Thomas Davenport

Big Data aan het werk

Wat je als manager moet weten

vakmedianet

Big data aan het werk



Thomas Davenport

Big Data
aan het werk

Wat je als manager moet weten

vakmedianet

Ontwerp omslag en binnenwerk: www.neonontwerpers.nl

NUR 801
© 2015 Vakmedianet,Deventer
www.overmanagement.nl
ISBN 978 94 62 76 0370

Oorspronkelijke titel: Big data@work
Original work copyright © Harvard Business School Publishing Corporation 2014
Published by arrangement with Harvard Business Review Press

Vertaling: Yvonne Halink & Ed Kerkman (www.taallijn.nl)

Alle rechten voorbehouden: niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van kopieën uit deze uitgave is toegestaan op grond van artikel 16B Auteurswet 1912 jo het Besluit van 20 juni 1974, Stb. 351, zoals gewijzigd bij het Besluit van 23 augustus 1985, Stb. 471 en artikel 17 Auteurswet 1912, dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 3060, 2130 KB Hoofddorp). Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet 1912) dient men zich tot de uitgever te wenden.

Inhoud

9 Dankwoord

13 1. Waarom big data belangrijk zijn voor jou en je organisatie
43 2. Hoe big data je baan, je bedrijf en je bedrijfstak zullen veranderen
69 3. Het ontwikkelen van een big-datastrategie
95 4. De menselijke kant van big data
123 5. Technologie voor big data
145 6. Wat ervoor nodig is om te slagen met big data
163 7. Wat je kunt leren van startups en online bedrijven
183 8. Wat je kunt leren van grote bedrijven: Big data en analytics 3.0

213 Bijlage – Test: is je organisatie klaar voor big data?

219 Noten

227 Register

241 Over de auteur

1

Waarom big data belangrijk zijn voor jou en je organisatie

Big data zijn onmiskenbaar big, maar de benaming is ook een beetje misleidend. Het is een vergaarbak voor data die niet in de normale hokjes passen. De term 'Big data' verwijst naar een hoeveelheid data die te groot is om op één enkele server te kunnen draaien en te ongestructureerd om in een database met rijen en kolommen te passen, of verwijst naar een zodanig continue toevoer van data dat ze niet in een statisch datawarehouse passen. Terwijl de omvang alle aandacht trekt, is in werkelijkheid het lastigste aspect van big data hun gebrek aan structuur.

Boeken als deze beginnen meestal met een verhaal over hoeveel data er wel niet in de wereld rondgaan. Je kent de cijfers en de vergelijkingen wel – het gemiddelde bedrijf heeft 427 keer de hoeveelheid data die zijn opgenomen in de Library of Congress. Facebook bevat meer fotografische data dan alle pixels die Kodak ooit heeft geproduceerd. We maken per dag meer video-opnamen dan er gemaakt zijn in de eerste vijftig jaar televisie. Dit zijn geen echte feiten over het duizelingwekkende karakter van de datavolumes en soorten data van vandaag de dag – ik heb ze uit mijn duim gezogen – maar ze zouden zomaar waar kunnen zijn.

Ik ben niet met zo'n soort opsomming begonnen, omdat ik vind dat die er niet echt toe doet. Zeker, er is een gigantische hoeveelheid data om ons heen. Volgens een onderzoek gebruikte de wereld in 2012 2,8 zettabyte

aan data – dat is 2,8 triljoen gigabyte, een onvoorstelbaar grote hoeveelheid.[1] Dat is inderdaad meer dan alles wat we ooit hebben gekend, en het zal in de loop van de tijd alleen nog maar toenemen. Het mag dan leuk zijn om op feestjes te vertellen hoeveel data we hebben, voor organisaties die big data moeten managen en er hun voordeel mee moeten doen, is die omvang het punt niet.

Het punt is niet dat je overdonderd raakt door het datavolume, maar veel meer dat je de data moet analyseren – om ze te kunnen omzetten in inzichten, innovaties en waarde voor je business. In het eerder aangehaalde onderzoek wordt gezegd dat slechts 0,5 procent van de 2,8 zettabyte aan data op enigerlei wijze wordt geanalyseerd. De grootste drempel bij die analyse is dat we de data eerst moeten structureren. Het grootste deel van de 2,8 zettabyte staat nog niet in een format met rijen en kolommen. Er wacht ons een gigantische taak: een begin maken met het structureren van de data, ze analyseren en de potentiële waarde ervan benutten. Niet alles zal bruikbaar zijn – in het onderzoek wordt geschat dat ongeveer 25 procent van potentiele waarde is – maar wat het cijfer ook mag zijn: op dit moment benutten we een mimimaal deel van de mogelijkheden.

Voorbij de big-datahype

Het is zaak om sceptisch te zijn over big data en de hype eromheen. Ik weet dat ik ook sceptisch was, in elk geval tot ik het onderwerp aan nader onderzoek onderwierp. Ik had veel samengewerkt met bedrijven rond het gebruik van business analytics en had verschillende boeken over het onderwerp geschreven – inclusief *Competing on Analytics* (met Jeanne Harris) en *Analytics at Work* (met Jeanne Harris en Bob Morison). Ik had met meer dan honderd bedrijven samengewerkt op het gebied van analyticsgebruik om concurrentievoordeel te behalen. Toegegeven, in eerste instantie dacht ik dat big data gewoon oude analyticswijn in nieuwe zakken was. De term kwam op in het vierde kwartaal van 2010, en er waren toen buiten Silicon Valley nog niet veel voorbeelden. Daardoor dacht ik dat het gewoon weer een hype was van analisten uit de hoek van de consultants, verkopers of technici. Ik overwoog

kort om mijn boeken over analytics van de plank te halen, de term ´analytics´ te vervangen door 'big data' en voilà: weer een paar nieuwe boeken erbij (grapje!).

Dat mijn scepsis onterecht was, ontdekte ik toen ik onderzoek naar het onderwerp begon te doen, in 2011. Ik voerde verschillende systematische onderzoeken uit, waaronder een onderzoek naar datawetenschappers en de menselijke factor in big data, een ander onderzoek over big data in grote concerns, nog een over big data in de reisbranche en een over het proces van *datadiscovery* bij big data.[2] Ik hield meer dan honderd interviews voor deze projecten bij startups op het gebied van big data, bij gevestigde online bedrijven en bij grote organisaties in traditionele sectoren. In grote, gevestigde bedrijven was het niet ongebruikelijk dat managers mijn scepsis en wantrouwen over de hype deelden. Ze hadden vaak het gevoel dat ze al jaren met big data te maken hadden gehad (zeker als je de grote omvang als belangrijkste kenmerk van big data ziet) en dat er niets nieuws onder de zon was. Maar in een latere fase van de interviews, gaven de meesten van hen toe dat het gebrek aan structuur van veel van de tegenwoordige data nieuwe uitdagingen en nieuwe business opportunity's oplevert.

Ik concludeerde uiteindelijk, op basis van mijn onderzoek, dat er reële verschillen zijn tussen conventionele analytics en big data – hoewel je dat niet altijd zou zeggen als je andere boeken en artikelen over het onderwerp leest, waarin de verschillen vaak vaag blijven (zie tabel 1.1 voor een samenvatting van deze verschillen). Ik zal de verschillen en de belangrijkste overeenkomsten verderop in dit hoofdstuk en in dit boek uitleggen. Dat neemt niet weg dat ik ondanks mijn reserves over de benaming zal betogen dat big data geen tijdelijk verschijnsel zijn en dat ze van substantieel belang zijn voor veel organisaties.

TABEL 1.1

Big data en traditionele analytics		
	Big data	Traditionele analytics
Soort data	Ongestructureerde formats	Ingedeeld in rijen en kolommen
Datavolume	100 terabyte tot petabyte	Maximaal enkele tientallen terabyte
Toevoer van data	Constante toevoer van data	Statische verzameling van data
Analysemethode	Machineleren	Op basis van hypothesen
Primair doel	Op data gebaseerde producten	Ondersteuning van diensten en interne beslissingen

Ik hoop dat ik je kan overtuigen en dat je mijn mening gaat delen, maar geef je scepsis op dit moment nog niet op. Als je vindt dat big data belangrijk zijn voor jou en je organisatie, dan moet je er immers wat mee. Je moet beslissen welke aspecten het zinvolst zijn om toe te passen in je business, en daarmee aan de slag gaan. Je moet de mensen aannemen, inhuren of opleiden die ervoor zorgen dat big data voor je werken. En je zou ook je technologische architectuur moeten veranderen. Nogmaals: voor mijn gevoel zou dit boek succesvol zijn als je met die dingen zou beginnen, maar alleen als ze je echt zinvol lijken – je zou hoe dan ook eerst overtuigd moeten zijn van de noodzaak.

