

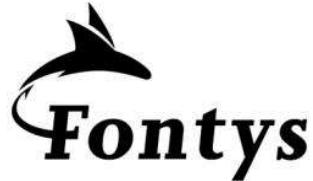


# Big Data

Eric van Tol



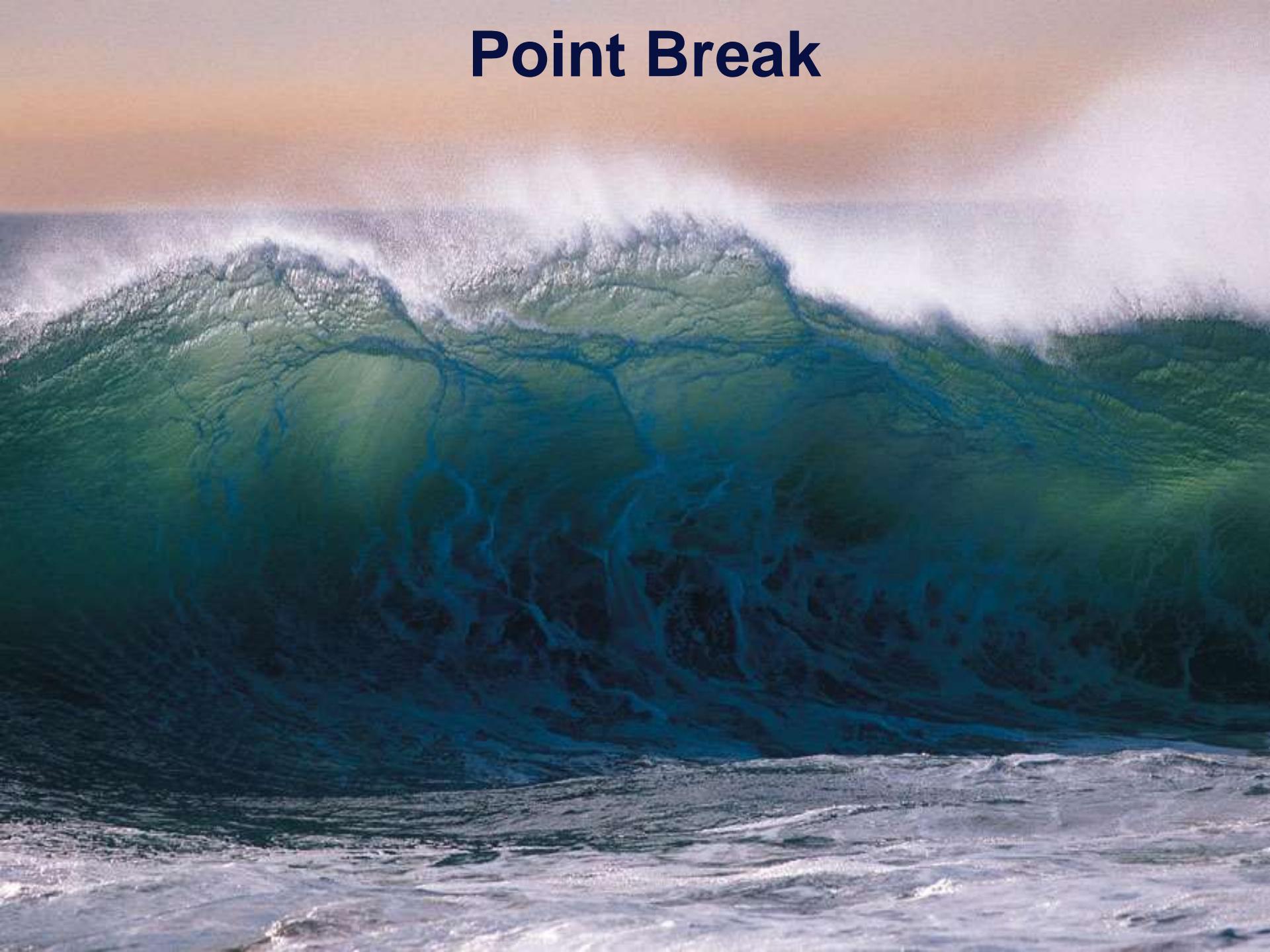
Ministerie van Economische Zaken,  
Landbouw en Innovatie



**absorberen en effectief maken  
van beschikbare Big Data kennis**



# Point Break

A photograph of a massive, curling green wave breaking at sunset. The wave's face is a vibrant green, transitioning to white spray at the crest. The background is a warm, orange and yellow sky. The water in the foreground is dark and choppy.

