networkpro.nl

VAKBLAD VOOR
IT-BEHEER EN
SMART SOLUTIONS

1e jaargang, nummer 4 2012

BIG DATA

Big business in het petabyte tijdperk



Features

CLOUD COMPUTING
Er is meer dan Google apps

OPENSTACK
De slag om de API

HET BESTE DATACENTER
Onherkenbaar en onvindbaar?

coverNP0412.indd 1 12-12-12 11:41

COLOFON

Hoofdredacteur
Marcel Beelen
marcel@networkpro.nl

Content Manager
Ramon Wieleman

Eindredactie
Jaap de Wreede

Auteurs
Menno Hagens, Abram Wagenaar, Steff Gyssels, Ruud Koster, Erik de Ruijter, Sander van Vugt, Bram Dons, Anon Murphy, Jaap Veenstra

Vormgeving
Daniel Amoako

Uitgever
Martin Smelt

Traffic
Marco Verhoog

Media Order
Mirella van der Willik

Marketing
Ramon Wieleman
r.wieleman@hub.nl

Drukkerij
Senefelder Misset, Doetinchem

Distributie
Sandd

Advertenties
Janet Robben (j.robben@hub.nl)
Ivo Meijer (i.meijer@hub.nl)
Aart van der Giezen (a.vd.giezen@hub.nl)
Telefoon: 023 543 00 00
Fax: 023 535 96 27

Abonnementen
Network Pro verschijnt 6 maal per jaar en een jaarabonnement kost € 37,50 (België € 40,50).
Abonneren kan via www.hubstore.nl/networkpro.
Een nieuw abonnement wordt gestart met de eerst mogelijke editie voor een bepaalde duur. Het abonnement zal na de eerste (betalings)periode stilzwijgend worden omgezet naar abonnement van onbepaalde duur, tenzij u uiterlijk één maand voor afloop van het initiële abonnement schriftelijk (per brief of mail) opzegt. Na de omzetting voor onbepaalde duur kan op ieder moment schriftelijk worden opgezegd per wettelijk voorgeschreven termijn van 3 maanden. Een abonnement is alleen mogelijk in Nederland en België. U kunt mailen naar abonnementen@hub.nl of schrijven naar HUB Uitgevers, abonnementenadministratie, postbus 3389, 2001 DJ Haarlem.

Op werkdagen kunt u bellen tussen 10 en 14 uur naar telefoonnummer 023-5364401.

Verhuisberichten of bezorgklachten kunt u doorgeven via www.hubstore.nl (Klantenservice).

Abonnementenadministratie
Ingrid van der Aar, Tanja Ekel, Brigitte Hetem

AFSCHEID EN VOORUITBLIK

Onno Louwen, hoofdredacteur van dit in juni gelanceerde nieuwe vakblad, heeft het stokje overgedragen en HUB Uitgevers verlaten. Onno, bedankt voor je inzet! De basis voor Network Pro is gelegd, en de eer is aan mij om samen met het hele team het nieuwe vakblad in 2013 verder vorm te geven en voor de inhoud nog meer de focus te leggen op de B2B-doelgroep. Een blad voor IT-beheerders in de breedst mogelijke zin, met informatie over IT-producten, IT-diensten, hypes, trends en de ontwikkelingen op de markt. Maar daarbij vergeten we niet dat je vandaag de dag het werk ook meeneemt naar huis en andersom.

Dit is nummer 4 van dit jaar, waarmee we meteen 2012 afsluiten. Dit nummer staat in het teken van de cloud en datacenters. In een boeiend artikel beschrijft Bram Dons de open source-oplossing OpenStack, waarmee je zelf een cloud bouwt. Menno Hagens heeft goed in de cloud rondgekeken om handige cloud-apps te vinden. Erg interessant is ook de uitgebreide beschouwing van onze Belgische collega Stef Gyssels over hoe we een goed datacenter selecteren om onze machines onder te brengen of te hosten. In dit nummer deel 1 van dit tweeluik.



