

# Whitepaper

**Big Data**

**Henk Brands**



**Hoofdkantoor**

Kruisboog 42  
3905 TG Veenendaal  
Tel. +31(0)318 - 55 20 20  
Fax +31(0)318 - 55 23 55

**Kenniscentrum**

De Smalle Zijde 39  
3903 LM Veenendaal  
Tel. +31(0)318 - 50 11 19  
Fax +31(0)318 - 51 83 59

info.nl@infosupport.com  
www.infosupport.com  
K.v.K. 3013 5370  
BTW NL8062.30.277.B01

IBAN NL92 RABO 0305 9528 89  
BIC RABONL2U  
IBAN NL74 INGB 0004 7385 93  
BIC INGBNL2A

# Whitepaper

## Big Data

**Meer informatie** Voor vragen of meer informatie over deze whitepaper kunt u contact opnemen met Info Support door te bellen naar +31 (0) 318 55 20 20 en te vragen naar Sales Support & Marketing (Nederland) of te bellen naar +32 (0) 15 28 63 70 (België). U kunt ook een e-mail sturen naar [sales@infosupport.com](mailto:sales@infosupport.com).

© Info Support B.V., Veenendaal 2015

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze ook, zonder voorafgaande toestemming van **Info Support B.V.**

No part of this publication may be reproduced in any form by print, photo print, microfilm or any other means without written permission by **Info Support B.V.**

Prijsopgaven en leveringen geschieden volgens de Algemene Voorwaarden van **Info Support B.V.** gedeponeerd bij de K.v.K. te Utrecht onder nr. 30135370. Een exemplaar zenden wij u op uw verzoek per omgaande kosteloos toe.

## Inhoudsopgave

1. Big data is geen big issue	3
2. Wanneer zijn data 'big'?	4
3. Sla op	5
4. Analyseer	6
5. Cases	7
6. Conclusie	8
7. Over Info Support	9

## 1. Big data is geen big issue

IT-hypes volgen elkaar steeds sneller op. Wat gisteren nog totaal onbekend was, kan vandaag het onderwerp zijn van een levendig debat.

Big Data is zo'n hype die er ineens leek te zijn, en waar iedereen over mee wil praten. Maar praten we eigenlijk wel over hetzelfde? Of wordt er juist zo veel over Big Data gediscussieerd omdat we verschillende definities hanteren en dus langs elkaar heen praten?

Wie alle artikelen over het onderwerp leest, kan zich niet aan de indruk onttrekken dat de term inderdaad met veel vaagheid omgeven wordt. Dat zorgt ervoor dat bedrijven niet weten wat ze eigenlijk met Big Data aan moeten. Big Data is een soort vesting, die veel moois lijkt te beloven maar vooralsnog onneembaar lijkt.

Big Data zit bij veel bedrijven in de 'hier moeten we iets mee-fase'. In deze whitepaper willen we daar verandering in brengen, door eerst voor eens en voor altijd af te kaderen wat Big Data nu eigenlijk precies is (en dus ook wat niet). Daarna wordt besproken hoe je er als bedrijf vandaag al mee kunt beginnen en wat voor mogelijkheden het morgen biedt.

De Big Data-vesting is namelijk een stuk toegankelijker dan je in eerste instantie zou denken; als je eenmaal ziet welke loopbruggen er al liggen, dan lijkt de weg ernaartoe ineens een stuk minder ingewikkeld.

## 2. Wanneer zijn data 'big'?

De meeste onduidelijkheid rondom Big Data zit besloten in de eerste helft van de term. Bij 'big' denk je al gauw aan grote hoeveelheden (data). Dat is ook niet zo verwonderlijk: we slaan steeds meer data op. De verwachting is dat we in 2020 vijftig keer zo veel gegevens opslaan als in 2010 en dat die exponentiële groei voorlopig nog wel even door zal zetten.

Veel data dus. Toch is er meer. Ook zonder giga- of terabytes kun je spreken over Big Data. Gartner definieert de term als volgt:

*'High-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.'*

Het gaat dus niet alleen maar om grote hoeveelheden (high-volume) data, maar ook om data met een grote omloopsnelheid en met verschillende formats, uit meerdere bronnen.

Er wordt hierin vaak een verschil gemaakt tussen **gestructureerde** en **ongestructureerde** data. Gestructureerde data is alle informatie die in een vast format past; denk aan gegevens uit databases, financiële cijfers en spreadsheets. Onder ongestructureerde data valt alles wat niet eenvoudig kan worden geclassificeerd of in een hokje past, zoals foto's, video's, websites, PDF-bestanden, PowerPoint-presentaties, e-mails en Word-documenten.

Het tweede deel van Gartners definitie gaat over wat je met deze data dóet: het verwerken en analyseren van gegevens, om meer inzicht te krijgen en betere beslissingen te kunnen maken.