# Motivatie om...

gerichte en gelokaliseerde diensten te leveren

detail inzicht in gedrag van de klant te verkrijgen

proces optimaliseren over bedrijfsketens heen

**niet de boot willen missen**

# **De angst om...**

**jonge technologie te omarmen**

**waardevolle data te delen**

**de klant te verliezen**

**af te gaan**



**64% van de organisaties investeert of heeft  
plannen te investeren in big data  
technologie**

**terwijl minder dan 8% het echt doet**

September 2013 survey of 720 Gartner Research Circle members worldwide

**Gartner®**

**Big Data & Open Data verbeteren Bruto  
Nationale Product Europa met 1.9%**



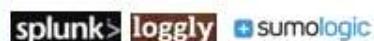
*The Warsaw Institute of Economic Studies*

# Productoverzicht

## Vertical Apps



## Log Data Apps



## Ad/Media Apps



## Business Intelligence



## Analytics and Visualization



## Data As A Service



## Analytics Infrastructure



## Operational Infrastructure



## Infrastructure As A Service



## Structured Databases

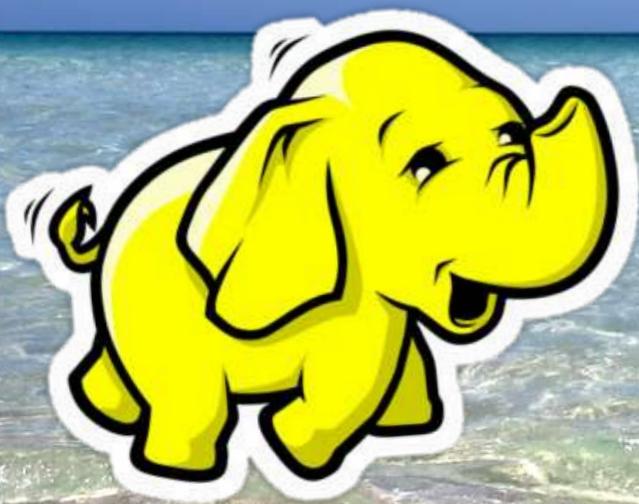


## Technologies



The big data landscape (Matt Tuck, 2012)

Hadoop open source omgeving analyse en bewerken van grote hoeveelheden veelal ongestructureerde data.





**Apache Hadoop is een open source omgeving voor analyse en bewerken van grote hoeveelheden ongestructureerde data.**

Hadoop omgeving omvat veel toepassingen en wordt door veel commerciële leveranciers (zoals IBM, Microsoft en Oracle) bedrijfsgeschikter gemaakt

**Open source verbeteringen vanuit de universitaire wereld:**  
Asterix en Stratosphere

**De ‘Hadoop’ Redhats:** Cloudera, Hortonworks en mapR

**Van preprocessing (staging archiving) tool naar central stage**

**Van 100.000 \$ per terabyte naar 1000 \$ per terabyte**

A large, powerful ocean wave is shown in the background, breaking towards the shore. The water is a vibrant greenish-blue, with white spray and foam at the crest. The wave's face is curved, showing the energy and movement of the water.

