

Tijdperk '3.0' in grootschalige gegevensanalyse

T.H. Davenport is hoogleraar IT en Management aan Babson College, fellow aan het MIT Center for Digital Business, senior-adviseur bij Deloitte Analytics en mede-oprichter van het International Institute for Analytics.¹

Het begrip *big data* raakt ook bij het grote publiek meer en meer ingeburgerd. Minder bekend is misschien de ontwikkelingsgeschiedenis van de technologie achter het begrip, en van de jongste trendbreuk daarin: *big data* en grootschalige gegevensanalyse worden nu ook de 'motor' voor de producten en diensten voor de massa.

Wie zich al jarenlang verdiept in ondernemingen die slim omgaan met gegevens zal het beamen: we maken alweer de derde omslag mee in grootschalige kwantitatieve of gegevensanalyse (*analytics*). De tijdperken die we achter ons laten zouden we 'vóór *big data*' respectievelijk 'na (de opkomst van) *big data*' kunnen noemen. Of, om een goed bij het onderwerp passende aanduiding te gebruiken: Analytics 1.0 werd gevolgd door Analytics 2.0. In de regel gaat het bij een versie 2.0 niet alleen om wat kleine aanpassingen, aanvullingen en verbeteringen; dan heet de volgende versie 1.1. Bij een versie 2.0 gaat het om een omvangrijkere herziening van het product op grond van nieuwe prioriteiten en technische mogelijkheden. En dat was zeker het geval toen veel ondernemingen nieuwe bronnen van ongestructureerde en snel stromende informatie gingen gebruiken: 'big data'.

De veranderingen waar we nu inmiddels mee te maken hebben, zijn dermate fundamenteel en verstrekkend dat we met recht kunnen spreken van 'Analytics 3.0'. De essentie is dit: er worden nu niet alleen meer op grote schaal gegevens verzameld en geanalyseerd om de processen in een organisatie te verbeteren, maar ook het aanbod. Er wordt nu 'intelligentie' verpakt in de producten en diensten die de klant afneemt.

In dit artikel zal ik deze nieuwe ontwikkeling schetsen. Ik zal betogen dat net zoals de eerste toepassingen van *big data* (Analytics 2.0) een grote verandering betekenden ten opzichte van Analytics 1.0, de innovaties die nu plaatsvinden bij enkele koplopers in het bedrijfsleven ook de

tekenen aan de wand zijn van een nieuw tijdperk. Wanneer een nieuwe manier van denken postvat over een krachtig concurrentiewapen en de toepassing ervan, moeten managers daar op allerlei manieren op inspelen. Verandering manifesteert zich razend snel op alle terreinen voor een onderneming. Nieuwe partijen worden actief, concurrentieverhoudingen gaan verschuiven, mensen moeten nieuwe technologie onder de knie krijgen, het talent in de arbeidsmarkt zoekt de spannendste werkgevers op.

Dat alles zullen managers de komende maanden en jaren meemaken. Om er doeltreffend op te kunnen inspelen, zullen zij verbanden moeten kunnen leggen. Bovendien zullen zij moeten onderkennen dat er zich een fundamentele verandering voltrekt in het denken over concurreren op basis van grootschalige gegevensanalyse. De ondernemingen die als eerste begrijpen welke richting die verandering uit gaat – die dus als eerste enig zicht weten te krijgen op Analytics 3.0 – zullen zich in de beste uitgangspositie bevinden om die verandering mede vorm te geven.

De ontwikkeling van grootschalige gegevensanalyse

Dit artikel is niet bedoeld als een brede schets van de ontstaansgeschiedenis van grootschalige gegevensanalyse. Maar het is wel nuttig om heel even stil te staan bij de vorige grote verandering en de context waarin die zich voltrok. Besluitvorming op basis van feiten is als zodanig natuurlijk al zo oud als besluitvorming zelf. 'Business analytics'

1. De ideeën in dit artikel werden ontwikkeld voor het laatstgenoemde instituut.

De auteur is tevens mede-auteur van *Keeping Up with the Quants*, in 2013 gepubliceerd bij Harvard Business Review Press. Bij dezelfde uitgever is een volgend boek van zijn hand in voorbereiding: *Big Data at Work*.

als discipline dateert echter van halverwege de jaren vijftig. Toen kwamen er hulpmiddelen ter beschikking om grotere hoeveelheden informatie mee te genereren en te verzamelen en daar bovendien veel sneller patronen in te herkennen dan een mens op eigen kracht kan doen.