Het doel van dit boek is niet om big data aan je te 'verkopen', maar om je te helpen effectieve beslissingen te nemen op dat vlak. Ik ga je vertellen waarvan ik onder de indruk was op big-datagebied, maar ik vertel je ook welke aspecten ervan wat al te zwaar worden aangezet. Ik vertel je welke bedrijfstakken en organisaties erdoor zullen transformeren, maar ook welke er waarschijnlijk niet veel invloed van zullen ondervinden, op zijn minst voorlopig niet. En als je besluit om verder te gaan met big data, laat ik je zien wat de verstandigste en zuinigste aanpak is.

En hoewel ik respectvol omga met je tijd en managementaandacht, hoop ik toch dat je openstaat voor het idee om je te mobiliseren rond een groot dataproject. Misschien zou dat slechts een verkennend project kunnen zijn om wat vaardigheden op te bouwen, maar ik denk wel dat de meeste

organisaties zich in die richting zouden moeten begeven. Je managementteam zou op zijn minst het onderwerp ter sprake moeten brengen hoe big data in je business passen.

Het is waarschijnlijk het best om te beginnen met mijn aannames over wie jij als lezer bent. Daarna zal ik een paar van mijn bedenkingen bij big data uiteenzetten die te maken hebben met de term en met de modieusheid waarmee het onderwerp omgeven is. En dan zal ik, in de rest van dit hoofdstuk, vertellen waarom big data zo'n big deal zijn.

Wie ben jij?

Wie ben je, dat je meer wilt weten over de implicaties en het nut van big data voor je business? Daar heb ik een paar aannames over (waarbij ik er even niet van uitga dat je dit boek misschien wel gewoon zag liggen op een vliegtuigstoel of dat je het verkeerde e-book downloadde). De lezers van mijn vorige boeken over analytics waren vooral intelligente, ambitieuze mensen uit het bedrijfsleven die geloven in het vermogen van data en technologie om de regels van het spel te veranderen. Ik vermoed dat jij dezelfde eigenschappen hebt.

Ik ben erachter gekomen dat leidinggevenden die het meest geïnteresseerd zijn in big data, werken in data-intensieve domeinen als marketing, logistiek en, in toenemende mate, finance en human resources. Leidinggevenden met functies die big data ondersteunen, bijvoorbeeld in de IT, willen vaak ook graag meer te weten komen over het concept. De meest geïnteresseerde leidinggevenden werken in sectoren die al heel sterk op data gericht zijn, zoals online bedrijven, of die het potentieel hebben om te worden omgevormd. Dat laatste geldt voor bijna alle takken van sport maar in het bijzonder voor bedrijfstakken met veel consumentendata, zoals retail, reizen en vervoer, telecommunicatie, media en entertainment, en financiële dienstverlening.

Tot slot: als je een student bent die zich voorbereidt op een carrière – of op zijn minst een baantje of twee – in de big data, dan mag je jezelf feliciteren met je scherpzinnigheid. Want big data gaan waarschijnlijk jarenlang boomen.

Als je bent zoals de meeste managers en professionals die ik tegenkom, ben je je zonder meer bewust van de term 'big data' en weet je dat de hoeveelheid data met een verbazingwekkende snelheid groeit. Maar je hebt misschien niet helemaal helder wat er nu precies anders is aan big data en hoe big data zich verhouden tot traditioneel datamanagement en analytics. Belangrijk: de meeste van jullie organisaties doen er nog niet veel mee. In een enquête uit 2013 onder bijna duizend lezers van *Harvard Business Review* bijvoorbeeld, zeiden veel respondenten dat ze bekend waren met het concept big data. Maar slechts 28 procent zei dat hun organisatie 'momenteel big data gebruikt om betere beslissingen over hun business te nemen of nieuwe business opportunity's te creëren'. Slechts 23 procent zei dat hun organisatie een strategie had voor big data. En een magere 6 procent was het 'zeer eens' met het statement 'mijn organisatie heeft nagedacht over de impact van big data op centrale functies van onze business'.

Daarom heb je waarschijnlijk je toevlucht genomen tot dit boek – om jouw organisatie te helpen om zulke vragen aan de orde te stellen, en misschien om je eigen carrière daarmee een duwtje in de rug te geven. Ik flatteer mezelf dan graag met de gedachte dat je precies naar de juiste plek bent gekomen.

Deconstructie van de term 'big data'

Net als veel leidinggevenden met wie ik dit verschijnsel heb besproken, vind ik alles aan het big-dataconcept geweldig – behalve de naam. Het concept is, zoals ik al aanstipte, revolutionair en draagt de mogelijkheid in zich om praktisch elke business te transformeren. De term zelf echter is problematisch, om uiteenlopende redenen.

Allereerst is er de kwestie dat *big* niet het enige aspect is dat kenmerkend is voor de nieuwe vormen van data, en voor veel organisaties is het niet de belangrijkste eigenschap. Een enquête uit 2012 van NewVantage Partners onder ruim vijftig executives in grote organisaties maakte duidelijk: voor grote bedrijven staat meer op de voorgrond hoe je het gebrek aan structuur van deze data aanpakt dan dat de hoeveelheid nu zo'n issue is. In deze

enquête zei 30 procent van de respondenten dat het 'analyseren van data uit diverse bronnen' het primaire issue op datagebied was waardoor ze gingen nadenken over big data; 22 procent was vooral gefocust op 'het analyseren van nieuwe soorten data' en 12 procent was gefocust op 'het analyseren van *streaming* data'. Slechts 28 procent was gericht op het analyseren van datasets van 1 terabyte of meer, en daarbinnen was een groot deel (13 procent van alle respondenten) georiënteerd op datasets van tussen de 1 en 100 terabyte – geen grote hoeveelheid in termen van big data.[3]

Er zijn nog een paar dingen aan de hand met de term 'big data'. Het woord 'big' is duidelijk relatief: wat vandaag nog gigantisch is, zal morgen al zo groot niet meer zijn. En zoals in de eerder aangehaalde enquête wordt gezegd: wat veel is voor de ene organisatie, is weinig voor de andere. Bij 'big' denk ik in de regel aan iets in de orde van een tiende petabyte of meer. Maar de enige reden dat omvang er inderdaad toe doet op datagebied, is de hoeveelheid hardware die je moet kopen om je data op te slaan en te verwerken.

Sommigen hebben 'big data' gedefinieerd in termen van de drie V's: *volume* (de hoeveelheid data), *variety* (diversiteit) en *velocity* (snelheid waarmee data binnenkomen en worden opgevraagd). Anderen hebben daar zelfs nog V's aan toegevoegd: *veracity* (waarheidsgetrouwheid) en *value* (waarde, nut) – en misschien is *venality* (corruptie) de volgende – maar ook dit is problematisch. Toegegeven: dit zijn belangrijke kenmerken. Maar wat als je slechts een of twee van de V's hebt? Betekent dat dan dat je maar voor een derde of twee vijfde aan big data hebt?

Een ander probleem is dat te veel mensen – vooral leveranciers (verkopers) – de term 'big data' al hanteren in de betekenis 'gebruik van analytics', of in extreme gevallen zelfs in de betekenis van rapportage en conventionele *business intelligence*. Het is een gevestigd gebruik dat verkopers en consultants elke nieuwe, trendy term gaan toepassen op hun bestaande aanbod – iets wat onmiskenbaar ook is gebeurd met 'big data'. Wees op je hoede als je boeken, artikelen of advertenties over big data leest: als ze gaan over rapportage, datagestuurde besluitvorming of conventionele analytics neem je weliswaar ideeën tot je die nuttig en waardevol zijn maar ze zijn niet speciaal nieuw.

Gezien deze definiëringsproblemen voorzie ik (net als andere experts die ik erover sprak) een relatief korte levensduur voor de ongelukkige benaming. De term werd omarmd door de media en startups. Maar ik heb al gezien dat op technologie georiënteerde medewerkers van grote bedrijven – vooral bedrijven die al vele jaren met grote hoeveelheden data werken, zoals banken en transportbedrijven – terughoudend zijn om de term over te nemen. Zoals ik verder uiteen zal zetten in hoofdstuk 8 vinden ze dat de huidige generatie van nieuwe databronnen en -soorten slechts een van de vele nieuwe generaties is die we al hebben zien opkomen. Dat betekent natuurlijk niet dat het 'verschijnsel dat voorheen bekend stond als big data' zal verdwijnen. Als je het brede areaal van de grote hoeveelheden nieuwe data wilt beschrijven die het laatste decennium het licht hebben gezien, dan weet ik geen betere verzamelterm dan 'big data'.