**Point Break**

# 1997

IBM Deep Blue verslaat G. Kasparov

## Alle combinaties doorrekenen



# 2011

IBM Watson wint Jeopardy quiz

## Meest waarschijnlijke antwoord bepalen uit Big Data



**Watson analyze the thousands of textbooks every second.**

**Watson can understand the nature of a question and review large amounts of information,**

**patient's electronic medical record, textbooks and journal articles,**

**then offer a list of suggestions with a confidence level assigned to each**

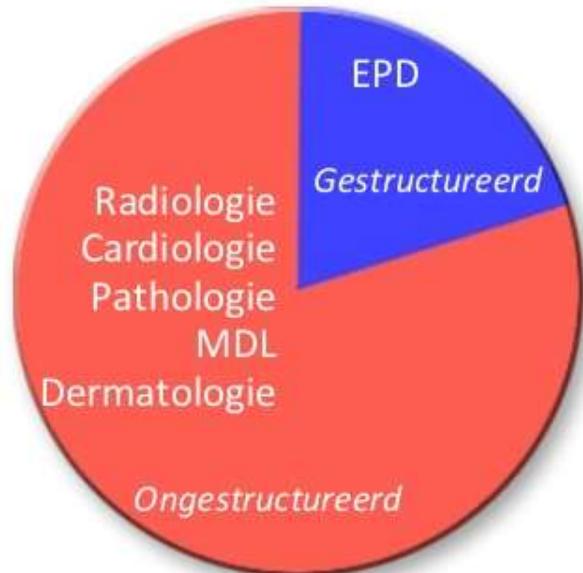
**Date: 04-12-2012**

**Source: The New York Times**

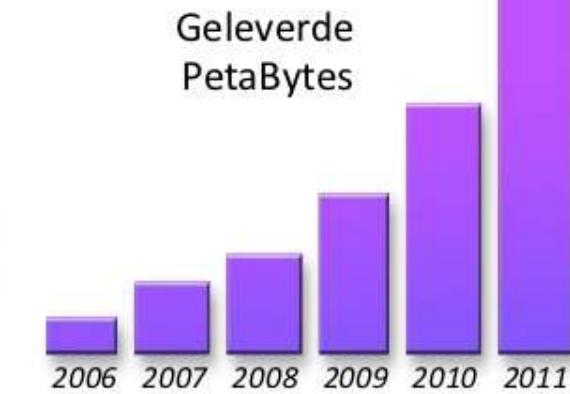
**Subject: For Second Opinion, Consult a Computer?**



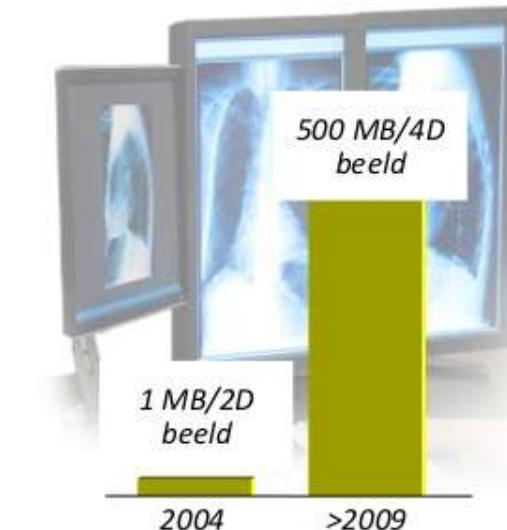
# Zorgdata – hoge volumes en enorme variatie



Ongeveer **80%** van alle opgeslagen zorgdata is *ongestructureerd*<sup>1</sup>



Zorgdataopslagcapaciteit groeit met **35%** per jaar<sup>2</sup>



**30%** van de wereldwijde gegevensopslag bestaat uit medische beelden<sup>3</sup>

Data jaarlijks uitgewisseld tussen zorginformatiesystemen:

In 2010: 283 terabyte, en in 2020: 78 petabyte - 774 miljoen verbonden apparaten<sup>4</sup>

