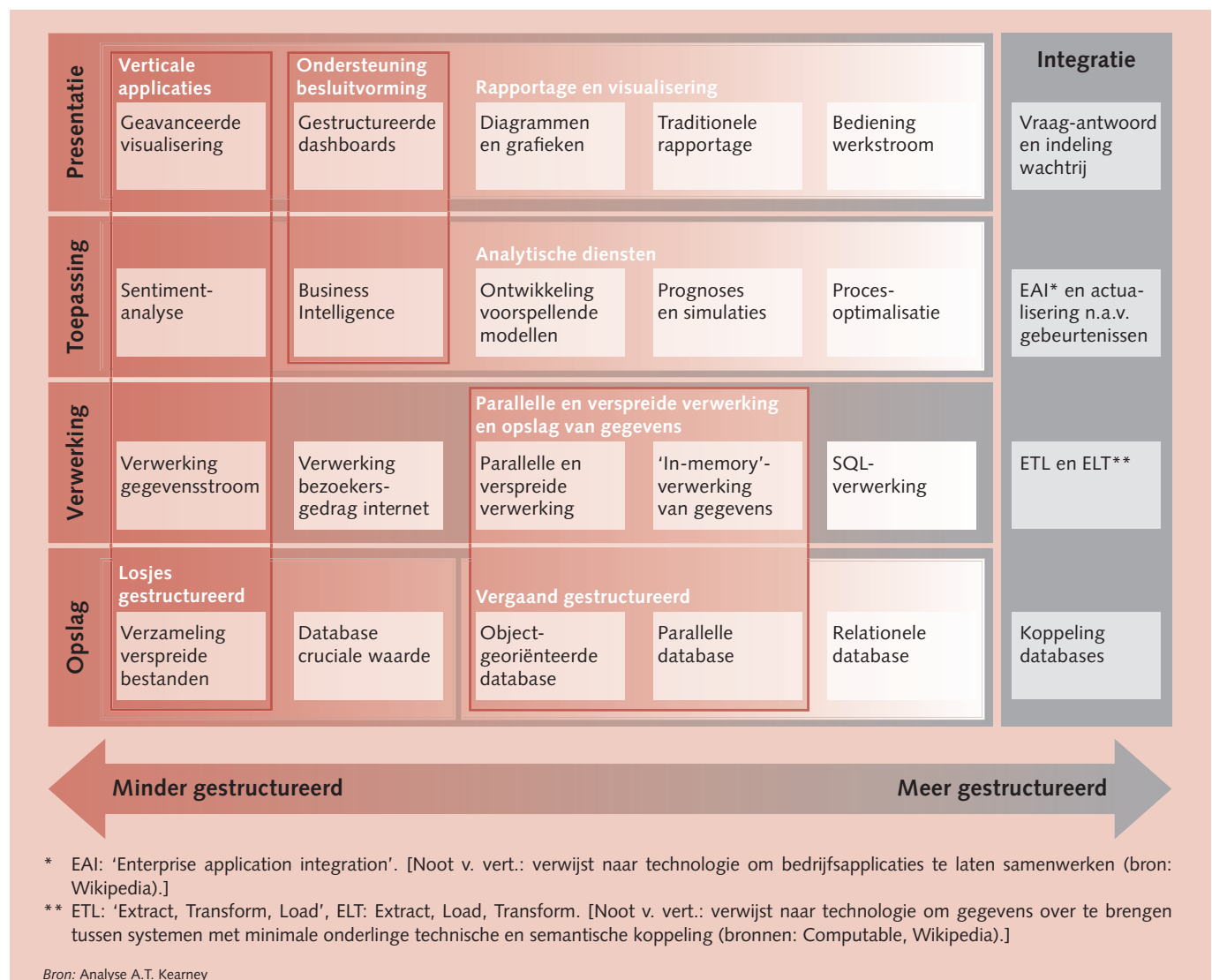
16. 'Data Is Useless Without Skills to Analyze It', *Harvard Business Review*, 13 september 2012.