Dat neemt niet weg dat organisaties de term 'big data', omdat die zo onnauwkeurig is, een beetje moeten ontleden om hun strategieën te verfijnen en om aan stakeholders duidelijk te maken wat ze precies interessant vinden om te doen met deze nieuwe soorten data en welke datasoorten het belangrijkst zijn. Er zijn uiteraard veel soorten big data waaruit je kunt kiezen – er is zelfs een ware menukaart aan mogelijkheden, zoals tabel 1.2 laat zien. Je zou kunnen beginnen met één datasoort te kiezen.

TABEL 1.2

Het menu van big-datamogelijkheden			
Datastijl	**Bron van de data**	**Betrokken bedrijfstak**	**Betrokken functie**
Groot volume	Online	Financiële dienstverlening	Marketing
Ongestructureerd	Video	Gezondheidszorg	Supply chain
Continue toevoer	Sensor	Fabricage	Human resources
Diverse formats	Genomisch	Reizen/vervoer	Finance

Met andere woorden: in plaats van dat je zegt 'We gaan aan de slag met een big-data-initiatief' is het waarschijnlijk constructiever om te zeggen: 'We gaan de videodata van onze geldautomaten en filialen analyseren om meer te weten te komen over klantrelaties.' En als je actief bent in de gezondheidszorg, zou je kunnen besluiten om 'elektronische patiëntendossiers en genomische data te combineren om gepersonaliseerde behandelingen te ontwikkelen voor patiënten.' Dit verschaft niet alleen helderheid over je bedoelingen en strategieën, het voorkomt ook eindeloze discussies over of de betrokken data al dan niet 'big' zijn. (In feite geven maar weinig organisaties toe dat ze werken met 'small data', ook al is het een heel respectabele activiteit – een term is pas echt bruikbaar als het tegengestelde ook steekhoudend is, heb ik geleerd.)

Natuurlijk kun je zo'n trendy benaming ook in je voordeel gebruiken. Als jij in een organisatie werkt die zich graag bovenop fonkelnieuwe managementobjecten stort – en als big data nog steeds relatief nieuw zijn tegen de tijd dat je dit leest – start dan gewoon wel een 'big-dataproject' of BDP. Ofwel: als de term 'big data' enthousiasme en energie vrijmaakt in jouw bedrijf, ga er dan voor. Maar: wees er klaar voor om te 'pivoteren' (over trendy termen gesproken!) naar de volgende hippe benaming voor diezelfde activiteiten zodra die komt bovendrijven. IBM heeft al een onderzoekscentrum voor 'massive data' – en termen als 'gargantuesk' en 'kolossaal' laten vast niet lang op zich wachten!

In dit boek blijf ik 'big data' gebruiken, omdat er momenteel niets anders voorhanden is dat de verzameling karakteristieken typeert die hier aan de orde is. Maar ik ben ervan overtuigd dat je de echte waarde van deze bron-met-de-slordige-naam pas blootlegt als je er een paar niveaus dieper induikt.

Zijn big data blijvend?

We zijn het er nu misschien over eens dat de naam 'big data' ongelukkig is. Het volgende issue is: zijn big data iets blijvends of niet meer dan een modegril? Zijn ze de hoelahoep of Tamagotchi van de informatiemanagementwereld? Als dat zo is, kan het management van organisaties big data zonder

kopzorgen negeren. Maar als big data een langetermijnonderdeel zijn van de bedrijfsomgeving, dan is het gevaarlijk voor organisaties en managers om ze te negeren.

Het big-data-idee heeft onmiskenbaar modieuze trekjes, vooral als je bedenkt dat het bredere concept erachter niet speciaal nieuw is. Het idee om data te analyseren met als doel wijs te worden uit wat er in je business gebeurt, kennen we al een tijdje (in bedrijven op zijn minst sinds 1954, toen UPS een analyticsafdeling startte). Waarom moeten we dan met nieuwe namen komen om dit te beschrijven? De algemene activiteit van het wijs worden uit data heeft al vele namen gekregen: *decision support*, *executive support*, *online analytical processing (OLAP)*, *business intelligence*, *analytics* en nu *big data* (zie tabel 1.3).[4] Elke generatie terminologie heeft zeker enige nieuwe elementen, maar ik betwijfel of de zaken zodanig zijn geëvolueerd dat dit zes generaties van benamingen rechtvaardigt.

TABEL 1.3

Terminologie voor het gebruiken en analyseren van data		
Term	**Tijdsbestek**	**Specifieke betekenis**
Decision support	1970 - 1985	Gebruik van data-analyse om de besluitvorming te ondersteunen
Executive support	1980 - 1990	Focus op data-analyse voor beslissingen van senior executives
Online analytical processing (OLAP)	1990 - 2000	Software voor het analyseren van meerdimensionale datatabellen
Business intelligence	1989 - 2005	Tools ter ondersteuning van datagestuurde besluitvorming, met de nadruk op rapportage
Analytics	2005 - 2010	Focus op statistische en mathematische analyse ten behoeve van de besluitvorming
Big data	2010 - nu	Focus op zeer grote hoeveelheden ongestructureerde, fast moving data

De reden dat de nieuwe naam 'big data' tot op zekere hoogte wel terecht is, is dat er nieuwe en omvangrijkere datasoorten bij betrokken zijn – volgens één schatting gaat het om 2,5 triljard (2,5 gevolgd door 18 nullen)

gegenereerde bytes per dag wereldwijd.[5] Minder gestructureerde datasoorten zijn, zoals ik al aangaf, die nieuwe benaming en aanpak nog meer waard. Sommige datasoorten, zoals tekst en stemgeluid, kennen we al een hele tijd – maar hun omvang op internet en andere digitale kanalen kondigt een nieuw tijdperk aan, net als de nieuwe technologieën die analyse van zulke data mogelijk maken. Echt nieuw zijn socialmediadata, van Facebook- en Pinterest-pagina's tot tweets. Ik weet niet of al deze datasoorten overeind zullen blijven. Het gebruikelijke patroon voor de functies die deze tools vervullen, is dat ze worden geïncorporeerd in bredere applicaties. Maar de data die ze genereren en de inzichten die ze blootleggen over hun auteurs … die verdwijnen niet.

In algemene zin kun je echter wel stellen dat *sensordata* een blijvend verschijnsel is. Het aantal 'genetwerkte' apparaten overtrof in 2011 het aantal mensen op aarde. Analisten schatten dat er in 2025 50 miljard sensoren verbonden zijn met internet (*the internet of things*), die stuk voor stuk een heleboel data kunnen voortbrengen. Hoewel vroege voorspellers dachten dat met internet verbonden sensoren voornamelijk zouden worden gebruikt in apparaten voor consumenten, is er in dat opzicht slechts beperkte vooruitgang geboekt. onze koelkasten mogen dan nog niet op afzienbare termijn verbonden zijn met internet (dan kunnen ze bijvoorbeeld automatisch een nieuw pak melk laten afleveren als de melk bijna op is), maar onze tv's, beveiligingssystemen en thermostaten zijn in toenemende mate genetwerkt. Elk van die apparaten genereert data die kunnen worden benut om het gebruik, de consumptie en levering van diensten te optimaliseren.

Sensoren worden ook bevestigd op bezielde objecten als koeien en mensen. Roger Parks, chief information officer van landbouw- en veebedrijf F.R. Simplot, beschrijft iets wat hij een ´digitale koe´ noemt. Het bedrijf experimenteert met sensoren in koeienmagen om de temperatuur te meten (koeien hebben vier magen, en om een of andere reden lijkt de tweede maag de optimale locatie). Als een koe ziek is, laat de sensor de veearts weten dat er een probleem is op een moment dat de ziekte nog kan worden behandeld. Andere onderzoekers experimenteren met sensoren die de aanwezigheid van E. colibacteriën opsporen in koeienmagen. Laten we hopen dat er in koeienmagen nog plaats is voor voer met al die sensoren!