<sup>1</sup>AIIM website, geaccepteerd percentage

<sup>2</sup>Recente studie door de Enterprise Strategy Group

<sup>3</sup>IBM Global Technology Outlook for 2005

<sup>4</sup><http://www.machinaresearch.com/healthcare2020.html>



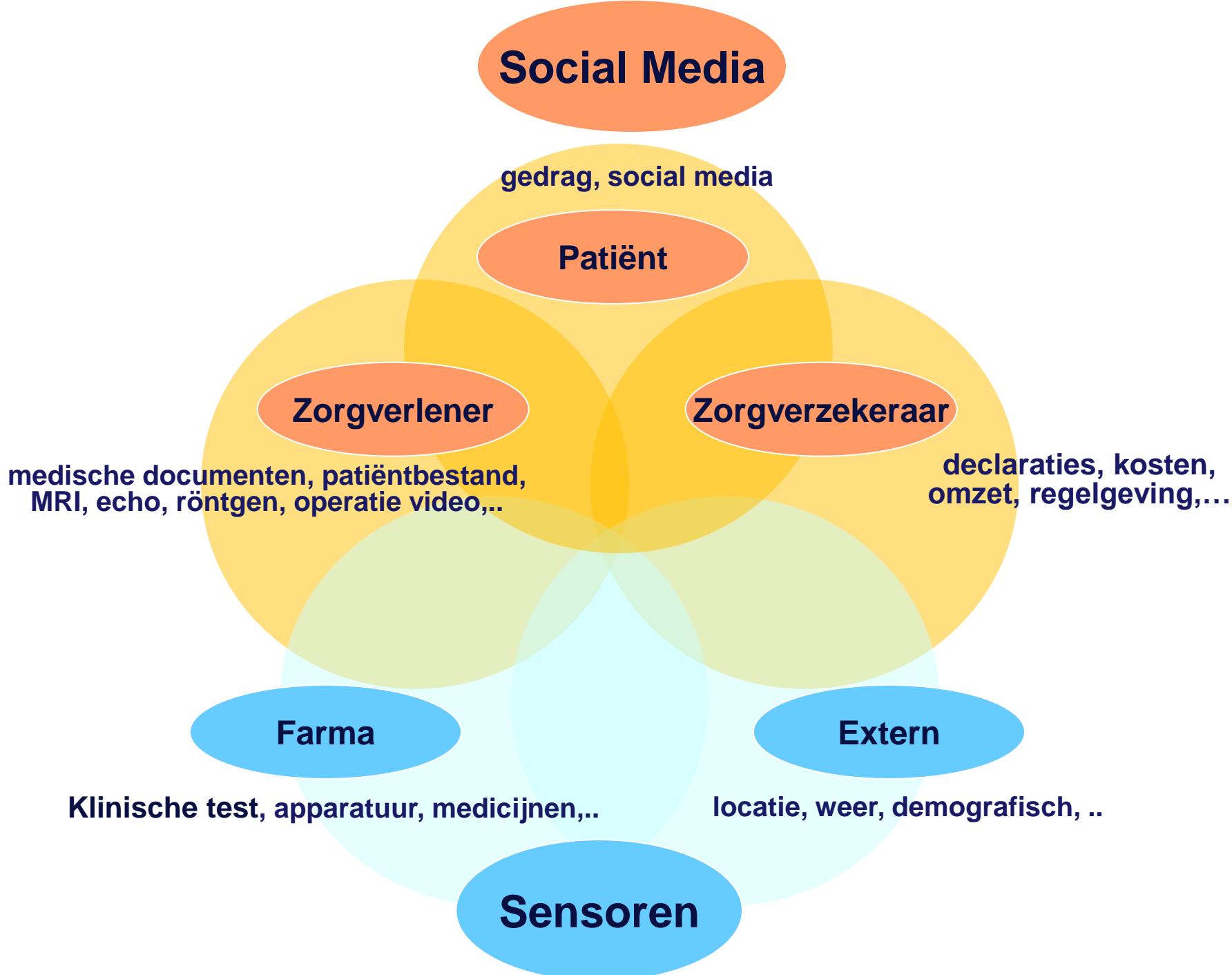
Quantified Self

**The Quantified Self is a movement to incorporate technology into data acquisition on aspects of a person's daily life in terms of inputs (e.g. food consumed, quality of surrounding air), states (e.g. mood, arousal, blood oxygen levels), and performance (mental and physical).**

**Such self-monitoring and self-sensing, which combines wearable sensors (EEG, ECG, video, etc.) and wearable computing, is also known as lifelogging.**

**Other names for using self-tracking data to improve daily functioning are “self-tracking”, “auto-analytics”, “body hacking” and “self-quantifying”.**