Analytics 1.0 – het tijdperk van ‘business intelligence’

In de tijd van Analytics 1.0 werden belangrijke stappen vooruit gezet in het ontwikkelen van objectief, diepgaand inzicht in belangrijke aspecten van bedrijfsvoering. Managers kregen de beschikking over inzichten gebaseerd op feiten en hoefden in hun besluitvorming minder te vertrouwen op hun intuïtie. Voor het eerst werden gegevens over onder meer productieprocessen, verkoop en het contact met de klant geregistreerd, samengebracht en geanalyseerd.

Nieuwe computertechnologieën gaven de doorslag. De eerste informatiesystemen werden op maat gemaakt door bijzonder kapitaalkrachtige ondernemingen, later kwamen er algemenere systemen die meer bedrijven zich konden veroorloven. Dit was de tijd van het ‘enterprise data warehouse’, waarin informatie werd opgeslagen, en van de ‘business intelligence’-software waarmee die gegevens werden onderzocht en gerapporteerd.

Er waren ook nieuwe competenties nodig, allereerst om gegevens te beheren. Gegevensbestanden waren nog vrij klein en statisch, en konden dus ten behoeve van de analyse apart worden opgeslagen in de ‘warehouses’. Het was echter een hele klus om ze daarvoor geschikt te maken; analisten besteedden vooral daar veel tijd aan, en relatief weinig aan de analyse zelf.

Het was vooral belangrijk om de paar cruciale vragen te bepalen waar men zich op kon concentreren, want de analyse was bijzonder tijdrovend; vaak waren er weken of zelfs maanden mee gemoeid. De rapportages, die veruit de meerderheid vormden van het ‘business intelligence’-werk in die dagen, hadden bovendien alleen betrekking op het verleden. Er werden geen verklaringen of voorspellingen ontwikkeld.

Gold: grootschalige gegevensanalyse destijds als een bron van concurrentievoordeel? In algemene termen wel. Maar niemand had het toen al over ‘concurreren op basis van *analytics*’. De winst zat vooral in grotere operationele efficiëntie: op bepaalde belangrijke punten werden betere beslissingen genomen om het prestatieniveau te verbeteren.

Analytics 2.0 – het tijdperk van ‘big data’

In essentie heeft het tijdperk Analytics 1.0 ongeveer een halve eeuw geduurd, tot rond 2005.

Toen begonnen internetbedrijven en sociale netwerken, hoofdzakelijk bedrijven in Silicon Valley – Google, eBay e.d. – nieuwe soorten informatie te verzamelen en analyseren.

Het begrip ‘big data’ is toen niet meteen al bedacht. Maar door de nieuwe realiteit waar dat

begrip voor staat veranderde wel heel snel de rol van gegevens en grootschalige gegevensanalyse bij die bedrijven. Het onderscheid dat ontstond tussen *big data* en *small data* was ook ingegeven door de herkomst van gegevens: die kwamen niet langer uitsluitend uit de interne transactiesystemen maar ook van buiten. Er werden nu gegevens ontleend aan allerlei verschillende bronnen: internet, allerlei sensoren, openbare initiatieven als het menselijk genoomproject, geluids- en beeldopnamen.

Toen het tijdperk van Analytics 2.0 aanbrak, werd al snel duidelijk dat er krachtig gereedschap voor nodig was – en ook dat het heel winstgevend zou kunnen zijn om daarin te voorzien. Ondernemingen begonnen uit alle macht nieuwe capaciteiten op te bouwen en klanten aan te trekken. Natuurlijk gingen velen ervan uit dat ‘wie het eerst komt, het eerst maalt’. Die gedachte leidde tot een indrukwekkende rage, maar ook tot een ongekende versnelling in de ontwikkeling van nieuw aanbod. Kijk eens naar de verschillende gegevensproducten die LinkedIn heeft ontwikkeld: People You May Know, Jobs You May Be Interested In, Groups You May Like, Companies You May Want to Follow, Network Updates, Skills and Expertise. Daarvoor werd een krachtige infrastructuur ontwikkeld en werden slimme en productieve *data scientists* aangetrokken. Het bijzonder succesvolle Year in Review van LinkedIn, dat laat zien wie er in je netwerk allemaal van baan is veranderd, werd ontwikkeld in slechts één maand. LinkedIn is niet de enige die haast heeft. De CEO van een starter in *big data* zei tegen mij: ‘We hebben het geprobeerd met *agile* [software-ontwikkeling], maar dat ging te traag.’