Ook menselijke wezens worden in toenemende mate 'gesensord' (een woord dat ik net heb verzonnen), in de regel om gezondheidsredenen of voor fitnessdoeleinden. We leven in het tijdperk van *zelfanalyse*, van het verzamelen en analyseren van data over je persoonlijke fitheid, productiviteit en gezondheid.[6] De vroegste toepassing van persoonlijke analytics die ik ken, was de Nike+/iPod-combinatie die het licht zag in 2006. De Nike+-schoen die met een iPod verbonden was registreerde de tijd die je gerend had, de afstand, de snelheid en de verbruikte calorieën en liet dat allemaal zien op de display van de iPod. Nike en Apple hebben snel allerlei andere manieren ontwikkeld om data over sportieve inspanningen te verzamelen (hartslagsensoren, links naar cardiofitnessapparatuur, kleding met ingebouwde sensoren enzovoort) en de vele Nike+-gebruikers (naar verluidt meer dan 5 miljoen) uploaden hun gegevens vaak naar hun laptop of de Nike+-website. Je kunt je loopprestaties monitoren, vrienden uitdagen en online coaching krijgen voor trainingsprogramma's.

Sinds 2006 zijn allerlei sensorprogramma's voor het vastleggen van persoonlijke data als paddenstoelen uit de grond geschoten. Net als bij Nike+ draait het vaak om sport. Garmin Connect bijvoorbeeld, stelt verschillende soorten atleten in staat om een logboek bij te houden van hun activiteiten, nieuwe routes te plannen en hun prestaties met anderen te delen. Zed9 maakt een trackrecord van je 'sociale fitheid', CycleOps van je fietskracht, Concept2 van je roeiworkouts.

Maar persoonlijke analytics reiken in toenemende mate verder dan sport alleen, en beslaan ook domeinen als gezondheid, welstand, werk en tevredenheid over je leven in het algemeen. Withings introduceerde de met wifi en Twitter verbonden weegschaal en lijkt zich nu richting bloeddrukmeting en andere apparaten op gezondheidsterrein te bewegen. Allerlei leveranciers – waaronder MyZeo, WakeMate, BodyMedia en Fitbit – bieden slaapanalytics aan (mijn Fitbit informeerde me onlangs dat ik een slaapefficiency van 97% had – waar ik erg blij van werd!). Me-trics, een startup uit Michigan, biedt nu een algemene tool aan voor het meten en monitoren van persoonlijke analytics in zo ongeveer elke levenssfeer, met inbegrip van gezondheid, stemming, financiën, fitheid en online activiteit. De dag dat ik inlogde op de website van Me-trics, was Mary erotische ontmoetingen aan het

volgen (nu heb ik je aandacht!) en Ryan waterconsumptie. Er zullen zonder twijfel sensoren op maat worden ontwikkeld voor dit soort meetgegevens.

Alsof dat nog niet genoeg data zijn over menselijke activiteiten: we hebben onze smartphones bijna altijd bij de hand. Ze registreren onze locaties, conversaties en, steeds vaker, onze aankopen. Computers kunnen zo ongeveer elk aspect van ons werkleven bijhouden, zoals Stephen Wolfram (de CEO van Wolfram Research) heeft laten zien voor zijn eigen werk.[7] Wolfram weet bijvoorbeeld precies wanneer elk van de ruim 300.000 e-mails de deur uitging die hij sinds 1989 verstuurde. Sensoren volgen ook onze huisdieren. Misschien worden ook onze stemmingen en hersengolven binnen afzienbare tijd opgevangen en geanalyseerd. De bulk van sensordata zal echter uiteindelijk afkomstig zijn van het 'industriële internet' – een zeer grote hoeveelheid van genetwerkte apparaten in fabrieken, transportnetwerken, energienetten enzovoort. GE beschouwt deze ontwikkeling als in potentie revolutionair in termen van datavolumes en van mogelijke voordelen. Het monitoren van de schoepen in energieproducerende gasturbines alleen al kan 588 gigabyte aan data per dag opleveren – zeven keer het dagelijkse aantal Twitterdata.

Mochten deze omvangrijke bronnen van big data nog niet genoeg zijn om mensen te motiveren om big data te omarmen, dan zijn er nog softwareleveranciers die ze een duw in de rug geven – of misschien worden zij daar wel toe aangezet door hun klanten. De softwareleveranciers mogen dan soms oude wijn in nieuwe big-datazakken schenken, over het algemeen is er een ommekeer te zien in de aard van bedrijfssoftware. We zijn aan het bewegen van het automatiseren van transacties naar het analyseren van de data die ze voortbrengen. Als SAP meer geld verdient aan business intelligence en analytics dan aan zijn transactionele applicatiesuite, dan heeft er zich een grote verandering voltrokken. Als bedrijven als HP, EMC en Oracle grote acquisities doen en producten aankondigen in de hoek van big data en analytics, dan is er iets nieuws aan het gebeuren. Als IBM rond de 20 miljard dollar heeft gestoken in aan analytics gerelateerde acquisities, dan is de situatie definitief veranderd. Zelfs Microsoft – het bedrijf dat werd opgericht om small data te ondersteunen op personal computers – heeft diverse producten aangekondigd die te maken hebben met big data. Ik vertel verderop

in dit boek meer over de technologie van big data (vooral in hoofdstuk 5). Op dit moment hoop ik dat we het erover eens kunnen zijn dat het concept 'big data' waarschijnlijk niet gaat verdwijnen als zoveel grote bedrijven zich erachter scharen.

Meer data, meer technologie – wat is er nog meer nodig om zeker te weten dat big data geen gril zijn? Zoals ik in dit boek zal betogen, zijn het *mensen* die er daadwerkelijk voor zorgen dat big data werken. De rol van *datawetenschapper*, die ik uitgebreider bespreek in hoofdstuk 4, is in mijn ogen de primaire sleutel tot het welslagen van big data in een organisatie. De data zijn vaak gratis of goedkoop, de hardware en software zijn gratis of niet duur – maar de mensen zijn duur en moeilijk te vinden. Ik vertel in hoofdstuk 4 veel meer over waarom dit zo'n lastig probleem is – mijn punt is hier dat het veel gemakkelijker gaat worden. Veel universiteiten bieden masters of specialisaties aan op het gebied van business intelligence of analytics, en een flink aantal voegt big-datakwesties en -vaardigheden toe aan het curriculum. Op scholen worden momenteel verschillende big-data- en datawetenschapsprogramma's ontwikkeld. Het zal niet lang meer duren voor universiteiten tienduizenden gekwalificeerde studenten gaan afleveren. Dat zal het veel makkelijker maken om big-dataprojecten op te zetten in bedrijven en het voorkomt dat de big-databeweging in de kiem wordt gesmoord door een tekort aan gekwalificeerde arbeidskrachten.

Al deze factoren aan de aanbodkant wijzen erop dat big data en de ermee samenhangende concepten en technologieën bepaald geen trendverschijnsel zijn, maar ons vele tientallen jaren zullen vergezellen. Als organisaties deze ideeën willen negeren, kunnen ze net zo goed meteen ophouden zich te bekommeren om hoe ze geld kunnen besparen, hoe ze meer producten kunnen verkopen en hoe ze klanten blij kunnen maken – en dat zullen ze waarschijnlijk niet zo gauw doen.

Wat is er nieuw vanuit managementperspectief?

Ik zet in hoofdstuk 5 uiteen wat er nieuw is aan de big-datatechnologie. Zoals wel vaker het geval is, kan technologie je weliswaar voor uitdagingen stellen,

maar is toch de management- en mensenkant het ingewikkeldst. Sommige kwesties op die terreinen zijn zo nieuw dat er nog nauwelijks antwoorden voor zijn. Bij andere zie je inmiddels de eerste oplossingen doorschemeren.

Ik heb een van deze kwesties al aangestipt, namelijk dat je de juiste mensen moet binnenhalen om je big-datawerk te doen. Ik kan niet genoeg benadrukken dat mensen de sleutel zijn in dit werk. Hoewel het niet volledig anders of nieuw is dat je mensen nodig hebt voor analytische activiteiten, verschilt het soort mensen dat nodig is voor big-data-initiatieven – datawetenschappers – zonder meer van conventionele analisten. Ze zijn vaardiger met data, ze zijn experimenteler en ze zijn meer gefocust op producten. In hoofdstuk 4 vertel ik daar veel meer over.