En weet u al wat Big Data is? De term komen we te pas en te onpas tegen, maar de achterliggende gedachten en oplossingen hebben wel degelijk potentie. Deze Network Pro bevat dus weer veel boeiende informatie.

Marcel Beelen
Hoofdredacteur Network Pro

Verantwoording

Network Pro is een uitgave van HUB uitgevers. Zonder uitdrukkelijke toestemming van de uitgever mag niets uit deze uitgave op enigerlei wijze worden overgenomen. Copyright © 2012 HUB uitgevers. HUB uitgevers en bij deze uitgave betrokken redactie en medewerkers aanvaarden geen aansprakelijkheid voor mogelijke gevolgen die zouden kunnen voortvloeien uit het gebruik van de in deze uitgave opgenomen informatie.

Wij nemen uw gegevens, zoals naam, adres en telefoonnummer op in een gegevensbestand. De verwerking van uw gegevens is aangemeld bij het College Bescherming Persoonsgegevens in Den Haag door HUB Uitgevers b.v., de verantwoordelijke voor uw gegevens. Uw gegevens worden gebruikt voor de uitvoering van met u gesloten overeenkomsten, zoals de abonnementenadministratie. Daarnaast kunnen wij uw gegevens gebruiken om u op de hoogte te houden van interessante informatie en/of aanbiedingen. Uw gegevens kunnen ook aan door ons zorgvuldig geselecteerde partijen ter beschikking worden gesteld. Uw gegevens kunnen, samen met hun informatie over u, worden geanalyseerd om de aanbiedingen en/of informatie zoveel mogelijk op uw interesses af te stemmen. U kunt bij het opgeven van uw gegevens bezwaar maken tegen beschikbaarheidsstelling van uw gegevens aan derden. Ook kunt u uw eigen gegevens opvragen en verzoeken ze te corrigeren of te verwijderen. Stuur hiertoe een kaartje aan de abonnementenadministratie. Postbus 3389, 2001 DJ Haarlem.

NP_INHOUD.indd 3 12-12-12 11:43

06 Alles in de cloud

Google levert met diverse apps een flinke bijdrage in de hedendaagse cloud computing, maar er zijn ook andere mogelijkheden. We bekijken in dit artikel enkele van Google Apps en de beste third party apps. Welke voordelen bieden deze cloud computing software voor jouw organisatie en kun je er kosten mee besparen? En hoe hoe kan je het beste samenwerken hiermee?



14 Bezoek aan Silicon Valley

"Silicon Valley is waar het allemaal gebeurt", hoor je altijd. Silicon Valley is de geboorteplaats van alle grote technologiebedrijven en de vele start ups in deze regio dromen allemaal van een dergelijke toekomst. Network Pro besloot eens rond te gaan neuzen in het Mekkah van de technologie en innovatie en brengt hierover verslag uit.



24 Big Business met Big Data

Big Data is de trend van 2012 en de grote bedrijven zoeken uit hoe ze hiermee om dienen te gaan. Maar wat is Big Data precies en hoe moet je als organisatie hiermee omgaan? En hoe zit de toekomst van Big Data eruit? Wij zochten het voor je uit en geven tips waar je op dient te letten bij implementatie en bij het gebruik van Big Data.



32 Datacenters: onherkenbaar en onvindbaar?

Je wilt als organisatie gebruik gaan maken van een datacenter. Maar hoe kies je uit alle datacenters de juiste voor jouw organisatie? Network Pro zoekt in dit eerste deel van een serie voor je uit waar je hierbij op dient te letten en wanneer je gillend dient weg te rennen. Dit eerste deel zal gaan over de uiterlijke kenmerken van een datacenter.