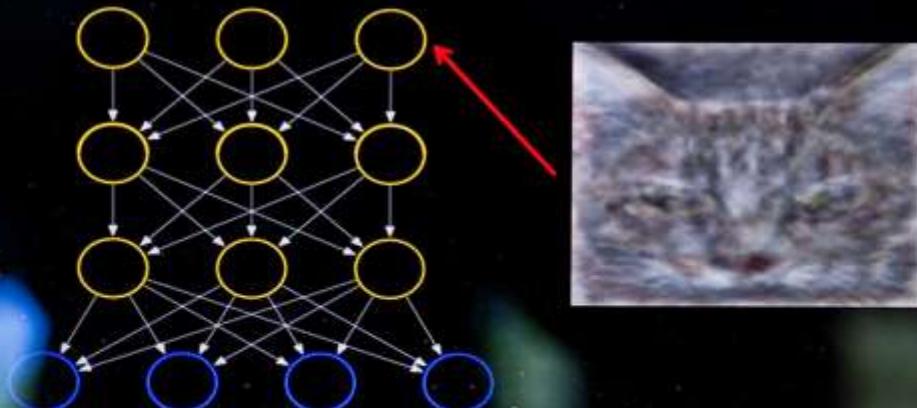




**leeralgoritme van  
Google toegepast op  
You-Tube**

**leert zichzelf wat een  
kat is**

## Deep Learning



Verwerken van 10 miljoen beelden met 16.000 processoren 10 dagen lang

Date: 08-05-2013

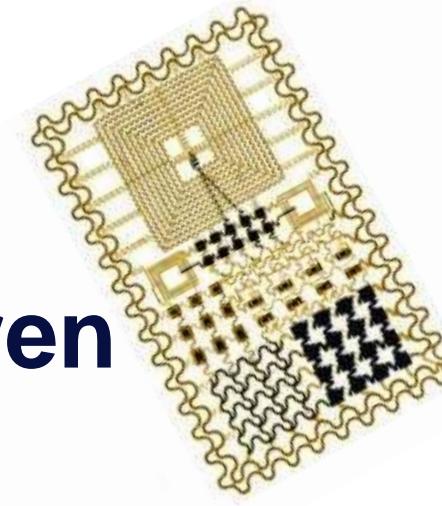
Source: WIRED

Subject: The Man Behind the Google Brain: Andrew Ng and the Quest for the New AI

**Stanford professor Andrew Ng, de man in het centrum van de Deep Learning beweging.**

**Is de menselijke intelligentie gebaseerd op een enkel algoritme?**





# Big Data kernbronnen zijn gebruikers en sensoren

**Sensoren:** “internet of things”, elk apparaat een IP adres en veel mobiele apparaten

**Gebruikers:** toepassingen (Web 2.0) en Social Media



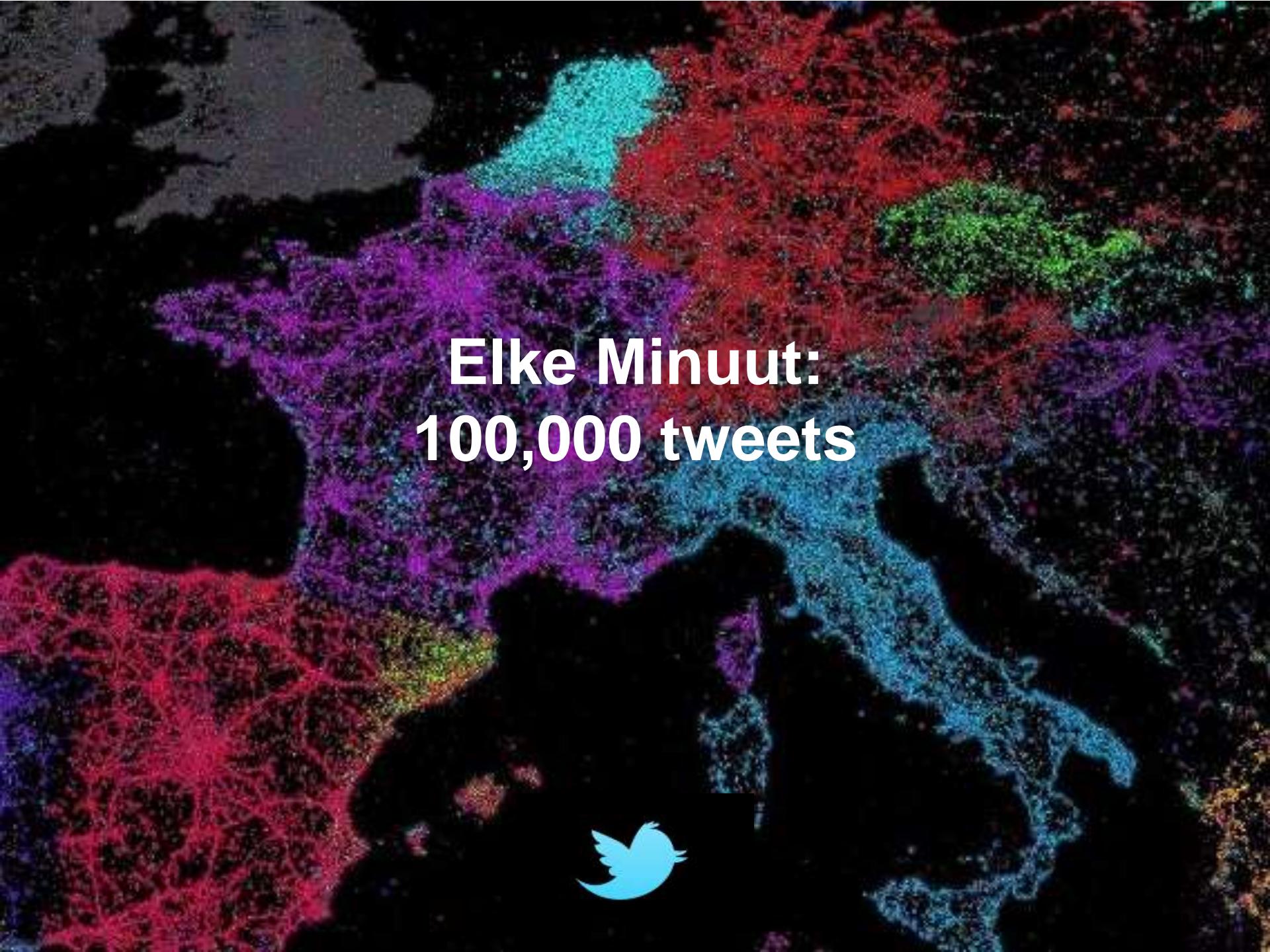


A world map against a dark blue background, where glowing blue lines represent global network connections between continents, forming a dense web of communication.