Aangezien bij big data zowel de data, de technologie als de mensen anders zijn dan bij traditionele analytics, is er behoefte aan nieuwe organisatorische structuren om big data een plaats te geven. Je kunt niet zomaar aannemen dat big data een zaak zijn voor de IT-afdeling. In grote organisaties kun je big-datagroepen aantreffen op terreinen als marketing, finance, productontwikkeling, strategie en IT. In hoofdstuk 3 ga ik dieper in op waar big data het best passen bij verschillende omstandigheden.

Traditioneel informatiemanagement en traditionele analytics hadden primair betrekking op het ondersteunen van interne besluitvorming. Wat dat betreft zijn big data anders. Ja, er zijn veel gevallen waarin ze ook voor dat doel worden gebruikt, vooral in grote organisaties. Maar datawetenschappers werken normaliter aan producten en diensten voor consumenten in plaats van dat ze rapporten of presentaties maken om senior executives te ondersteunen bij hun interne beslissingen. Dat geldt vooral voor big-datastartups, maar je ziet het ook in grote, gevestigde bedrijven. Een voorbeeld. Reid Hoffman, medeoprichter en rvb-voorzitter van LinkedIn, vormde met zijn datawetenschappers een productteam voor het bedrijf – ze ontwikkelden producten als People You May Know, Groups You May Like, Jobs You May Be Interested In en Who's Viewed My Profile. GE focust vooral op big data om zijn dienstverlening te verbeteren en gebruikt nu al datawetenschap om servicecontracten en onderhoudscontracten voor industriële producten te optimaliseren. En dan is er natuurlijk Google – het ultieme big-dataconcern – dat datawetenschappers inzet om zijn centrale

zoekfunctie en de algoritmen voor de advertentiedienst te verfijnen. Zynga gebruikt datawetenschappers om games en daarmee samenhangende producten af te stemmen op klanten. Netflix riep de bekende Netflix Prize in het leven voor het datawetenschapsteam dat de filmaanbevelingen van het bedrijf voor klanten kon verbeteren. Het testbedrijf Kaplan gebruikt zijn datawetenschappers om klanten aan het begin van het proces te adviseren over effectief leren en testvoorbereidingsstrategieën. De inspanningen van deze bedrijven op het gebied van big data zijn direct gericht op producten, diensten en klanten. Dit heeft uiteraard belangrijke gevolgen voor de plaats waar big data organisatorisch worden ondergebracht en op de processen en snelheid van productontwikkeling.

Voor big-datagebruik waarbij wel interne besluitvorming aan de orde is zijn óók andere managementmethoden nodig, al zijn die in de praktijk nog niet volledig opgehelderd. Het punt is dat big data gewoon blijven binnenstromen. In traditionele situaties waarin data de besluitvorming ondersteunen, neemt een analist een verzameling data, analyseert die, komt met een model en adviseert de beslisser op grond van de resultaten. Maar bij big data gaat het niet om een vastomlijnde set data, maar om een voortdurende, snelle toevloed van data. Daarom is een meer continue benadering nodig van het verzamelen, analyseren en handelen op basis van data. Dat geldt in het bijzonder voor toepassingen waarbij data non-stop worden gemonitord, zoals bij analyse van sentimenten in social media. *Sentimentanalyse* maakt het voor organisaties duidelijk of commentaren op hun merken en producten in blogs, tweets en op Facebook-pagina's gemiddeld genomen positief of negatief zijn. Een potentieel probleem bij zulke monitortoepassingen is dat managers de neiging hebben om een continue stroom van analyses en rapporten te bekijken zonder tot actie over te gaan of beslissingen te nemen. 'Het sentiment schiet omhoog .... nee, het zakt ... hoera, het stijgt weer!' Voor dit continue monitorwerk zouden er processen moeten zijn om te bepalen welke specifieke beslissingen en acties nodig zijn – bijvoorbeeld als datawaarden buiten bepaalde grenzen vallen. Dergelijke informatie helpt te bepalen wie de stakeholders zijn bij beslissingen, wat de besluitvormingsprocessen zijn en wat de criteria en tijdskaders zijn waarbinnen beslissingen moeten worden genomen.

Zelfs de Verenigde Naties – een organisatie die niet bepaald bekendstaat om zijn wendbaarheid – is bezig met deze nieuwe manier van beslissingen nemen. Het Global Pulse-innovatielab van de VN heeft onder de naam HunchWorks een aan big data gerelateerde tool ontwikkeld, die duidelijk een op monitoring georiënteerde toepassing van big data is. Het lab omschrijft HunchWorks als 'het eerst sociale netwerk ter wereld voor hypothesevorming, het vergaren van bewijsmateriaal en collectieve besluitvorming'.[8] Het idee is dat de data een patroon of vondst onthullen. Stel dat meteorologische gegevens op een droogte wijzen die tot honger in een deel van Afrika zou kunnen leiden. Een analist zou dan een *hunch* (voorgevoel) posten met daarbij de data waarop die hunch gebaseerd is. Zulke suggestieve hypothesen zijn al aangeduid als 'digitale rooksignalen'.[9] Het doel van die rooksignalen is: bepalen hoe waarschijnlijk het is dat de hunch gedetailleerde analyse verdient en bepalen of er actie op moet worden ondernomen. Het idee dat de VN beschikt over een systeem voor het verspreiden van door data aangestuurde voorgevoelens, markeert een omslag in de cultuur van de organisatie.

Of de analyse- en besluitvormingsprocessen nu sociaal of individueel zijn, de continue big-datastroom dwingt organisaties om na te denken over nieuwe manieren van beslissingen nemen. Als het de moeite loont om te investeren in het vergaren en analyseren van big data, is het ook zinvol om na te denken over de manier waarop de uitkomsten van de analyse impact zullen hebben op beslissingen en handelingen.

Big data en nieuwe managementoriëntaties

Big data veranderen niet alleen technologie- en managementprocessen maar ook de primaire oriëntatie en cultuur van organisaties. We kunnen gewoon niet meer op dezelfde manier over business denken nu deze nieuwe resource er is. Eén noodzakelijke verandering is dat de focus moet verschuiven in de richting van *discovery* ('ontdekking') en van het experimenteren met data.[10] Tot nu toe waren bedrijven en technologieorganisaties vooral gefocust op het duurzaam automatiseren van processen als marketing, ver-

koop en service. Analytics werd gebruikt om meer inzicht in de processen te krijgen en ze af te stellen, om het management geïnformeerd te houden en te waarschuwen bij afwijkingen (het rapporteren van uitzonderingen was een centraal punt in de business intelligence). De business- en de technologiearchitectuur zijn vaak een weerspiegeling van dit proces, dat begint met transacties en operaties en dan – mag je hopen – leidt naar analyse en inzicht. Bedrijven evalueren hun performance, plannen verbetering en implementeren die in technologieprojecten die vele jaren beslaan.

Big data zetten die aanpak op zijn kop. Het grondbeginsel is dat de wereld – en de data die de wereld beschrijven – voortdurend verandert en in beweging is, en dat organisaties die dat herkennen en snel en intelligent reageren een voorsprong hebben. De winnende bedrijfsmatige en IT-vaardigheden zijn eerder *discovery* en *agility* (wendbaarheid) dan stabiliteit. Datawetenschappers die met big-datatools en -technologieën werken, zijn in staat om continu nieuwe en bestaande databronnen te ontginnen op patronen, gebeurtenissen en kansen – op ongekende schaal en met nog niet eerder vertoonde snelheid. Die nieuwe focus is relatief simpel voor startups, en ik kwam hem inderdaad vaak tegen bij de startups die ik onderzocht heb en die ik beschrijf in hoofdstuk 7. Maar in grote organisaties vergt die focus een forse omslag in gedrag en activiteiten op IT-gebied. Kenmerkend is dat 75 tot 80 procent van de tijd van analisten gaat zitten in het 'sourcen', opschonen en voorbereiden van data voor analyse. Dat is het geval in zowel bedrijven als overheidsorganisaties. Carey Schwartz, hoofd van het strategisch initiatief 'Data to Decisions' bij het Pentagon, denkt dat de datavolumes van volgende-generatiesensoren en de complexiteit van geïntegreerde systemen straks de vermogens van analisten ver te boven zullen gaan, als er geen ingrijpende verbeteringen plaatsvinden.[11]