FEATURES

06 CLOUD COMPUTING
Er is meer dan Google Apps

10 SECURITY
Malware checken met Blue Coat

14 WORK & SOCIAL
Bezoek aan Silicon Valley

18 CLOUD COMPUTING
Bouw je eigen cloud

24 BIG DATA
De 3 V's verklaard

32 DATACENTERS
Een gebouw met zonnepanelen

RUBRIEKEN

22 SYSADMIN TOOLS
Linux tools

30 INFOSEC CORNER
Security nieuws

37 NICE TO HAVE
Y-cam Bullet HD 1080

OVERIG

28 COLUMN
Alles in de cloud?

38 GASTCOLUMN
Fresh air cooling

Volg NetworkPro ook op:

- www.networkpro.nl
- www.twitter.com/networkpro
- www.networkpro.nl/linkedin
- www.networkpro.nl/googleplus

NP_INHOUD.indd 5 12-12-12 11:42

Het petabyte-tijdperk

DE BIG DATA-UITDAGING

Big Data is de trend van 2012. McKinsey spreekt van 'Big Data: the next frontier' en Forbes heeft het Big Data Landscape uitgewerkt. Gartner heeft het thema Big Data toegevoegd aan het Hype Cycle Report 2012 over emerging technologies. En nu de eerste Dilbert over Big Data een feit is, mogen we toch wel concluderen dat Big Data 'hot' is. **MARCEL BEELEN**



24-27 Big_data.indd 24 05-12-12 14:53

Tien jaar geleden sprak niemand over Big Data. De term bestond niet en het bijbehorende probleem evenmin. De term Big Data is ontstaan ergens tussen 2005 en 2008; de bronnen geven hierover tegenstrijdige gegevens. Wel is duidelijk dat de eerste zoekmachines (Google, Yahoo, AltaVista) ook de eerste waren die met Big Data te maken kregen. De webcrawlers verzamelden en verwerkten immers voortdurend extreem veel data om de zoekresultaten zo snel en goed mogelijk weer te geven.

Big Data gedefinieerd

Big Data is op dit moment geen hype meer, maar een ware trend: het is de nieuwe uitdaging voor veel organisaties. Zo spreken we anno 2012 plotseling van een Big Data-uitdaging, een Big Data-strategie, een Big Data-specialist, een Big Data-oplossing, Big Data-producten en een Big Data-markt. Sommige bedrijven hebben zelfs al een Big Data Competence Center.

Maar wat is Big Data, anders dan een leuk klinkende marketingterm? Helaas heeft – net als bij BIG DATA - elke fabrikant die mee wil doen op deze groeimarkt, een eigen definitie die past bij zijn productportfolio. Big Data is dus feitelijk een containerbegrip. Maar we kunnen het eigenlijk simpel uitleggen: we spreken van Big Data als je gegevensverzameling niet meer te hanteren is met gangbare hardware of software. Een eenvoudiger definitie is eigenlijk niet te geven. Onder het hanteren van data valt het versturen en ontvangen van de data, het opslaan van de data, raadplegen en gebruiken van data, evenals het verwerken, analyseren en (grafisch) weergeven ervan.

Het digitale tijdperk

De voornaamste veroorzakers van Big Data vormen het groeiend aantal mobiele apparaten (smartphones, tablets), snelle online verbindingen (mobiel en vast) en social media. De laatste tien jaar zijn we wereldwijd massaal digitaal én online gegaan en is het digitale tijdperk ingeluid. Messaging, bellen over IP, downloaden van alles en nog wat, online games, online winkelen, digitale films en muziek (de laatste vooral dankzij Apples iTunes), digitale foto's en niet te vergeten social media als LinkedIn, Twitter, Facebook en Hyves: we leven in een tijd waarin we extreem veel data produceren en verplaatsen.