worden gestructureerde en ongestructureerde gegevens opgeslagen zodra deze ontstaan, of kort daarna. Op het integratieniveau, tot slot, worden de verschillende bronnen verspreid over de hele organisatie ontsloten voor de 'big data'-architectuur. De verschillende niveaus, elk met specifieke functies om gegevens te verwerken, grijpen in elkaar om de gewenste uitkomsten op te leveren. Op elk niveau zijn specifieke taken en functies met betrekking tot gegevens ondergebracht. Deze worden geclusterd conform de marktsegmenten voor nieuwe 'big data'-producten:

- *Verticale toepassingen* of productpakketten (Hadoop bijvoorbeeld) leveren en verwerken gegevens en bevatten ingebouwde instrumenten voor rapportage en visualisering.
- *Besluitvormingsondersteuning* biedt traditionele dashboards en systemen voor 'business Intelligence (BI) versterkt door 'big data'-toepassingen en visualiseringsinstrumenten.
- *Rapportage- en visualiseringsinstrumenten* geven de uitkomsten van 'big data' op een begrijpelijke manier weer.
- *Analytische diensten* bieden opslag, verwerking en toepassingen ter ondersteuning van de door 'big data' mogelijk gemaakte oplossingen, zoals voorspellend modelleren, het ontwikkelen van prognoses, simulaties.
- *Parallel verspreide verwerking en opslag* maken het mogelijk om op grote schaal gelijktijdig gegevens te verwerken ('massively parallel processing', of MPP) en meer gestructureerde gegevens

Figuur 5. 'Big data'-architectuur

ILLUSTRATIEF



direct via geheugens te analyseren ('in-memory analytics').

- *Losjes gestructureerde opslag* verzamelt minder ver- gaand gestructureerde gegevens en slaat deze op.
- *Vergaand gestructureerde opslag* verzamelt traditi- onele databases en legt deze vast, met inbegrip van hun parallelle en verspreide versies.

Belangrijke technologieën

De markt wordt overspoeld met innovatieve pro- ducten voor het beheren, verwerken en analyseren van 'big data', waaronder het zoëven al genoemde MPP, 'columnar databases', en NoSQL-databases voor het beheren van gestructureerde gegevens.¹⁷ Vooraanstaande leveranciers van MPP-oplossingen – Greenplum, Vertica, Aster Data, Netezza en Splunk, bijvoorbeeld – bieden hard- en software in één gecombineerd systeem. Om ongestructureerde gegevens te verwerken – bijvoorbeeld gegevens uit sms'jes, sociale media, mobiele technologie en sen- soren in apparaten – kloppen ondernemingen aan bij het 'open source'-netwerk Hadoop, dat honder- den terabytes, petabytes of zelfs zettabytes aan ge- gevens kan verwerken.

De meeste ondernemingen kiezen voor de com- merciële Hadoop-producten van Greenplum, MapR, Cloudera, IBM en andere aanbieders. Daarmee kunnen gemakkelijk allerlei verschillen- de bestanden met ongestructureerde gegevens worden verzameld. Zo is het inmiddels mogelijk om aan de hand van gegevens ontleend aan Facebook en Twitter de actuele klanttevredenheid in te schatten. Die informatie kan vervolgens wor- den gecombineerd met andere gegevens, zoals ge- ografische locatie, doorklikgegevens op websites en informatie uit winkelkassa's, om nog meer nieuwe toepassingen op basis van 'big data' te ont- wikkelen.

Ten behoeve van deze 'big data'-toepassingen is een nieuw 'ecosysteem' aan software-applicaties ontstaan (zie kader 'Trends in Big Data-techno- logie'). Deze applicaties zijn ontwikkeld voor uiteenlopende doelen, variërend van prognoses en simulaties tot voorspellende modellen, en ze wor- den ondersteund door werkstroom-engines, dash- boards, batch-rapportages en instrumenten voor het maken van diagrammen en grafieken. Verder zijn er allerlei visualiseringsinstrumenten op de markt verschenen, onder meer om diagrammen te maken voor netwerken met meervoudige knoop- punten, om geometrische informatie grafisch weer te geven, en voor multidimensionale grafische overzichten.