Het Pentagon en Schwartz maken zich terecht zorgen over deze kwestie. Net als veel ondernemingen in de private sector is het militaire apparaat bezig om sneller data te verzamelen dan men kan analyseren. Neem de drones die de laatste tijd in zwang zijn bij het Amerikaanse leger. Die vuren niet alleen raketten af op terroristen maar leggen ook videodata vast van het gebied waar ze overheen vliegen. Die videodata zouden heel nuttig kunnen zijn voor het leger, maar alleen als ze geanalyseerd zijn. En blijkbaar zijn er

niet genoeg menselijke analisten om al die videodata te verwerken. In 2012 klaagde minister van luchtmacht Michael Donley dat het de luchtmacht jaren gaat kosten om alle videodata te analyseren die de drones vastleggen. Naar verluidt is de luchtmacht nu bezig om deze data te analyseren met minder menselijke tussenkomst.[12]

Bij het Pentagon onderzoekt Schwartz big-datatechnologie met als doel analisten (van video- en andere data) honderden malen productiever te maken. Hij benadrukt de noodzaak van 'analytics die werken', die hij omschrijft als analytics die betrouwbaar en krachtig zijn en die kunnen worden geautomatiseerd. Analytics, algoritmen en gebruikersinterfaces moeten worden verbonden om te voorzien in nieuwe methoden voor interactie met en ondersteuning van de 'mens in de lus'. Misschien is die druk op automatisering deels ontstaan doordat op zijn minst één 'mens in de lus', Edward Snowden, data uit de militaire en intelligence-machinerie gebruikte om 'het duurste lek van Amerikaanse geheimen in de geschiedenis van de Republiek' te worden, in de woorden van voormalig directeur van de CIA en National Security Agency Michael Hayden.[13]

Eén element van de nieuwe architectuur voor big data is dat discovery en analyse worden beschouwd als de belangrijkste agendapunten. Datawetenschappers (en algemene analisten in bedrijven) hebben continue toegang nodig tot een analyticsplatform of *sandbox* (een term die de discovery-achtige missie van big data aangeeft) waardoor snel inzicht in bedrijfsdata en externe data wordt ondersteund. Dit platform moet integratie van nieuwe data, ad-hocvragen (*query's*) en visualisatie faciliteren om menselijk inzicht sneller mogelijk te maken. De waardevolle inzichten die boven komen drijven via dit platform, worden de grondstof voor veranderingen aan productiesystemen en -processen.

Bedrijven moeten ook nieuwe methoden adopteren voor het verkrijgen van inzicht en voor op data gebaseerde productontwikkeling. Traditionele watervalmethoden – vergaand gestructureerde aanpakken die pas een resultaat opleveren aan het eind van een lang proces – worden in systeemontwikkelingsprocessen steeds meer verdrongen door snellere en flexibelere *agile* processen. De agile aanpak, waarbij relatief weinig tijd wordt besteed aan het vooraf specificeren van een systeem en meer aandacht uitgaat

naar het snel opleveren van kleinere resultaten, kan ook worden toegepast op analytics en big data. Het onnauwkeurig en langzaam vergaren van de vereisten voor een nieuw analytisch systeem of proces wordt dan vervangen door iteratief experimenteren, inzicht verkrijgen en bekrachtigen.

Natuurlijk is niet alle IT discovery: hoe nuttig een inzicht precies is, wordt meestal pas duidelijk als het wordt toegepast in productieprocessen en -systemen. Als eenmaal een inzicht uit de data is gedestilleerd, moet dat worden geclassificeerd – bijvoorbeeld als 'irrelevant voor de business', 'interessant maar niet bruikbaar' of 'basis waarop actie moet worden ondernomen'. Iets dergelijks geldt als de analyse van big data heeft geleid tot een nieuw product of productkenmerk: dat idee moet dan worden overgenomen (of niet).

Het eerder besproken HunchWorks-voorbeeld van de VN illustreert een aspect van big data dat noopt tot verandering van managementoriëntaties. HunchWorks wordt wel beschreven als 'een mechanisme om de membranen tussen kennissilo's binnen en buiten de VN poreuzer te maken'. Een belangrijk kenmerk van big data is dat ze meestal afkomstig zijn van buiten de organisatie die ze gebruikt. Of je nu internetdata gebruikt, data over het menselijk genoom, socialmediadata of data van het internet der dingen, ze zijn waarschijnlijk niet afkomstig van de interne transactiesystemen van je bedrijf. Uitzonderingen op deze tendens – die ik beschrijf in hoofdstuk 2 – kom je het vaakst tegen in de telecommunicatiesector en de financiele dienstverlening, die gezegend zijn met gigantische hoeveelheden intern gegenereerde data die ze kunnen analyseren. Zelfs daar is het echter vaak nuttig om interne data aan te vullen met externe data.

Dit markeert een grote omslag in de oriëntatie van bedrijven. Peter Drucker zei in 1998 dat de meeste informatiesystemen gefocust zijn op interne accountingdata en hebben 'verergerd wat altijd al de degeneratieve tendens was in het management, zeker in grote ondernemingen: om naar binnen gericht op de kosten en eigen inspanningen te focussen in plaats van op kansen, veranderingen en bedreigingen … Hoe meer interne informatie het topmanagement krijgt, hoe meer het die in evenwicht zal moeten brengen met externe informatie – en zoiets bestaat nog niet.'[14]

Nu, met big data, begint die externe focus wel te bestaan. Zoals een leidinggevende van een startup het uitdrukte (Christopher Ahlberg, CEO van Recorded Future): "Het lijkt erop dat we al het sap uit de interne informatie hebben geperst. Misschien is het tijd dat we ons eens op de hele wereld aan externe informatie richten."[15] Ik denk zelf dat er echt nog wel wat sap uit interne bronnen te persen valt, maar er is absoluut meer informatie te verkrijgen buiten organisaties dan binnen organisaties.

Om externe data te kunnen verwerken in beslissingen, producten en diensten, zullen veel managers hun mindset en gewoonten moeten veranderen. Ze zullen geregeld externe databronnen moeten onderzoeken om te zien wat er beschikbaar is en hoe die informatie hun organisatie zou kunnen ondersteunen. Ze moeten eenvoudigweg hun blik naar buiten richten, naar hun leveranciers en de leveranciers van hun leveranciers, naar hun klanten en de klanten van hun klanten, en naar zakelijke en politieke risico's. En zoals ik al zei: ze moeten systematische criteria ontwikkelen waarmee ze beslissingen kunnen nemen en actie kunnen ondernemen als ze de externe informatie belangrijk vinden.

Big data en nieuwe kansen

Om substantieel aan te slaan in het bedrijfsleven moeten big data uiteraard nieuwe kansen bieden. Blijven doorpraten over hoeveel data je op Facebook of Twitter vindt of over hoeveel gigabyte één enkel menselijk genoom heeft, helpt leidinggevenden niet om te bepalen wat voor hen het nut is van het exploiteren van big data. Big data kunnen op drie manieren nuttig zijn. Ze kunnen leiden tot: 1) kostenreductie, 2) betere beslissingen, of 3) verbeteringen op het gebied van producten en diensten. De kansen die big data bieden voor kostenreductie bespreek ik verderop in dit boek; hier volstaat de opmerking dat die kansen zeer substantieel zijn. Op het gebied van de besluitvorming is het primaire nut van big data dat er nieuwe databronnen worden toegevoegd aan verkennende en voorspellende modellen. Veel big-datafans betogen dat het waardevoller is om nieuwe databronnen toe te voegen aan een model dan om het model zelf te verfijnen. Neem Anand

Rajaram, werkzaam bij @WalMartLabs en docent aan de Stanford University. Hij deed een soort live-experiment in een van zijn Stanfordcolleges in het kader van de Netflix Prize, de wedstrijd die iedereen die er zin in had uitnodigde om het algoritme voor videovoorkeuren van klanten te verbeteren – en waarmee je een miljoen dollar kon winnen.[16] Een van de collegegroepen van Rajaram gebruikte de data van Netflix en paste daarop zeer geavanceerde algoritmen toe. Een andere groep vulde de data aan (illegaal, gelet op het wedstrijdreglement) met data over filmgenres uit de Internet Movie Database. De voorspellingen van de laatste groep waren veel succesvoller. Rajaram schrijft veel van Googles succes ten opzichte van eerdere zoekmachines toe aan het gebruik van aanvullende hyperlinkdata. Zoals Googles directeur onderzoek Peter Norvig zegt: "We hebben geen betere algoritmen. We hebben gewoon meer data."[17]

Ook in meer standaardorganisaties kunnen allerlei beslissingen worden verbeterd door big data toe te voegen. Als je over data beschikt die voorspellen wat klanten wel of niet aantrekt op grond van wat ze bij je hebben gekocht, kun je die waarschijnlijk verbeteren door data toe te voegen over hun transactiegeschiedenis. Als je een model hebt dat voorspelt welk volgende item afzonderlijke klanten waarschijnlijk zullen kopen op basis van hun aankoopgeschiedenis en demografie, kun je dat waarschijnlijk verbeteren door hun commentaren en *likes* op social media te analyseren. Een deel van de aanvullende data die je gebruikt, zullen 'big' zijn in die zin dat het om een grote hoeveelheid gaat of om ongestructureerde data, maar sommige van die data zullen kleinschalig en gestructureerd zijn. Het gaat erom dat je verder om je heen kijkt naar nieuwe databronnen die je kunnen helpen om beslissingen te nemen.