Wereldwijd is tussen de 2 en 3 zettabytes aan data opgeslagen. Dagelijks voegen we daar wereldwijd zo'n 2,5 * 10^18 bytes aan data (2,5 miljoen petabytes) aan toe. Dit is zoveel dat negentig procent van alle data pas de laatste twee jaar ontstaan is! We creëren elke twee dagen net zoveel data als we in de periode vanaf het ontstaan van de mensheid tot en met 2003 hebben gedaan. Deze feiten zijn eigenlijk schrikbarend.

We leven in het petabyte-tijdperk: organisaties zijn op dit moment data aan het hanteren in de orde van grootte van petabytes. Facebook heeft onlangs laten weten meer dan 100 petabytes aan gebruikersdata op te hebben geslagen, waarvan één petabyte foto's is (10 miljard foto's). Elk halfuur wordt er door Facebook 105 terabyte aan data gescand. De Internet Archive Wayback Machine bevat 2 petabyte data en groeit elke maand met 20 terabyte. IBM is vorig jaar begonnen aan een server die 120 petabyte aan data kan opslaan (voorzien van 200.000 harde schijven). Apple heeft vorig jaar een extra opslagsysteem van 12 petabyte gekocht voor iTunes. Experimenten met de deeltjesversneller bij CERN genereren dit jaar alleen al 22 petabyte aan data (en dat is nadat 99 procent van de meetgegevens al is verwijderd).

Internet of Things

Het petabyte-tijdperk duurt slechts kort en zal worden opgevolgd door het exabyte-tijdperk. Big Data is dus een blijvend fenomeen. En onze Big Data-uitdaging wordt groter en de explosie van data omvangrijker. Triljoenen bytes aan gegevens over klanten, toeleveranciers, producten, transacties en click-streams. Triljoenen bytes aan profielen, berichten, locatiegegevens, likes en un-likes, user-generated content, (streaming) HD video, audio en foto's. Triljoenen bytes afkomstig van sensoren, RFID-chips, NFC-communicatie in smartphones, auto's of winkels. Binnen niet al te lange tijd is elk apparaat dat stroom verbruikt, aangesloten op internet en produceert het enorme hoeveelheden extra data. We noemen dit 'The Internet of Things'.

kB	kilobyte	103
MB	megabyte	106
GB	gigabyte	109
TB	terabyte	1012
PB	petabyte	1015
EB	exabyte	1018
ZB	zettabyte	1021
YB	yottabyte	1024

Verschillend voor iedereen

Omdat het bij Big Data gaat om het niet meer kunnen 'hanteren' van gegevensverzamelingen, kun je op allerlei manieren een Big Data-uitdaging hebben. Het Big Data-probleem ontstaat omdat (database) bestanden veel te groot worden of omdat het aantal bestanden veel te omvangrijk is. Maar het kan ook zijn dat je zoveel gegevens in databases opslaat dat ze niet meer te verwerken zijn of dat het analyseren ervan onmogelijk wordt. Moet je snel gegevens verwerken (misschien wel real-time) en lukt dat niet meer, dan heb je eveneens een Big Data-probleem. Het kan bij Big Data gaan om gestructureerde gegevens (bijvoorbeeld in databaseomgevingen), maar ook om ongestructureerde gegevens (bijvoorbeeld mappen met bestanden of social media-content).

Hierdoor verschilt het per organisatie wanneer het probleem van Big Data ontstaat. De grens (in hoeveelheid opslagruimte, de hoeveelheid gegevens of verwerkingssnelheid) hangt af van je infrastructuur en hoe je gekozen hebt gegevens op te slaan en te verwerken. Zo zal de ene organisatie afhaken bij 1TB aan data en een ander bij 1PB aan data. De ene organisatie raakt de controle kwijt bij 10.000 records in een gestructureerde database; een ander heeft met 1 miljard records nog geen probleem. Zelfs als kleine zzp'er kun je een Big Data-probleem ervaren.

> "Big Data is een containerbegrip. We spreken van Big Data als je gegevensverzameling niet meer te hanteren is met gangbare hardware of software."