Er moesten allerlei soorten technologie worden ontwikkeld, aangetrokken en bedwongen. *Big*

Tegenwoordig kunnen niet alleen online- en informatiebedrijven producten en diensten ontwikkelen uit gegevensanalyse, maar iedere onderneming in iedere branche

data paste niet, of kon niet snel genoeg worden geanalyseerd, op één enkele server. Dus werden de nieuwe grote gegevensvolumes verwerkt met Hadoop, een openbaar (*open source*) software-

Hoe zou het Ron Johnson zijn vergaan als hij bij J.C. Penney kleine experimenten had laten doen in plaats van al die grote (mislukte) veranderingen?

raamwerk waarmee verschillende ‘plukken’ aan gegevens tegelijkertijd, ieder op een andere server, konden worden verwerkt. Om tamelijk ongestructureerde gegevensvolumes te kunnen verwerken gingen ondernemingen aan de slag met een nieuw soort databanken: NoSQL. Veel informatie werd opgeslagen en geanalyseerd in openbare of afgeschermdes *cloud*-omgevingen. Andere technologieën die in deze periode het licht zagen waren onder meer ‘in memory’- en ‘in database’-analyse om snel kwantitatieve analyses uit te voeren. Met methoden om apparatuur te laten leren van uitgevoerde bewerkingen (semi-automatisch bouwen en testen van modellen) werden snel modellen ontwikkeld aan de hand van de snel stromende gegevens. Zwart-op-witrapporten maakten plaats voor kleurrijke, complexe diagrammen.

Kortom, Analytics 2.0 vereiste geheel andere competenties dan Analytics 1.0. De nieuwe kwantitatief-analisten heetten nu *data scientists* en beschikten over zowel computervaardigheden als analytische vaardigheden. Zij namen al snel geen genoegen meer met een ondersteunde rol achter de schermen, maar wilden werken aan het productaanbod en de verdere ontwikkeling van de onderneming.

Analytics 3.0 – het tijdperk van aanbod verrijkt met gegevens

Iemand met een scherpe blik kon in tijdperk 2.0 al de aanzet zien naar het volgende tijdperk. De pionierende *big data*-bedrijven in Silicon Valley gingen investeren in grootschalige gegevensanalyse om het product- en dienstenaanbod aan de klant te ondersteunen en invullen. Ze trokken meer bezoekers naar hun websites met betere zoekalgoritmen, aanbevelingen van vrienden en collega's, productaanbevelingen en uiterst scherp gerichte reclameboodschappen. Dit alles werd aangedreven door analyses van ontzaglijk grote hoeveelheden gegevens.

Analytics 3.0 markeert het moment waarop andere grote organisaties zich aansloten bij deze ontwikkeling van grootschalige gegevensanalyse.

Inmiddels kunnen niet alleen maar informatie- en onlinebedrijven producten en diensten ontwikkelen door gegevens te analyseren. Nu gaat het om iedere onderneming in iedere bedrijfstak. Iedere onderneming die iets maakt, vervoert, verbruikt of met klanten werkt, beschikt over steeds grotere volumes aan gegevens over die activiteiten. Elk apparaat, iedere zending en iedere consument laat een spoor aan gegevens achter. Ondernemingen zijn nu in staat om die gegevens te analyseren en op grond daarvan iets te ontwikkelen waar de klant en de markt hun voordeel mee kunnen doen. Iedere onderneming kan nu beslissingen aan de ‘voorkant’ van de bedrijfsvoering nemen en optimaliseren op basis van grootschalige gegevensanalyse.

Evenals de eerste twee tijdperken van grootschalige gegevensanalyse brengt ook dit derde tijdperk nieuwe uitdagingen en kansen met zich mee. Dat geldt zowel voor de ondernemingen die willen concurreren op basis van grootschalige gegevensanalyse als voor de leveranciers die er de gegevens en het gereedschap voor leveren. Hoe een onderneming kan profiteren van deze jongste ontwikkeling bespreek ik zo dadelijk. Laten we echter eerst eens bekijken hoe Analytics 3.0 er uitziet bij enkele bekende ondernemingen die veruit het grootste deel van hun lange bestaan hebben geopereerd in de traditionele fysieke markt.