Ook zijn er allerlei testomgevingen beschikbaar ('proof of concept'-omgevingen). Zo'n 'PoC- omgeving' moet een afzonderlijke maar parallelle omgeving bieden waarin onbelemmerd kan wor- den getest zonder de bestaande systemen te ver- storen. Veel producten zijn de moeite waard om te overwegen, maar Hadoop staat op de meeste testlijsten vanwege de kostenefficiënte verwerking en geschiktheid voor minder gestructureerde gegevens.

'Big data' zal vermoedelijk niet lang een aparte technologie zijn. Naarmate het een onderdeel wordt van het gehele analytische 'ecosysteem' van een onderneming, zal het verschil tussen 'big data' en traditionele gegevens vervagen. 'Big data' wordt een onderdeel van een alomvattende en geavan- ceerde analyse-strategie.

Trends in 'big data'- technologie

Ondanks de vele zelfstandige ondernemingen op het gebied van 'big data' verandert deze markt van dag tot dag naarmate meer partijen in 'data management' hun leveranciers overnemen: EMC kocht Greenplum, IBM nam Netezza over, en HP kreeg Vertica in handen. Ook ontstaan er allerlei samenwerkingsverbanden om bestaande capaciteiten aan te vullen met verwerking van ongestructureerde gegevens; prominente voor- beelden hiervan zijn de samenwerking van Microsoft met Hortonworks en de alliantie van Oracle met Cloudera.

Hardware-fabrikanten integreren verticaal door soft- ware-oplossingen in huis te halen waarmee zij kunnen inspelen op de vraag in de markt op korte termijn. Op langere termijn zullen echter grotere integraties en investeringen nodig zijn. Bestaande BI-leveranciers zullen vermoedelijk inzetten op 'big data'-oplossingen waarbij de beste technologieën voor gestructureerde en ongestructureerde gegevens worden gecombineerd. Voor deels gestructureerde gegevens wordt Hadoop ingezet op standaard hardware en tegen minimale kosten, vergelijkbaar met Apache en Linux zo'n tien jaar geleden.* Hadoop kan echter niet zonder meer worden ingezet in een onderneming; voor de im- plementatie en ondersteuning ervan zijn hoog opge- leide professionals nodig. Er komen nu op Hadoop gebaseerde toepassingen, en Greenplum, IBM en andere partijen werken aan de ontwikkeling van standaardproducten voor de verwerking van slechts ten dele gestructureerde of volstrekt ongestructureerde gegevens. Om een voorbeeld te geven: Cassini, de

17. *MapReduce and MPP: Two Sides of the Big Data Coin*, *ZDNet*, 2 maart 2012.

nieuwe zoekmachine van eBay, is gebaseerd op 'big data'-technologie en geschikt voor 97 miljoen actieve kopers en verkopers, 250 miljoen zoekopdrachten per dag en 200 miljoen actuele artikelen in ruim 50.000 categorieën.**

Uiteindelijk zal de informatie-infrastructuur voor de gehele onderneming voldoen aan de eisen voor opslag, verwerking en toepassing van 'big data'. Dan zullen de bestaande systemen voor gegevensopslag en -verwerking er geschikt voor zijn gemaakt. Het analytisch vermogen van applicaties zal sterk toenemen. In systemen voor het inschatten van de vraag in de markt zullen bijvoorbeeld naast aspecten als promoties en kortingen ook prijzen van concurrenten, verstoringen in de vraag, nul-voorraden en toeleveringsinformatie worden meegewogen om prijzen onmiddellijk ('real time') te kunnen aanpassen.

* 'Hadoop Could Save You Money Over a Traditional RDBMS', *Computerworld UK*, 10 januari 2012.