24-27 Big_data.indd 25 05-12-12 14:53



Datamining

Lang geleden leerden aankomende IT'ers tijdens hun studie al dat gegevens (data) iets anders is dan informatie. Aan pure gegevens heb je namelijk niets; pas als je de gegevens gaat verwerken, analyseren, classificeren en onderling relateren, ontstaat hieruit informatie. Een CAD-model is dus geen data maar informatie; een Word-document bevat geen data maar informatie. Informatie is kostbaar, terwijl data die automatisch of willekeurig worden gegeneerd, weinig specifieke waarde hebben. Een simpel voorbeeld: 250.000 tweets op Twitter hebben geen daadwerkelijke waarde (data). Als je ze gaat scannen en analyseren, ben je in staat bijvoorbeeld de populariteit van politici te achterhalen en de uitslag van de verkiezingen te voorspellen. Dit proces van het ontginnen van data tot informatie wordt ook wel datamining genoemd. Grote organisaties die aan datamining doen - zoals banken en overheden - hebben vaak een Data Warehouse. Dit is een gestructureerde en gecontroleerd database waarin de gegevens samenkomen uit allerlei processen, procedures en andere databases van de organisatie en zelfs daarbuiten. Datamining wordt losgelaten op het Data Warehouse en levert veel informatie op. Datamining is vaak een onderdeel van Business Intelligence (BI). Met BI zijn organisaties in staat uit het Data Warehouse informatie te filteren met datamining-technologie om direct hun bedrijf bij te sturen en belangrijke beslissingen te nemen. Je zou kunnen zeggen dat BI in staat is informatie om te zetten naar kennis. Datamining en BI bestaan al vele jaren. Sommige bedrijven doen er veel mee en halen er zelfs aantoonbaar concurrentievoordeel uit.

Andere bedrijven (en zeker het MKB) zijn niet met BI bezig, maar doen wel beperkt aan datamining (zonder dit waarschijnlijk te beseffen). Datamining en andere processen van BI beginnen echter het probleem te ervaren dat de data steeds moeilijker te hanteren worden: Big Data is geboren en traditionele data warehouses voldoen niet meer. Moet je data zelfs real-time analyseren, dan is er zelfs geen sprake meer van BI. Want bij BI kijken we in de achteruitkijkspiegel naar de aanwezige data. Data Analytics kijkt door de voorruit en kan voorspellende scenario's doorrekenen. Het niet (meer) kunnen hanteren van je bedrijfsdata betekent dat je als organisatie minder winst maakt en achterop gaat lopen in vergelijking met de concurrentie die het beter onder controle heeft. Je data zijn een goudmijn: door slim in alle databronnen te zoeken, real-time data te relateren, ze te combineren maar ook door ze op te schonen, krijg je zeer waardevolle bedrijfsinformatie in handen.

De drie V's

De verschillende aspecten van Big Data worden vaak samengevat onder de drie V's: volume, velocity en variety (vrij vertaald: hoeveelheid, snelheid en diversiteit). Het zijn drie aspecten die vooral te maken hebben met de business van datamining. Hoe maak je van data concrete informatie? Met behulp van Big Data Analytics.

Volume (hoeveelheid). Hoe meer data, des te beter of nauwkeuriger het wordt om voorspellingen te doen en trends te bepalen. Veel data daadwerkelijk analyseren is beter dan met een statistisch model en een kleine hoeveelheid data een voorspelling te doen. Bij volume gaat het dus om miljoenen of zelfs miljarden compacte of omvangrijke gegevens.