De volgende ‘kraker’, in bèta

Het Duitse Bosch is al 127 jaar oud. In de toepassing van grootschalige gegevensanalyse bevindt deze onderneming zich echter allesbehalve in een voorbije eeuw. In alle business units zijn bij Bosch initiatieven genomen om met behulp van veel gegevens en de analyse daarvan ‘intelligent’ aanbod te ontwikkelen voor de klant. Denk aan intelligent wagenparkbeheer, intelligente infrastructuur voor het opladen van elektrische auto's, intelligent energiebeheer, intelligente analyse van beelden van beveiligingscamera's en nog veel meer. Om de kansen voor dergelijke innovatieve diensten in het vizier te krijgen en te ontwikkelen heeft Bosch een software-innovatiegroep ingesteld die zwaar inzet op *big data*, *analytics* en het ‘*internet of things*’.

Het Franse Schneider Electric is 170 jaar oud. Deze voormalige ijzer-, staal- en wapenfabrikant legt zich tegenwoordig hoofdzakelijk toe op energiebeheer, waaronder optimalisering van energieverbruik, slim netwerkbeheer en automatisering

voor gebouwen. Schneider Electric heeft allerlei software- en *data*-starters overgenomen of ontwikkeld in Silicon Valley, Boston en Frankrijk. De onderneming heeft bijvoorbeeld een geavanceerd distributiebeheersysteem (ADMS) voor de energiedistributie bij nutsbedrijven. Zij kunnen met ADMS hun netwerkkapapparaat bewaken en beheeren, storingen behandelen en teams aansturen. Met ADMS kunnen miljoenen datapunten over netwerkprestaties worden verzameld; technici krijgen aan de hand van visuele analyses inzicht in de status van het netwerk.

Een van de indrukwekkendste ontwikkelingen van een aanbod op basis van gegevens en *analytics* is te vinden bij General Electric, 120 jaar oud. GE is van oudsher een fabrikant, maar zijn 'fabrieken' ontwikkelen zich steeds meer tot leveranciers van optimalisatiediensten voor bedrijfsmiddelen en -processen. GE krijgt via sensoren een voortdurende stroom aan gegevens binnen over turbines, locomotieven, vliegtuigmotoren en apparaten voor medische beeldtechniek. Aan de hand daarvan kunnen optimale onderhoudsschema's worden ingepland. Om de juiste mensen voor dit soort werk aan te trekken en te ontwikkelen werd ruim twee miljard dollar geïnvesteerd in een nieuw software- en analysecentrum bij San Francisco. GE verkoopt nu aan andere industriële bedrijven technologie voor de aansturing van *big data* en *analytics*. Ook heeft GE nieuw technologisch aanbod ontwikkeld op basis van *big data*-concepten, waaronder Predix (een platform voor het bouwen van 'industrial internet'-toepassingen²) en Predictivity (24 applicaties voor de optimalisatie van bedrijfsmiddelen en -processen die in allerlei bedrijfstakken draaien op het Predix-platform).

En dan UPS, slechts 107 jaar oud. Dat biedt misschien nog wel het beste voorbeeld van de toepassing van grootschalige gegevensanalyse in de markt, in dit geval de bestelroutes. Voor UPS is *big data* niets nieuws; pakjes en transacties worden hier al sinds de jaren tachtig op afstand gevolgd. UPS bezorgt dagelijks zo'n 16,3 miljoen pakjes en krijgt per dag ongeveer 39,5 miljoen *tracking*-aanvragen. Die gegevens worden allemaal verzameld. UPS heeft nu ook de ruim 46.000 vracht- en bestelwagens uitgerust met telematica-sensoren om hun snelheid, richting, remgedrag en aandrijving te meten. Met de informatie die dat iedere dag opnieuw oplevert, krijgt UPS niet alleen inzicht in de dagelijkse vervoersprestaties maar worden ook de bestelroutes opnieuw uitgezet. Dit initiatief,

ORION,³ is misschien wel het allergrootste onderzoeksproject in *operations research* ter wereld. Het doel is om de routes van de chauffeurs met behulp van geografische gegevens online en optimaliseringsalgoritmen nog tijdens de rit te optimaliseren. In 2011 wist UPS de chauffeursroutes met 85 miljoen mijl in te korten; hiermee werd ruim 8,4 miljoen gallon aan brandstof uitgespaard.

Al deze voorbeelden hebben het volgende met elkaar gemeen: het management besloot om met grootschalige gegevensanalyse niet alleen de interne besluitvorming te verbeteren, maar ook hoogwaardiger producten en diensten te ontwikkelen. Dat is waar het bij Analytics 3.0 allemaal om draait.