**Elke Minuut:  
204,166,677 e-mail messages**



**Elke Minuut:  
100 uur You Tube Filmpjes**



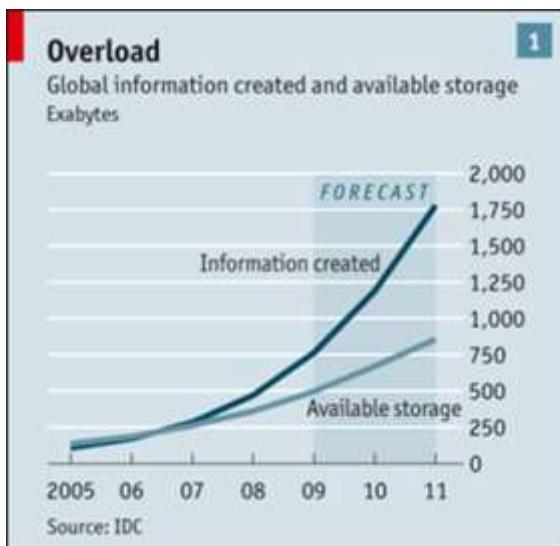
**Elke Minuut:  
100,000 tweets**





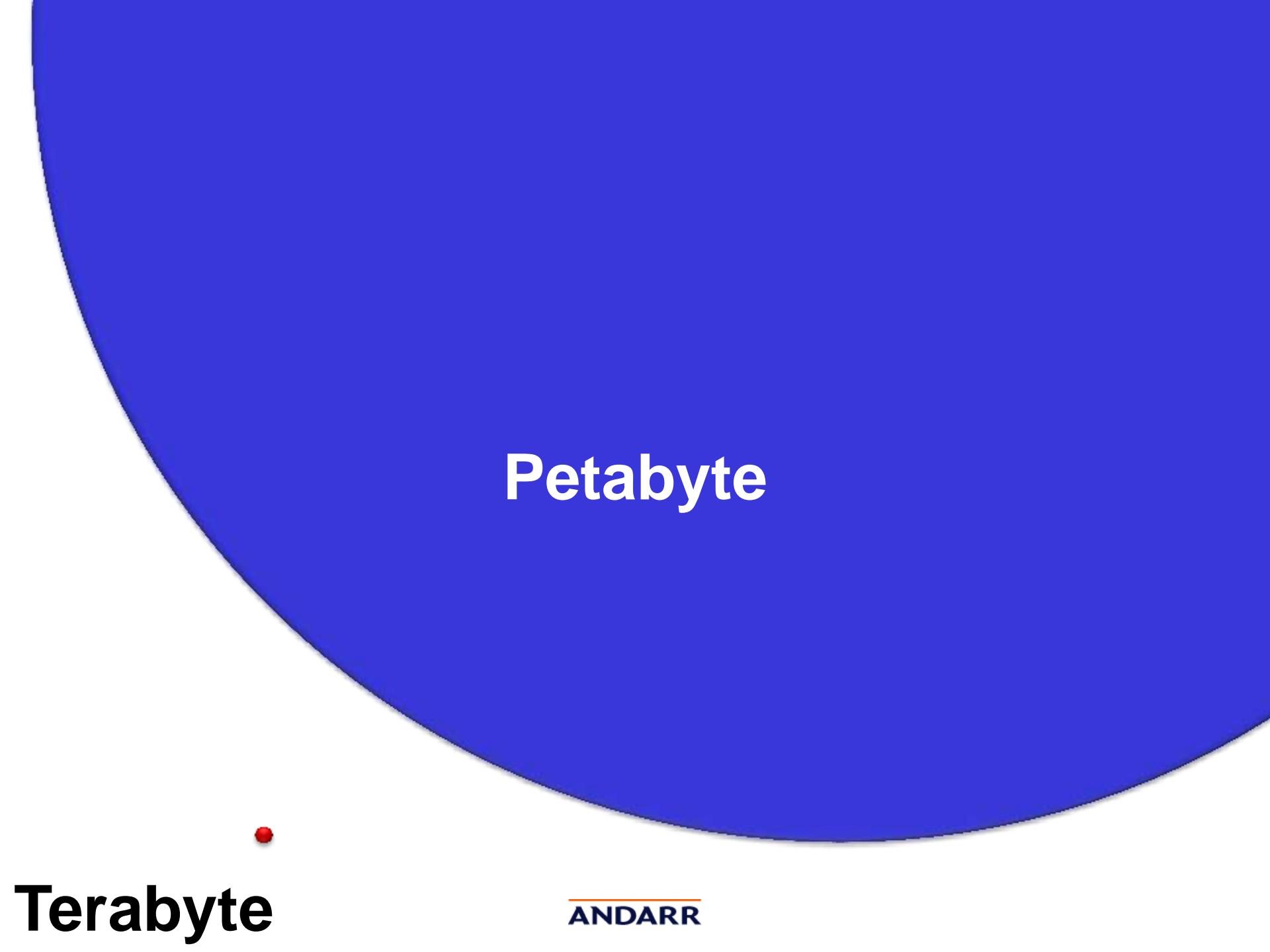
90% van de data  
is ongestructureerd

## Exponentiële groei data





<b>Kilobyte</b>	$10^3$ bytes
<b>Megabyte</b>	$10^6$ bytes
<b>Gigabyte</b>	$10^9$ bytes
<b>Terabyte</b>	$10^{12}$ bytes
<b>Petabyte</b>	$10^{15}$ bytes
<b>Exabyte</b>	$10^{18}$ bytes
<b>Zettabyte</b>	$10^{21}$ bytes
<b>Yottabyte</b>	$10^{24}$ bytes