Velocity (snelheid). De snelheid waarmee data gegenereerd worden, is zeer hoog en neemt net als de hoeveelheid data alsmaar toe. Data zijn vaak bijzonder waardevol als deze nagenoeg real-time geanalyseerd en gebruikt kunnen worden. Als een klant in een webshop iets bestelt, is het handig meteen het profiel en de eerder geplaatste bestellingen te raadplegen om een aanbieding te presenteren die precies past bij deze of een eerdere aankoop. Data komen bijna streamend binnen, moeten geanalyseerd worden, opgeslagen worden en later weer verwerkt worden. De enorme hoeveelheid data moet binnen minuten of zelfs seconden worden verwerkt.

Variety (diversiteit). Bij Big Data gaat het niet om gegevens van een enkel datatype, zoals creditcardgegevens of e-mailadressen. Het gaat om een grote diversiteit aan redelijk losstaande data van allerlei datatypen die gecombineerd moeten worden. Op Facebook vinden we data als foto's, berichten, likes, favorieten, locaties, profielinformatie, vrienden en alle clicks die je op de site maakt. Gecombineerd met andere sites waarop je met het Facebook-account inlogt (bijvoorbeeld TripAdvisor) is dit een bonte verzameling data die op zichzelf weinig voorstelt, maar gecombineerd een oliebron vormt die erom vraagt ontgonnen te worden.

Big Data-infrastuctuur

Om te voldoen aan de eisen die de drie V's voor Big Data stellen, is een zware maar dynamische en elastische ICT-infrastructuur nodig. Hierbij spelen opslagsystemen (Big Data Storage), processorcapaciteit (Big Data Servers) en netwerkcapaciteit (Big Data Networks) een rol. Soms zie je overigens ook wel eens de indeling Big Data Storage, Big Data Analysis en Big Data Speed.

Big Data Storage. Alle data moeten uiteraard worden opgeslagen en hiervoor is veel storage nodig. De data zelf zijn wel-

24-27 Big_data.indd 26 05-12-12 14:53

licht kleine brokjes maar alle data samen vragen om een enorm datacenter met opslagcapaciteit. Een extra probleem hierbij is dat analyses van data zelf ook weer nieuwe data genereren die opgeslagen worden. En omdat relevante data nooit opgeruimd maar gearchiveerd worden, is er nog meer opslagcapaciteit noodzakelijk. Met de jaarlijkse groei van data is het bijna onmogelijk een schaalbare omgeving op te zetten. De enorme hoeveelheid oorspronkelijke en verwerkte data dient op gedistribueerde wijze opgeslagen te worden. Storagesystemen maar ook traditionele relationele databases worden geoptimaliseerd en aangepast om te kunnen voldoen aan de Big Data-ontwikkelingen.

Big Data Servers. Aan de pure opslag van Big Data heb je niets. Met Big Data Analytics worden de data geanalyseerd en gecombineerd tot nieuwe data. Hiervoor is rekenpower nodig in de vorm van servers. Die rekenkracht is niet alleen nodig voor het pure rekenwerk dat in batch kan gebeuren, maar tegenwoordig willen organisaties real-time analyse loslaten op binnenkomende data. In Australië is er bijvoorbeeld een webshop die de beschikking heeft over 2 miljoen creditcardgegevens van klanten en het bedrijf probeert real-time te voorspellen hoe klanten zich zullen gedragen en reageren op nieuwe producten en diensten. Er is behoefte aan zeer zware machines die in staat zijn dit te verwerken. We spreken dan van High performance computing (HPC). Het is mogelijk een supercomputer aan te schaffen die dit rekenwerk verzet. Een alternatief dat schaalbaarder is en potentieel een hogere beschikbaarheid biedt, is een grid van servers te gebruiken die de data gedistribueerd (en dus parallel) verwerken. Dit is typisch de 'cloud-manier' om het probleem aan te pakken.