Voor sommigen zal hiermee een oude voorspelling uitkomen. In hun boek *2020 Vision* uit 1991 betoogden Stan Davis en Bill Davidson dat ondernemingen hun bedrijf moesten 'informationaliseren', ofwel: producten en diensten ontwikkelen op basis van informatie. Davis en Davidson constateerden dat er bij ondernemingen informatie uit de 'uitlaatpijp' kwam. Die informatie konden ze opvangen om hun aanbod 'op te voeren'. Destijds trokken hun ideeën alleen de aandacht van ondernemingen die toch al actief waren in de informatievoorziening, zoals Quotron (beursgegevens) en OAG (gegevens over verkeersvluchten). Nu kan echter ook een bank, fabrikant, zorginstelling of detailhandelaar – kortom, wie maar wil, in welke bedrijfstak dan ook – met de gegevens waarover men beschikt waardevolle producten en diensten ontwikkelen.

Davis en Davidson schreven hun boek toen het voor een informatiebedrijf nog volstond om *informatie* aan te leveren. Tegenwoordig worden we echter overspoeld met informatie en hebben we nauwelijks de tijd om er inzichten uit te putten. De informatieleveranciers van gisteren moeten inzichtleveranciers worden. Zij moeten met behulp van grootschalige gegevensanalyse informatie absorberen en ons vertellen wat we ermee moeten doen. Online-bedrijven die beschikken over grote hoeveelheden gegevens ontleend aan het *click*-gedrag van hun bezoekers hebben dit als eerste beproefd: Google, LinkedIn, Facebook, Amazon en anderen zijn niet succesvol omdat ze klanten informatie verstrekken, maar omdat ze klanten in staat stellen snel beslissingen te nemen en te handelen. Ondernemingen uit de traditionele informatiebranche zijn inmiddels ook lang en breed deze weg ingeslagen.

2. 'Industrial internet' is de benaming bij GE voor het 'internet of things', ook wel 'machine-to-machine' (M2M)-communicatie genoemd; noot v. vert.

3. ORION: On-Road Integrated Optimization and Navigation.

Tien voorwaarden om te kunnen profiteren van Analytics 3.0

De geschetste verandering in oriëntatie betekent dat grootschalige gegevensanalyse een andere rol krijgt in een organisatie. Ondernemingen zullen zich bewust moeten zijn van allerlei bijkomende uitdagingen. En ze zullen erop moeten reageren met nieuwe capaciteiten, posities en prioriteiten.

Verschillende soorten gegevens, vaak in combinatie met elkaar

Er moeten zowel grote als kleine volumes aan gegevens worden samengebracht, uit zowel in- als externe

bron, en zowel gestructureerd als ongestructureerd. Die moeten tot inzichten leiden, in voorspellende en aanbevelende modellen die de frontlinie van de onderneming vertellen hoe deze haar werk moet doen. Wegtrans-

portbedrijf Schneider National, bijvoorbeeld, optimaliseert zijn logistiek door zijn algoritmen te voeden met gegevens verkregen uit sensoren. Deze onderneming houdt zo belangrijke indicatoren bij, zoals brandstofniveaus, container-locaties en -capaciteit, en het rijgedrag van chauffeurs. Schneider National probeert zijn routenetwerken steeds efficiënter te maken, zijn brandstofkosten te verlagen en het risico op ongelukken terug te dringen.

Nieuwe opties voor gegevensbeheer

In het tijdperk 1.0 analyseerden ondernemingen gegevens in hun data warehouses. In het tijdperk 2.0 werkten ze met Hadoop-clusters en NoSQL-databanken. Nu, in tijdperk 3.0, komen al die mogelijkheden bij elkaar: data warehouses, databanken en big data-toepassingen. In zo'n omgeving wordt traditioneel gegevensonderzoek gecombineerd met onder andere Hadoop (soms Hadoop 2.0 genoemd) en verticale en grafendatabanken. De IT-architecten moeten veel meer en veel complexere keuzen maken over het gegevensbeheer. Bijna iedere organisatie komt uiteindelijk uit bij een hybride beheer-omgeving. De oude oplossingen zijn niet verdwenen, maar er zijn nieuwe procedures nodig om gegevens en analyse door de staging, evaluatie-, verkennings- en productietoepassingen te voeren.