De andere grote kans die big data bieden, is producten en diensten te creëren die klanten aanspreken. In wat ik hiervoor schreef over conventionele business intelligence en analytics stipte ik al aan dat ik niet vaak tegenkom dat deze kans wordt benut. We staan nog maar aan het begin van big data in algemene zin en van op big data gebaseerde producten en diensten in het bijzonder, maar er zijn tal van voorbeelden van aantrekkelijke producten en diensten die zijn voortgekomen uit big data. Ik zal er meer van bespreken in hoofdstuk 2, maar hier volgen alvast een paar voorbeelden om een idee te geven van de mogelijke voordelen.

Ik heb al de rol van big data en op data gebaseerde producten bij LinkedIn genoemd, maar een specifiek product dat voor de onderneming absoluut nuttig is geweest, is 'People You May Know' (PYMK). Zoals veel lezers bekend zal zijn stelt PYMK aan LinkedInleden (klanten) andere leden voor die zij misschien in hun netwerk willen opnemen. PYMK gebruikt een multifactorbenadering bij het identificeren van mogelijke nieuwe connecties, inclusief gemeenschappelijke scholen, werkplekken, connecties en geografische locaties. Veel mensen, en daar behoor ik zelf ook toe, vinden het vermogen van PYMK om lang vergeten relaties op te sporen bijna griezelig.

Belangrijker is dat PYMK veel nieuwe klanten heeft gegenereerd voor LinkedIn. Vergeleken met de andere reminders die LinkedIn aan mensen stuurt om de site opnieuw te bezoeken, wordt op de PYMK-berichten 30 procent meer geklikt. Miljoenen mensen die dat normaal niet zouden hebben gedaan, kijken hierdoor opnieuw op LinkedIn. Door dit ene feature stegen LinkedIns groeicijfers significant. Een andere factor die getuigt van de waarde van het PYMK-feature is dat veel andere *socials* – waaronder Facebook, Twitter en Google+ – vergelijkbare mogelijkheden hebben toegevoegd, al denk ik dat de PYMK van LinkedIn nog steeds het effectiefst is.

Een ander voorbeeld. Amadeus, een bedrijf dat online, realtime computerreserveringssystemen aanbiedt voor onder meer de reis- en hotelbranche, heeft een big-dataservice onder de naam Featured Results. De onderneming zag zich geconfronteerd met een groeiende uitdaging: het snel toenemende aantal boekingsaanvragen online, ofwel het groeiend aantal zoekopdrachten per geboekt vliegticket. Amadeus moest voor reisdistributeurs een of andere manier vinden om klanten een aantrekkelijk aanbod te kunnen doen. Op basis van databases met gebruikersquery's, 'honderden miljoenen' realtime vliegticketprijzen en een half miljard reserveringen geeft Featured Results vier reisroutes aan waarin klanten speciaal geïnteresseerd zouden kunnen zijn. De eerste bètatestresultaten – bij Vayama, reisagent en partner van Amadeus – laten een verbetering van 16 procent zien in Vayama's zoek-verkoopratio.

Veel executives beamen dat big data in potentie substantiële toegevoegde waarde kunnen hebben voor online businesses, maar ze zijn minder

stellig over de waarde buiten dat domein. Wat ze zou kunnen overtuigen van de relevantie van big data voor henzelf, zijn de acties en plannen van GE – een van 's werelds grootste en succesvolste bedrijven en een van de meest enthousiaste gebruikers van big data – en dat nog wel in de industriële sector. GE heeft in het San Francisco Bay-gebied een instituut opgezet om software- en big-datakwesties aan te pakken en neemt voor dat doel heel veel datawetenschappers aan. Zij werken in van oudsher data-intensieve sectoren van GE, zoals financiële dienstverlening en gezondheidszorg. Maar hier wil ik het hebben over de potentiële toegevoegde waarde die GE ziet in industriële toepassingen, zoals in GE's businesses voor locomotieven, straalmotoren en gasturbines. GE typeert de apparaten in bedrijfstakken als deze ook wel als 'dingen die draaien' en verwacht dat de meeste van die dingen, zo niet alle, binnenkort gegevens over dat 'draaien' kunnen vastleggen en communiceren.

Een van die draaiende dingen is de gasturbine en die de klanten van GE gebruiken voor energieopwekking. GE monitort nu al meer dan 1500 turbines vanuit een centrale faciliteit, dus een groot deel van de infrastructuur voor gebruik van big data om de prestaties te verbeteren is er al. GE schat dat het de efficiëntie van de gemonitorde turbines met minstens 1 procent kan verbeteren via software- en netwerkoptimalisatie, doeltreffenderer afhandeling van onderhoud en betere harmonisering van het gas-energiesysteem. Dat lijkt misschien niet veel, maar het zou neerkomen op een brandstofbesparing van 66 miljard dollar in de komende vijftien jaar.

GE verdient tegenwoordig veel van zijn geld met het onderhouden van industriële producten. Dus als GE van de eigenaren data zou verkrijgen over hoe die producten presteren en wanneer ze kapotgaan, zou het bedrijf klanten kunnen beloven de kosten en effectiviteit van onderhoudswerkzaamheden te optimaliseren. Stel, je wilt een turbine kopen voor een krachtcentrale. Je kunt kiezen uit twee turbines. Er is er eentje met big-data-output, continue monitoring van prestaties en geoptimaliseerd onderhoud (alleen wanneer het nodig is). De andere turbine heeft die kenmerken niet. Welke zou je kopen, en voor welke zou je ietsje meer willen betalen?

GE denkt dat optimalisering van onderhoudswerkzaamheden via big data ook goed zal werken voor veel van zijn andere dure industriële goe-

deren, met inbegrip van locomotieven, straalmotoren en apparatuur voor medische visualisatie. Natuurlijk kunnen andere bedrijven in die bedrijfstakken voor dezelfde aanpak kiezen, en dat zullen ze waarschijnlijk op een gegeven moment ook doen. Maar de schaalgrootte van GE, de ambitieuze investeringsplannen en de vroege start op big-datagebied zullen waarschijnlijk aanzienlijk concurrentievoordeel opleveren.

Wat we niet weten – en voorlopig ook niet zullen weten

Het is nog niet duidelijk of vooral gigantische concerns als GE profijt hebben van big data en hoe andere bedrijven gaan reageren. We weten niet wat de bedrijfstak- en concurrentiedynamiek wordt voor door big data mogelijk gemaakte industriële producten en diensten van GE en zijn concurrenten. En er zijn veel meer onzekerheden op het gebied van big data waar we nog een hele tijd mee moeten leven. Maar het loont de moeite om nu over zulke kwesties na te denken, omdat het beter is om met een goed voorbereide reactie te komen dan om ad hoc te reageren.

We weten bijvoorbeeld niet hoe big data organisatiestructuren gaan beïnvloeden. Er is reden om aan te nemen dat de beschikbaarheid van big data – over werkzaamheden, werknemers, klanten en bedrijfsrisico's – in het voordeel zal werken van organisaties die hun capaciteiten om data vast te leggen en te analyseren gaan centraliseren. Dat zie je al gebeuren bij small-data-analytics: veel bedrijven zijn bezig om centraal gecoördineerde analysestrategieën en -groepen op te zetten. Als big data zich in de hokjes en silo's van organisaties gaan nestelen, wordt het heel moeilijk om de puzzel nog in elkaar te passen met als doel kansen voor de business te zien en er actie op te ondernemen. Omdat big data nog zo nieuw zijn, weten we echter niet welke vormen de organisatorische transformaties eromheen zullen aannemen. Vroege resultaten van grote bedrijven, die ik evalueer in hoofdstuk 8, wijzen erop dat big data gecombineerd gaan worden met bestaande data- en analyticsgroepen. Maar dat zou de komende jaren zomaar kunnen veranderen.