Doen we dat niet al jaren dan? Zo lang als er IT is, wordt data toch al opgeslagen en geanalyseerd? Dat klopt, alleen is het grote verschil dat er nu zó veel verschillende soorten data met zo'n enorme snelheid op ons afkomt, dat het een veel grotere opgave is geworden om hier nog nuttige informatie uit te halen.

Maar áls het lukt om de juiste analyse toe te passen, dan zijn de mogelijkheden erg groot. Niet voor niets zijn veel commerciële en open source-partijen bezig met het ontwikkelen van tools voor Big Data; van het analyseren, tot het visualiseren en vervolgens weer verwerken van gegevens.

Kortom: Big Data is nog best ingewikkeld, dus is het niet zo vreemd dat er verwarring is en organisaties nog niet zo goed weten hoe ze het fort kunnen beklimmen. Toch is de weg ernaar toe vaak simpeler dan je zou denken: dat begint met het verzamelen van data - zowel gestructureerde als ongestructureerde.

### 3. Sla op

Het klinkt zo eenvoudig: wil je beginnen met (het analyseren van) Big Data, dan zul je ze eerst moeten opslaan. Ook als je nu nog niets met deze gegevens doet, is het verstandig om ze alvast te verzamelen - ook wel het aanleggen van een 'data lake' genoemd. Gelukkig worden veel gegevens ook al opgeslagen en bewaard: denk aan financiële cijfers, facturen, personeelsgegevens, weblogs en andere data die in een ERP-systeem kan worden opgeslagen.

Zoals we in het vorige hoofdstuk hebben gelezen, zijn dat gestructureerde data. De belangrijkste eerste stap die bedrijven kunnen zetten op weg naar een succesvolle toepassing van Big Data is het toevoegen van ongestructureerde data aan gestructureerde data.

Laten we als voorbeeld een bouwbedrijf nemen. In de bouwtekeningen die worden gebruikt, zit veel ongestructureerde informatie: van afmetingen en ophangsystemen tot aan het aantal stopcontacten dat aanwezig is. Het combineren van dit soort informatie met gestructureerde data kan interessante inzichten opleveren - maar daarover meer in het volgende hoofdstuk.

Ongestructureerde data uit het verleden kun je (met terugwerkende kracht) boven water halen met business discovery en data visualisatie tools. Deze oplossingen zorgen ervoor dat ongestructureerde data meer gestructureerd wordt, zoekbaar is en dus ook visueel analyseerbaar.

Minstens net zo belangrijk is het om te bepalen welke ongestructureerde data je vanaf nu wilt opslaan, en hoe deze het beste bewaard kan worden. De vraag welke data belangrijk is om te bewaren en welke niet, is nog best ingewikkeld en zal in elke organisatie anders beantwoord worden. Een woningcorporatie zal bijvoorbeeld bijzonder geïnteresseerd zijn in data die iets zegt over noodzaak onderhoud aan een woning te plegen. Een eigenaar van een webshop is waarschijnlijk nieuwsgierig naar het profiel van bezoekers van zijn website en wie het meest geneigd is om bepaalde producten te kopen. Ook speelt wet- en regelgeving met betrekking tot privacy een belangrijke rol als het gaat om wat er opgeslagen en bewaard mag blijven.

De eerste stap is dus de vraag stellen welke informatie belangrijk is voor de business. Het spreekt voor zich dat dit een nauw overleg moet zijn tussen de business en IT. Vervolgens kan pas worden bepaald welke oplossing hierbij gaat passen.

## 4. Analyseer

Zoals eerder beschreven is het analyseren van data niets nieuws onder de zon. Het gebeurt al zolang er administratie en boekhouding bestaat: een jaarrekening analyseert bijvoorbeeld hoeveel geld er in een periode is binnengekomen in een organisatie en hoeveel er uit is gegaan, om de hoogte van de winst te kunnen bepalen.

Je zou kunnen zeggen dat data-analyse zich tot voor kort bijna altijd richtte op het verklaren van het heden op basis van gegevens uit het verleden. Een logische volgende stap is het voorspellen van de toekomst op basis van gegevens uit het verleden en heden - ook wel predictive analytics genoemd. Voorspellen is de ultieme manier om het bedrijfsresultaat of te verbeteren; als je succesvol voorspelt wat je doelgroep gaat doen, kun je daar op inspelen en proactief je aanbod daarop aanpassen.

De potentie van Big Data zit hem vaak in het vinden van verbanden en patronen waar je ze misschien niet meteen verwacht had. Met analytics-technieken kun je correlaties testen tussen verschillende variabelen en kom je er misschien wel achter dat er een verband is tussen het aantal stopcontacten in een huis en de jaarlijkse onderhoudskosten.

Er is wel enige voorzichtigheid geboden als het gaat om correlatie tussen twee variabelen. Het feit dat een correlatie bestaat, betekent nog niet dat er ook een causaal verband is. Om een voorbeeld te geven: onderzoek zou zomaar eens kunnen aantonen dat er een correlatie is tussen het dragen van een zwemvest en de kans dat je verdrinkt. Dat wil natuurlijk niet zeggen dat het dragen van een zwemvest af te raden is; er is echter een derde factor die invloed heeft op de andere twee variabelen, en dat is of de onderzochte personen überhaupt gaan zwemmen. Voordat je conclusies trekt over verbanden, is het dus belangrijk om je gezond verstand te gebruiken en meetgegevens te interpreteren.