Snellere technologieën en analysemethoden

De *big data*-technologieën van het tijdperk 2.0 zijn aanzienlijk sneller dan eerdere generaties tech-

nologie voor gegevensbeheer en analyse. Ze worden nu aangevuld met nieuwe *agile* analytische methoden en geautomatiseerde leermethoden om veel sneller inzichten te ontwikkelen. Net als systeemontwikkeling met de *agile*-methode, worden bij deze methoden veelvuldig gedeeltelijke uitkomsten aangeleverd bij de betrokkenen bij een project; net als in het werk van de beste *data scientists* is de druk voortdurend groot. De uitdaging in het tijdperk 3.0 is om operationele, productontwikkelings- en besluitvormingsprocessen aan te passen om te kunnen profiteren van wat de nieuwe technologieën en methoden opleveren.

Ingebouwde analyse

Geheel in lijn met de grotere snelheid waarmee gegevens nu worden verwerkt en geanalyseerd, zijn de modellen in het tijdperk Analytics 3.0 vaak ingebouwd in de operationele en besluitvormingsprocessen. Daardoor nemen de snelheid en het effect van die modellen drastisch toe. Procter & Gamble, bijvoorbeeld, heeft een complete infrastructuur opgetuigd om grootschalige gegevensanalyse te integreren in de dagelijkse besluitvorming van het management: er zijn ruim vijftig *decision rooms* voor de verschillende *business spheres* en meer dan vijftigduizend *decision cockpits* op computers van medewerkers.

Sommige ondernemingen voeren de analyses volautomatisch uit in systemen die draaien op scoringsalgoritmen en regels gebaseerd op grootschalige gegevensanalyse. Bij sommige is grootschalige gegevensanalyse ingebouwd in het aanbod aan de consument. Hoe dan ook: integratie van grootschalige gegevensanalyse in systemen en processen levert tijdwinst op. Maar dat niet alleen: door die integratie kunnen managers in hun besluitvorming ook bijna niet meer om die analyses heen, en dat is meestal een goede zaak.

Gegevensontdekking

Ontwikkeling van producten en diensten aan de hand van gegevens vraagt om een effectief ontdekkingsplatform waar gegevens worden verkend met behulp van de vereiste vaardigheden en processen. Aanvankelijk waren *enterprise data warehouses* bedoeld om die verkenning en analyse te faciliteren. Maar bij veel ondernemingen zijn dit in de praktijk 'magazijnen' geworden waar productiegegevens enkel liggen opgeslagen; en zoals al gezegd kost de gegevensinvoer veel tijd. Met de ontdekkingsomgevingen kunnen zonder al teveel voorbereiding de wezenlijke kenmerken

Google en Amazon zijn niet succesvol omdat ze de klant informatie bieden, maar omdat ze hem helpen zo snel mogelijk te beslissen en handelen

van een bepaalde hoeveelheid gegevens worden bepaald.

Multidisciplinaire gegevensteams

Bij online-bedrijven en starters op het gebied van *big data* hebben de *data scientists* vaak de teugels in handen (of ze hebben in elk geval veel vrijheid in hun werk). Bij grotere en meer conventionele bedrijven zullen ze echter moeten samenwerken met allerlei verschillende partijen om ervoor te zorgen dat *big data* ook leidt tot *big analytics*. Vaak zijn de 'data scientists' bij dergelijke bedrijven in feite gewone kwantitatief-analisten die meer tijd kwijt zijn aan gegevensbeheer dan hun lief is (geen nieuw verschijnsel). Ondernemingen laten nu *data hackers*, die heel goed informatie weten te onttrekken en structureren, samenwerken met analisten, die precies weten hoe ze die informatie in modellen moeten onderbrengen.

Beide groepen moeten zich verstaan met IT – want dat bedrijfs onderdeel levert de *big data* en de analytische infrastructuur aan, faciliteert de 'proeftuinen' waarin de groepen de gegevens verkennen, en vertaalt de verkennende analyse in productiecapaciteiten. Gezamenlijk pakken ze op wat maar nodig is om de analytische klus te klaren, waarbij de afzonderlijke rollen elkaar veelvuldig overlappen.