** 'eBay Readies Next Generation Search Built with Hadoop and HBase', *InfoQ*, 13 november 2011.

De start van een 'big data'-traject

Achter iedere succesvolle 'big data'-implementatie schuilt een al even succesvol programma voor organisatieverandering. Dat was ook het geval bij een grote winkelketen die onze firma recentelijk assisteerde. Gecorrigeerd voor mutaties in het aantal verkooppunten stond de omzet bij deze keten al jaren onder druk. De concurrentie werd

alsmaar feller. 'De online-retailers eten ons de kaas van het brood,' klaagde een van de directieleden. Een zwakke conjunctuur, veranderend consumentengedrag, nieuwe concurrenten, meer distributiekkanalen en meer gegevens hadden allemaal een effect. Er was sterk ingezet op de ontwikkeling van e-commerce en het online-kanaal. Miljoenen dollars waren besteed aan eenmalige projecten die de problemen moesten oplossen. Maar niets werkte. Diverse factoren, waaronder investeringen van concurrenten, de sterke stijging van het volume aan gestructureerde en ongestructureerde gegevens, en de behoefte aan meer inzichtelijke gegevens, deden deze onderneming besluiten om te gaan concurreren op basis van gegevensanalyse.

Een wezenlijke ombouw van analytische capaciteiten en de inrichting van 'big data'-platforms begint met een doordachte aanpak bestaande uit drie speerpunten: bepaal waar de grote kansen liggen, ontwikkel capaciteitsscenario's, en stippel het ontwikkelingspad uit (zie ook figuur 6).

Bepaal op welke punten 'big data' de concurrentiestrijd wezenlijk kan doen veranderen

Om uit het dal te kunnen komen, moest onze grote winkelketen nieuwe capaciteiten zien te ontwikkelen. Die capaciteiten moesten zorgen voor concurrentievoordeel dat bestand was tegen de onophoudelijke verandering, volatiliteit en grotere concurrentie in de markt. Het team onderzocht

Figuur 6. Opstarten van 'big data'-traject



alle aspecten van de bedrijfsvoering: van assortimentsbeleid, opstellen van prognoses en inkoop – tot en met distributie, toewijzing en transport. Overal werd bekeken op welke specifieke aspecten analyse voor betere prestaties kon zorgen. De nadruk lag op voorspellen in plaats van verklaren. Een voorbeeld: in plaats van zich af te vragen *waarom* er minder afbakpizzas werden verkocht, probeerde men voortaan dergelijke veranderingen in het afzetvolume in de tijd en per afzetgebied te *voorspellen*. En in plaats van problemen met veiligheid pas aan te pakken wanneer ze zich voordeden, wilde men die problemen voortaan op voorhand kunnen zien aankomen. Aan de hand van gegevens uit sociale media zouden sluimerende of ontkiemende problemen voortaan proactief worden opgespoord. Hierdoor zou de onderneming niet alleen klantgerichter worden maar in de toekomst crises ook vóór kunnen zijn. Besloten werd om een organisatie voor bedrijfsinformatie op te zetten die vier doelen moest helpen bereiken:

Naarmate 'big data' een onderdeel wordt van het gehele analytische 'ecosysteem' zal het verschil tussen 'big data' en traditionele gegevens vervagen

– Informatievoorziening afgestemd op de specifieke behoeften in de hele organisatie.
– Nieuwe vaardigheden om aanhoudend met succes mee te concurreren.
– Een gezamenlijk analytisch platform voor de hele organisatie.
– Een consistent overzicht van de verkoop in de verschillende kanalen en geografische afzetgebieden.

Ontwikkel capaciteitsscenario's voor de toekomst

De winkelketen wilde toekomstscenario's voor benodigde capaciteiten ontwikkelen, gelet op de totale kosten, risico's en flexibiliteit, maar ook gelet op de bedrijfscultuur. Zijn, bijvoorbeeld, harde gegevens doorslaggevend bij een bedrijfsactiviteit? Of voelt men zich comfortabel met een analytische aanpak op basis van hypothesen en experimenten? Beide behoren tot de essentie van 'big data'. Ook hebben we voor elk scenario de afwegingen in kaart gebracht, zoals een vergelijking van capaciteiten, prioriteiten bij de omschakeling, en inschattingen van de doorlooptijd. Wat werkt bijvoorbeeld het beste: een wereldomspannende topologie voor gegevens op het hoofdkan-

toor, of een combinatie van lokaal, regionaal en mondiaal? Wat zijn de voor- en nadelen van Hadoop respectievelijk Cassandra voor een toekomstige 'big data'-architectuur? Deze afwegingen werden gemaakt met het oog op cruciale kansen. Denk aan de mogelijkheid om baanbrekende technologie in te zetten, een samenwerkingsplatform op te bouwen, geavanceerde analyse in te bouwen in zowel bestaande als toekomstige architectuur, en een platform voor uiteenlopende soorten analyses te ontwikkelen dat kon meegroeien met de organisatie. Deze technologie zou vijf cruciale capaciteiten mogelijk maken en de basis leggen voor toekomstige voordelen:

- Voorspellen van winkelbezoek- en aankooppatronen van klanten.
- Ontwikkelen van prijsstelling, ruimte en assortiment afgestemd op winkels.
- Bepalen en benutten van elasticiteiten, affiniteiten en tendensen bij prijsstelling.
- Optimaliseren van mondiale inkoop vanuit allerlei locaties en business-units.
- Ontwikkelen van modellen voor mogelijkheden om energieverbruik en uitstoot van broeikasgasen te verminderen.

Bepaal de voordelen en stippel de routekaart naar de toekomst uit

Na de beoogde capaciteiten moesten de benodigde middelen in kaart worden gebracht. Was het financieel verstandig om eigen middelen en mensen in te zetten? Of was het misschien kosteneffectiever om de 'big data'-analyses door externe partijen te laten doen, in elk geval in het begin? Deze beslissing hing natuurlijk af van de capaciteiten die in de onderneming zelf al voorhanden waren. De technologische behoeften werden vanuit twee invalshoeken gepland: de gegevens respectievelijk de architectuur. Er werd een 'data plan' opgesteld, van verwerving tot opslag en vervolgens presentatie, gebruik makend van een zelfbedieningsomgeving voor zowel gestructureerde als ongestructureerde gegevens. De systeemarchitectuur, waar een integratie op basis van Hadoop deel van kan uitmaken, werd gepland in het licht van de bestaande IT-architectuur, die zwaar leunde op relationele 'data warehouses' ontsloten via platforms van Teradata en Oracle. De routekaart naar de toekomst voorzag in investeringen van ettelijke miljoenen dollars die binnen vijf jaar moesten zijn terugverdiend. De onderneming is nu in de posi-

tie om vier belangrijke voordelen te behalen met haar 'big data'-strategie:

- Snellere en goedkopere oplevering van consistente informatie.
- Doeltreffender samenvatting van informatie en verspreiding ervan onder de business-units, met als doel prestatieniveaus en kansen beter te doorgronden en optimaal te profiteren van de mondiale organisatie.
- Repliceerbare ontwikkeling van 'business intelligence' (BI) en analyses, zodat de verschillende bedrijfsonderdelen niet voortdurend het wiel opnieuw hoeven uit te vinden om antwoorden te vinden op vergelijkbare vragen.
- Ontwikkeling van nieuwe waardescheppende inzichten door middel van geavanceerde analyse.

Kiezen voor creatieve destructie van businessmodellen

Is 'big data' de 'Industriële Revolutie' van de 21e eeuw? Wij denken van wel. Ondernemingen experimenteren en werken steeds meer met 'big data' om er zowel op korte als lange termijn concurrentievoordeel mee te behalen. Een eerste cruciale succesfactor is om gegevens te gaan beschouwen als een waardevol bezit van de onderneming, een activum dat de basis vormt voor marktproposities en businessmodellen. Vervolgens zal met zorg moeten worden gebouwd aan de capaciteiten waarmee de beloften van 'big data' kunnen worden waargemaakt. Maar het belangrijkste is misschien nog wel dat ondernemingen durven te kiezen voor de creatieve destructie van hun bestaande businessmodellen. ■

Reprinted from Christian Hagen, Khalid Khan, Marco Ciobo, Jason Miller, Dan Wall, Hugo Evans and Ajay Yadav, 'Big Data and the Creative Destruction of Today's Business Models', *Executive Agenda*, digital publication by A.T. Kearney (www.atkearney.com), with permission from the publisher. Copyright © 2012, A.T. Kearney, Inc. All rights reserved. Reprinted with permission. Vertaling: Raymond Gijsen.