We weten evenmin hoe big data klantrelaties gaan beïnvloeden. Maar de richting die het gebruik van big data op gaat, is meer kennis over klantgedrag, likes en niet-likes en een betere afstemming van advertenties, producten en diensten op basis van die kennis. Overigens zijn er duidelijke aanwijzingen dat klanten die betere afstemming helemaal niet willen – en dat ze zeer wantrouwig zijn over wat bedrijven doen met hun data. Zo zegt 68 procent van de Amerikaanse internetgebruikers dat ze maatwerkadvertenties op basis van hun zoek- en websitebezoekgeschiedenis afkeuren.[18] Ondanks deze theoretische tegenwerpingen zijn Amerikanen vaak bereid om een substantiële hoeveelheid persoonlijke informatie te geven in ruil voor sociale interactie (zoals op Facebook) of kortingen. Dat neemt niet weg dat er wel degelijk verzet kan zijn in de VS – formeel of in de zin van klantgedrag – tegen *targeting* op basis van big data. In de EU is er al een sterke wettelijke impuls geweest om de exploitatie van klantgegevens door bedrijven te beperken. Ik zie niets dreigen aan de Amerikaanse wettelijke horizon, maar organisaties moeten zeker zorgvuldig de klantensentimenten op dit vlak monitoren, inclusief de mogelijkheid dat ze de 'griezeligheidsgrenzen' bij klanten overschrijden.

We weten ook niet hoe big data het management gaan veranderen. Big data bieden ons de mogelijkheid om veel vaker datagestuurde beslissingen te nemen. Bedrijven en instellingen zullen steeds meer te weten komen over de omgeving waarbinnen ze opereren en zijn straks in staat om analytics in te zetten – geautomatiseerd of ondersteund door menselijke besluitvorming – om te beslissen en handelen op basis van wat ze weten. Het is nog niet duidelijk in welk tempo managers deze nieuwe aanpak zullen adopteren. Maar de geschiedenis wijst uit dat het onwaarschijnlijk is dat ze dat überhaupt gaan doen. Het analyseren van *small data* is immers al tientallen jaren gemeengoed, en nog steeds nemen veel managers beslissingen op grond van hun onderbuikgevoel – en dit zal niet een-twee-drie veranderen. Hoewel mijn vrienden Erik Brynjolfsson en Andy McAfee er vaak over spreken en schrijven dat de 'hippo' (*highest paid person's opinion*) sterk in belang afneemt, is die diersoort nog lang niet uitgestorven in de organisaties waar ik over de vloer kom (al zou het absoluut wenselijk zijn dat dit snel gebeurt).[19] Als je erop rekent dat macht en politiek geen rol meer zullen spelen

in besluitvormingsprocessen, zul je hoogstwaarschijnlijk worden teleurgesteld – zeker op korte termijn. Het is veilig om je voor te bereiden op meer datagestuurd management in de toekomst, maar het precieze traject van die verandering is onduidelijk en zal sterk verschillen per organisatie.

Hoe dit boek verder gaat

We weten kortom dat de kansen en impact van big data substantieel zullen zijn, maar we weten nog niet in detail hoe dit bedrijven en bedrijfstakken gaat beïnvloeden. Laten we nu vooruitkijken naar de inhoud van de volgende hoofdstukken (je kunt hoofdstukken overslaan als sommige dingen je meer interesseren dan andere).

In hoofdstuk 2 beschrijf ik wat mensen en bedrijven in de beginfase doen met big data in belangrijke bedrijfstakken en bedrijfsfuncties, met daarbij diverse scenario's van wat ze in de nabije toekomst zouden kunnen gaan doen – wat wellicht iets meer duidelijkheid verschaft over het transformationele karakter van deze nieuwe resource.

In hoofdstuk 3 ga ik in op het ontwikkelen van een strategie voor big data – hoe je organisatie kan bepalen welke bedrijfsdoelstellingen ze wil realiseren met deze resource, of je moet inzetten op datadiscovery of productietoepassingen, en hoe snel je moet reageren op big-datakansen.

Hoofdstuk 4 behandelt een van de grootste remmen op succes bij big data: de menselijke factor. Ik beschrijf precies welke vaardigheden een datawetenschapper in huis moet hebben en bespreek een nieuw onderwerp: hoe gaan big data managementgedrag veranderen?

Als je steeds hoort over Hadoop en MapReduce en je eigenlijk geen touw kunt vastknopen aan zulk jargon, wil je misschien meteen door naar hoofdstuk 5: een op managers gerichte gids voor big-datatechnologie. Dat hoofdstuk gaat niet alleen over infrastructurele technologieën voor het managen van big data, maar behandelt ook enkele analytische benaderingen die nuttig zijn in de big-dataomgeving, zoals machineleren en visual analytics.

In hoofdstuk 6 bespreek ik vanuit breed perspectief wat er nodig is in organisaties – los van slimme mensen en nieuwe technologie – om succesvol gebruik te kunnen maken van big data. Als je het boek *Analytics at Work* hebt gelezen, waarvan ik coauteur ben, weet je misschien al wat het DELTA-model is. Dat model pas ik in dit hoofdstuk toe op big data. (Ik gebruik dit model met een kleine aanpassing ook in de bijlage achterin dit boek, in het schema voor de beoordeling of je organisatie klaar is voor big-dataprojecten.)

De laatste twee hoofdstukken gaan over wat big data betekenen voor organisaties van verschillende leeftijden en groottes. Hoofdstuk 7 focust op wat we kunnen leren van startups en online bedrijven, hoofdstuk 8 op wat grote, gevestigde ondernemingen met big data doen. In hoofdstuk 8 beschrijf ik een idee dat ik ´Analytics 3.0´ noem, dat aangeeft hoe bedrijven het beste van small data en traditionele analytics kunnen combineren met de big-databenadering.

De beloning voor de moeite die je neemt om al deze hoofdstukken te lezen is: je krijgt diepgaand inzicht in hoe je big data kunt laten werken voor jou en je organisatie. Aan het eind van elk hoofdstuk vind je een actieplan met actiegerichte vragen, die jij en je managementteam moeten stellen en beantwoorden om deze krachtige resource te gebruiken. Je zult in dit boek maar weinig superlatieven aantreffen over hoe geweldig groot die enorme hoeveelheid data wel niet is, eerder een visie op hoe je die data kunt inzetten. Vergeet niet wat ik aan het begin van dit hoofdstuk zei: het gaat er niet om hoeveel data je hebt maar om wat je ermee doet.

Hoe belangrijk zijn big data voor jou en je organisatie?

- → Heeft je managementteam nagedacht over welke nieuwe datasoorten jullie business en bedrijfstak nu en in de komende paar jaar kunnen gaan beïnvloeden?
- → Hebben jullie het over de term 'big data' gehad, en of die goed aangeeft wat je organisatie met data en analytics doet?
- → Zijn jullie bezig om de besluitvormingsprocessen te veranderen in de richting van een meer continue benadering, aangestuurd door de voortdurende beschikbaarheid van data?
- → Heeft je organisatie snellere en meer agile benaderingen geadopteerd voor het analyseren van en actie ondernemen op belangrijke data en analyses?
- → Beginnen jullie je meer te richten op externe informatie over je business en marktomgevingen?
- → Hebben jullie zwaar ingezet op big data?

Big data

... iedereen heeft ervan gehoord, maar hoe je er waarde uithaalt voor je business is een heel ander verhaal. *Big Data aan het werk* geeft jou als manager de informatie die je nodig hebt om big data te laten werken voor jouw organisatie.

Dit boek laat je zien:

- Waarom big data belangrijk zijn voor jou en je organisatie.
- Wat cruciale succesfactoren zijn voor het implementeren van welk big-dataproject dan ook.
- Hoe je de mensen aanneemt, inhuurt of ontwikkelt die ervoor zorgen dat big data voor je werken.
- Welke technologie je nodig hebt om big data in goede banen te leiden.
- Hoe big data leiden tot een nieuwe aanpak van het managen van analytics.

Thomas Davenport is hoogleraar Informatietechnologie en management aan het Babson College, en wetenschappelijk onderzoeker aan het MIT Center for Digital Business. Hij is onderzoeksdirecteur van het International Institute for Analytics en senior adviseur bij Deloitte Analytics.

'Davenport heeft een intelligent en praktisch boek afgeleverd voor iedereen die de opportunity's van big data wil ontsluiten – en de valkuilen wil vermijden.'

Rob Bearden, CEO van Hortonworks

vakmedianet

www.overmanagement.nl