Nieuwe topfunctie: hoofd analyse ('CAO')

Als analytisch werk zo belangrijk wordt, dan moet het topmanagement er ook direct toezicht op uitoefenen. De topfunctie 'chief analytics officer' (CAO) begint te ontstaan. Dit is een lid van de hoofddirectie met eindverantwoordelijkheid voor het opbouwen en inzetten van capaciteiten voor grootschalige gegevensanalyse. Organisaties die deze rol in de top inmiddels hebben ingesteld zijn onder meer AIG, FICO, USAA, het academisch ziekenhuis van de universiteit van Pittsburgh, Wells Fargo en Bank of America (evenals, een tijdje terug, de herverkiezingscampagne voor president Obama). De lijst wordt ongetwijfeld nog langer.

Analyses ter bepaling van de werkwijze

Er zijn altijd al drie soorten analyses geweest: beschrijvend (het verleden); voorspellend (prognoses op basis van modellen en gegevens uit het verleden); en voorschrijvend – specificaties voor de optimale werkwijze in de praktijk, op basis van modellen. Analytics 3.0 omvat alle drie deze vormen, maar de nadruk ligt op de laatste. Bij voorschrijvende modellen gaat het om grootschalige toetsing en optimalisatie; ze kunnen worden

gebruikt om *analytics* in te bouwen in cruciale processen en belangrijke handelingen van medewerkers. Dit kan grote operationele voordelen opleveren, maar het vergt wel een kwalitatief uitstekende planning en uitvoering. Als bijvoorbeeld het ORION-systeem van UPS onjuiste route-informatie doorgeeft aan de chauffeurs, dan gaat het niet lang mee. Topmanagers bij UPS zeggen dat zij veel meer tijd kwijt waren aan aandachtspunten in het veranderingsmanagement dan aan de ontwikkeling van algoritmen en systeemontwikkeling.

Gegevensanalyse op industriële schaal

Voor ondernemingen die met kwantitatieve analyse vooral hun interne besluitvormingsprocessen willen verbeteren, biedt Analytics 3.0 nu de mogelijkheid om die processen op volle industriële kracht te laten draaien. Dankzij het zelflerend vermogen van apparaten kunnen veel meer modellen worden gebouwd. Dat stelt ondernemingen in staat om veel gedetailleerder en nauwkeuriger prognoses te ontwikkelen. Neem bijvoorbeeld het jaarlijkse proces voor 'vraagontwikkeling' bij IBM, waarin wordt bepaald aan welke klanten de verkopers van IBM meer tijd en energie moeten gaan besteden. Vroeger had IBM 150 modellen voor dit proces. Vervolgens is samen met een klein bedrijf, Modern Analytics, een 'modellenfabriek' met een 'lopende band' voor gegevens ontwikkeld. Nu bouwt en onderhoudt IBM elk jaar zo'n 5.000 (vijfduizend!) van dergelijke modellen – en kan dat af met slechts vier medewerkers. De nieuwe systemen ontwikkelen 95 procent van die modellen zonder dat er ook maar één menselijke handeling aan te pas komt. Nog eens drie procent vergt minimale verfijning door een analist. Deze modellen hebben bovendien betrekking op zeer specifieke producten, klantsegmenten en geografische afzetgebieden. Een test uitgevoerd in een markt in Azië liet zien dat de respons van klanten ermee kon worden verdubbeld ten opzichte van een niet-statistische segmentering.

Nieuwe vormen van besluitvorming en management

De gegevenseconomie in de onderneming laten draaien op grootschalige analyse vraagt om nieuwe vormen van besluitvorming en management.

Bij UPS waren topmanagers veel meer tijd kwijt aan veranderingsmanagement dan aan de ontwikkeling van algoritmen en systeemontwikkeling

Bij een test bleek de respons van klanten met modellen te kunnen worden verdubbeld ten opzichte van niet-statistische segmentering

Die zullen vaak meer zekerheid bieden voordat tot actie wordt overgegaan. Managers moeten zich aanwennen om aan de hand van gegevens experimenten te laten uitvoeren. Ieder belangrijk initiatief zou altijd vooraf moeten worden gedaan door kleinschalige maar stelselmatige experimenten. Deze experi-

menten moeten onderworpen zijn aan strenge controles om oorzaken en gevolgen in kaart te kunnen brengen. Hoe zou het Ron Johnson zijn vergaan als CEO van J.C. Penney als hij dit had laten doen, in plaats van al die grote veranderingen te lanceren die voor het merendeel verkeerd uitpakt⁴?