Welke vaardigheden zijn er eigenlijk nodig binnen een organisatie om Big Data toe te passen? Het spreekt voor zich dat het bouwen en onderhouden van een dataset een basisvoorwaarde is. Los daarvan heb je analytische skills nodig, die zowel bij de business als bij IT aanwezig moeten zijn. Tenslotte is de betrokkenheid van het management van cruciaal belang: de visie van het bedrijf bepaalt immers waar je naartoe wilt met de data.

## 5. Cases

Door de enorme populariteit van de afgelopen jaren zou je bijna vergeten dat er al succesvolle voorbeelden zijn van de toepassing van Big Data. In dit hoofdstuk bespreken we een paar van deze voorbeelden.

### **Belastingdienst**

Soms zijn er verbanden tussen gegevens waar je ze niet zou verwachten. Zo ontdekte de Belastingdienst dat er een grotere kans was dat iemand belastingfraude pleegde bij een bepaalde combinatie van cijfers in het sofinummer. Bij een analyse kwam boven water dat dit verband te verklaren was: de sofinummers werden in 1975 per wijk uitgedeeld. Wat bleek: de sofinummers met een verhoogde kans op belastingfraude hoorden bij mensen die in 1975 in Wassenaar woonden.

### **Uitkeringsfraude**

Big Data lenen zich sowieso uitstekend om fraude op te sporen. Verschillende gemeentes in Nederland maken gebruik van de Smartbox-oplossing, waarin data van verschillende gemeentelijke databases wordt gecombineerd om uitkeringsfraude op te sporen. Als iemand recht heeft op een uitkering maar bijvoorbeeld in een jaar verschillende bouwvergunning aanvraagt, dan krijgen ambtenaren een signaal dat er een verhoogde kans is op uitkeringsfraude. Vervolgens kunnen met deze gegevens aan de slag om te onderzoeken of er daadwerkelijk sprake is van fraude.

### **Voorspellen van misdaad**

Politiekorpsen in Amerika gebruiken predictive analytics om te voorspellen waar misdaden zullen gaan plaatsvinden. Hiervoor gebruiken ze allerlei sensoren, zoals camera's op straat, maar ook aangiftes die zijn gedaan van misdaden en observaties van mensen waarvan crimineel gedrag wordt vermoed. Al die informatie wordt gecombineerd en in meldkamers wordt op stadsplattegronden real time geprojecteerd waar zich risicogebieden bevinden. De politie probeert dus te voorspellen waar misdaad gaat plaatsvinden en kan meer politie worden ingezet in wijken waar er sprake is van een verhoogd risico. IBM claimt zelfs dat met behulp van predictive analytics-oplossingen het misdaadcijfer in de stad Memphis met 30 procent is gedaald.



## 6. Conclusie

De complexiteit rondom Big Data schrikt veel organisaties af om 'er iets mee te gaan doen', terwijl de drempel in werkelijkheid niet zo hoog hoeft te zijn. Als je maar weet hoe je eraan moet beginnen: met het bewaren van (ongestructureerde) data. Het analyseren van deze data is een volgende stap.

Info Support denkt op strategisch niveau mee met wat er mogelijk is met Big Data: zowel in de uitvoering (hoe verzamel en analyseer je data) als op strategisch gebied (welke informatie wil je als organisatie uit de systemen halen en hoe pas je de processen hierop aan).

*Geschreven door Henk Brands, Info Support*

## 7. Over Info Support

Info Support is opgericht in 1986 en is met ruim 350 medewerkers in Nederland een vooraanstaand IT-dienstverlener op het gebied van IT-consultancy, software -ontwikkeling, opleidingen en beheer. Info Support is niet beursgenoteerd en financiert de verdere ontwikkeling van de organisatie op basis van een beheerste groei uit eigen middelen.

Onze drive achter de oplossingen die wij realiseren voor onze klanten is er sterk op gericht bedrijfsprocessen sneller en beter te maken. Info Support ontwikkelt en beheert solide en innovatieve softwareoplossingen die organisaties ondersteunen bij het realiseren van hun doelstellingen.



De kernwaarden Soliditeit, Integriteit, Vakmanschap en Passie typeren onze werkwijze, waarin we sociaal en solide management belangrijker vinden dan omzetmaximalisatie. Ons hoogste doel is dat we met opdrachtgevers en medewerkers willen bouwen aan langetermijnrelaties. Daarbij houden we ons aan gemaakte afspraken. Dit maken we in de praktijk waar, getuige de jarenlange relaties die we met onze klanten hebben. Info Support mag zich al 16 jaar op rij TOP-IT-werkgever van het jaar noemen.

Zie voor meer informatie [www.infosupport.com](http://www.infosupport.com).