Big Data Network. Veel Big Data worden getransporteerd vanaf webservers naar het back-office of blijft binnen een organisatie. Snelle netwerken zijn daar niet zo'n probleem. Als Big Data echter over het internet moeten worden verplaatst, kost dat capaciteit op vaste en mobiele netwerken. Om gigantische hoeveelheden data te transporteren zijn niet alleen snelle netwerken nodig, maar ook slimme apparatuur die de noodzakelijke hoeveelheid data die daadwerkelijk getransporteerd wordt, terugdringt en daardoor snelheid oplevert. Vooral dataverkeer naar smartphones en andere mobiele apparaten kan zeer problematisch zijn, omdat deze netwerken nog veel minder snel zijn dan het vaste net.

Big Data = veel open source

Een belangrijk product in Big Data-land is Apache Hadoop. Dit is een open source-framework voor Big Data waarmee grote datasets gedistribueerd worden verwerkt in een cluster bestaande uit duizenden computers. Het is dus een oplossing voor grid computing. Het algoritme dat parallelle verwerking mogelijk maakt, wordt MapReduce genoemd. Hadoop maakt gebruik van een gedistribueerd bestandssysteem dat HDFS heet.
Hadoop is een open source-project van Apache en wordt gebruikt en doorontwikkeld door onder meer Facebook. Maar ook bedrijven als IBM, LinkedIn, Microsoft, Rackspace, Sara, Spotify, Yahoo!, Adobe, Twitter, eBay en Microsoft werken met Hadoop. Twee organisaties geven de meeste ondersteuning aan Hadoop: Hortonworks en Cloudera. Beide bedrijven zijn honderd procent met Big Data bezig.
Voor zoekacties in Big Data bestaat een nieuw concept: NoSQL. De verzamelnaam geeft aan dat de database juist bedoeld is voor schaalbare niet-relationele databases met semigestructureerde data. Populaire NoSQL-databases zijn als open source te downloaden (zoals CouchDB en MongoDB) en er zijn commerciële spin-offs ontstaan (zoals DataStax). Als Apache-project is de NoSQL-database Cassandra te downloaden.

Emerging market

Op een opkomende markt als Big Data, waarop nog veel wordt gewerkt met open source-oplossingen, zien veel bedrijven kansen voor de toekomst. Aangezien de Big Data-infrastructuur draait om servers, storage en netwerken, promoten de meeste computer- en storagefabrikanten al Big Data en hebben ze Big Data-oplossingen (IBM, Intel, Dell, HP, Fujitsu, EMC, NetApp). Traditionele databaseontwikkelaars maken stappen richting Big Data (zoals Oracle).
Bij Big Data denk je meteen aan Big Brother die om de hoek komt kijken. Als allerlei databronnen als één geheel doorzocht kunnen worden en er relaties gelegd worden die nu nog niet mogelijk zijn, komt dit schrikbeeld van George Orwell wel erg dichtbij. Zaken als security en vooral privacy krijgen binnen Big Data-onderzoeken en -projecten nog veel te weinig aandacht en bedrijven die zich met security bezig houden doen nog niet mee. Grote afwezigen op de markt zijn ook fabrikanten die zich met netwerken en in het bijzonder met netwerkoptimalisatie bezig houden (F5, Juniper, Citrix, Cisco), maar dat is eveneens een kwestie van tijd.
Wikibon heeft een uitgebreide studie gedaan naar de Big Data-markt in 2011, waar per bedrijf is aangegeven wat de omzet is in Big Data. Big Data belooft de komende jaren voor veel fabrikanten, ontwikkelaars, dienstverleners en cloud-organisaties big money! ◘

> "De verschillende aspecten van Big Data worden vaak samengevat onder de drie V's: volume, velocity en variety."

Referenties en achtergrond

Orientdb
http://www.orientdb.org
Couchdb
http://couchdb.apache.org
MongoDB
http://www.mongodb.org
Hadoop
http://hadoop.apache.org
Hortonworks
http://hortonworks.com
Cloudera
http://www.cloudera.com
Wikibon
http://wikibon.org/wiki/v/Big_Data_Market_Size_and_Vendor_Revenues

24-27 Big_data.indd 27 05-12-12 14:53