Sommige veranderingen ten gevolge van de grote beschikbaarheid van gegevens brengen overigens nauwelijks meer zekerheid. *Big data* stroomt onophoudelijk; neem bijvoorbeeld analyses van merkbeleving ontleend aan sociale media. Die zullen onvermijdelijk fluctuaties te zien geven. De 'digitale rooksignalen', zoals ze wel zijn genoemd, kunnen duiden op ontkiemende problemen. Maar het zijn indicaties, geen hard bewijs. Managers zullen richtlijnen moeten invoeren om te bepalen wanneer alarmbellen vragen om beslissingen en actie.

Ook de aard van de *big data*-relaties brengt onze zekerheid met zich mee. Tenzij ze zijn ontleend aan formele toetsen, zijn de uitkomsten van *big data* doorgaans gebaseerd op correlaties en niet op oorzakelijke verbanden. Soms zullen die relaties bij toeval ontstaan (hoewel bij grotere hoeveelheden gegevens de kans toeneemt dat zwakke uitkomsten statistisch significant zijn). Sommige managers zullen zich hier ongemakkelijk bij voelen. Als het om een heel belangrijke kwestie gaat, dan kan het verstandig zijn om verder onderzoek te plegen alvorens een beslissing te nemen.

Bij analyses van grote hoeveelheden gegevens die tot gedragsspecificaties leiden, zal de aansturing van de medewerkers in de frontlinie van de onderneming vaak moeten worden veranderd. Ondernemingen zullen in ongekende mate in-

zicht krijgen in de activiteiten van chauffeurs, piloten, magazijnmedewerkers en iedere andere medewerker uitgerust met sensoren (en met de sensor in de smartphone misschien wel alle medewerkers). Medewerkers zullen zeker reserves hebben ten aanzien van een dergelijke controle. Gegevensanalyse die als nooit tevoren het gedrag van de individuele medewerker bloot legt, heeft iets 'griezeligs'. Al te gedetailleerde rapportages over hun doen en laten zal medewerkers een onprettig gevoel geven. In het tijdperk van Analytics 3.0 zullen er momenten zijn waarop het management 'even de andere kant op moet kijken'.

Waarde creëren in de gegevens economie

Vertegenwoordigt Analytics 3.0 de ultieme vorm van concurreren op basis van grootschalige gegevensanalyse? Misschien niet. Maar het lijkt er wel op dat dit de tijd is waarin ook de modale onderneming zich in de gegevens economie begeeft.

De online-bedrijven die de *big data*-sluizen hebben geopend, waren er van meet af aan op gebouwd. Zij hoefden *big data* niet te verenigen of integreren met traditionele informatiebronnen (en de analyses daarvan), want die hadden ze meestal niet. Evenmin hoefden zij *big data*-technologie te integreren in traditionele infrastructuren voor IT, want ook die hadden ze niet. Bij hen stond *big data* op zichzelf en was alles *big data*, zowel de gegevens als de analyse als de infrastructuur. Maar ieder van hen heeft nu een eigen versie van Analytics 3.0.

Eén ding is helder: zowel oudgedienden als starters hebben nieuwe capaciteiten nodig. Die kunnen niet worden ontwikkeld met de oude modellen voor analytische ondersteuning van de bedrijfsvoering. Het *big data*-model was een grote stap voorwaarts, maar het levert niet nog veel langer concurrentievoordeel op. Een goed prestatieniveau in de nieuwe gegevens economie vraagt om een fundamentele herziening van de manier waarop met gegevensanalyse waarde kan worden gecreëerd voor de onderneming en haar klanten. Analytics 3.0 is de veranderingsrichting en het nieuwe model voor concurreren op basis van grootschalige gegevensanalyse. ■

4. Ron Johnson, die eerder naam had gemaakt met de Apple stores, verliet J.C. Penney in april 2013 na een desastreus verlopen anderhalf jaar als CEO. Hedge-fundinvesteerder Bill Ackman, die hem had aangetrokken, zei n.a.v. zijn vertrek: 'Een van de grootste fouten was misschien wel dat er te snel te veel werd veranderd zonder eerst goed te toetsen wat het effect zou zijn.' (Noot v. vert., ontleend aan berichtgeving van Forbes en businessinsider).

Translated and reprinted by permission of Harvard Business Review. This article was originally published under the English title 'Analytics 3.0', by Thomas H. Davenport, in the December 2013 issue of the Harvard Business Review, vol. 91, issue 12, pp. 64-72. © 2013 Harvard Business School Publishing Corp. (Distributed by The New York Times Syndicate); all rights reserved. This translation, © 2014 Harvard Business School Publishing Corp. (Distributed by The New York Times Syndicate). Vertaling: Raymond Gijsen.