Petabyte

Terabyte

ANDARR



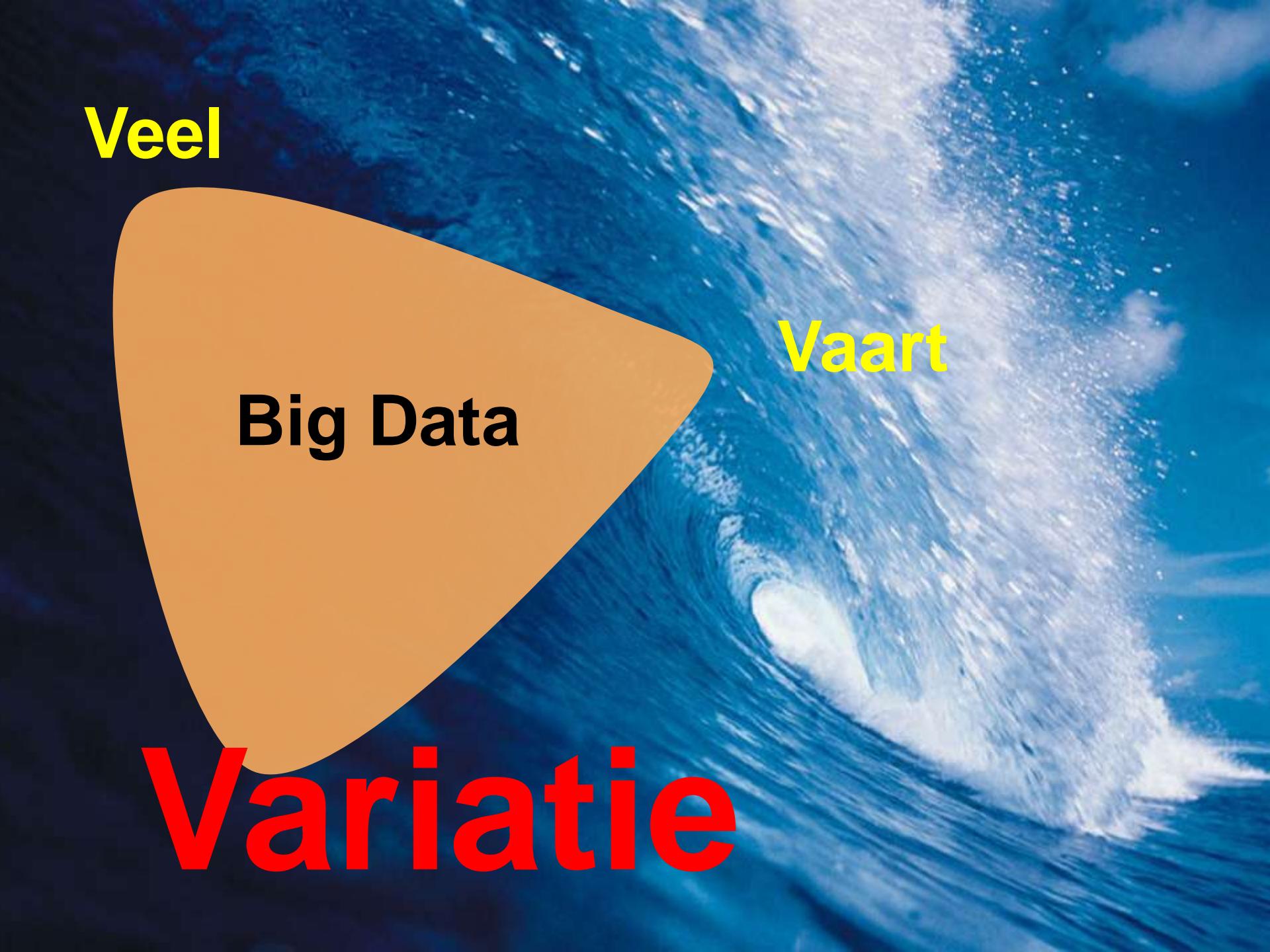
**Veel**

**Big Data**

**Variatie**



**Vaart**



**Veel**

**Big Data**

**Vaart**

**Variatie**

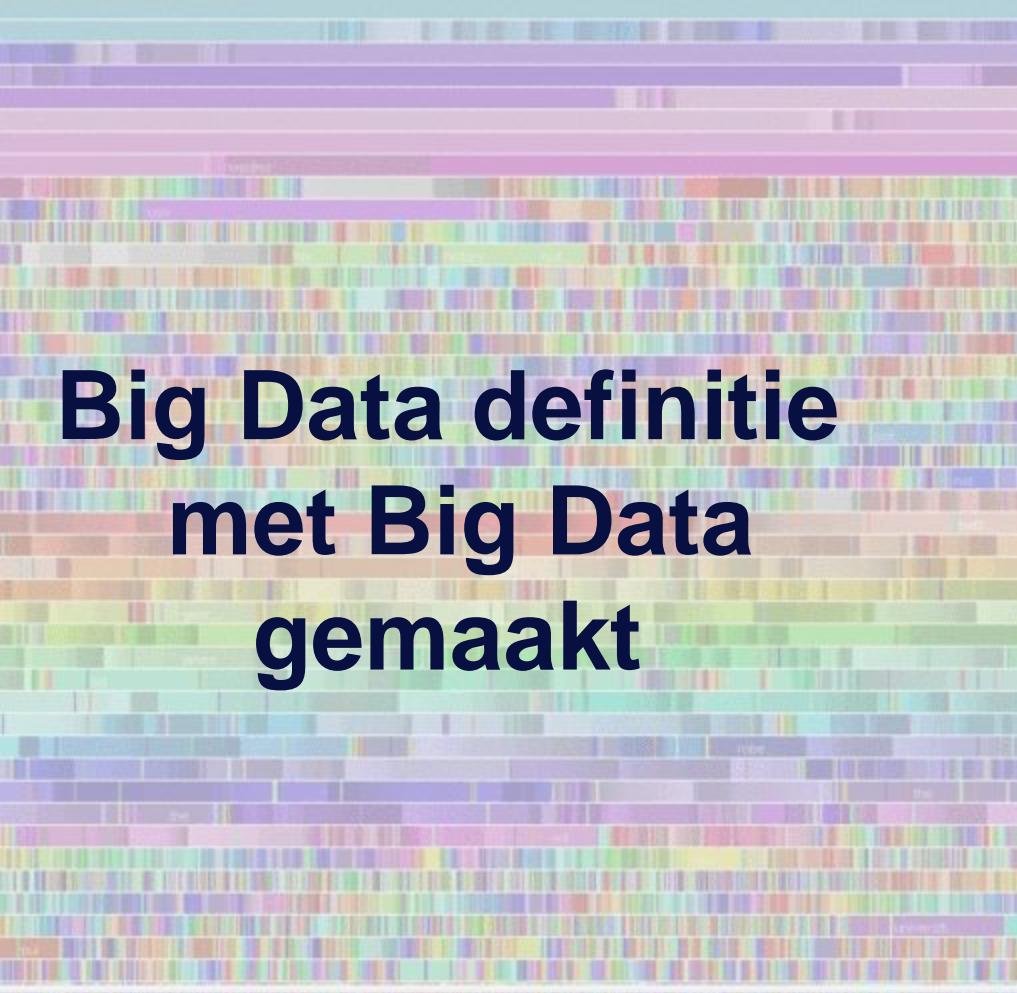


**Verzekeringsfraude vaststellen door automatische analyse van alle declaraties.**

**Aandelenwaarde voorspellen door sentiment analyse van twitter berichten.**

**Opzegging van telefoonabonnees voorkomen door alle belgegevens te verwerken.**

**Het profiel van de ‘zwevende’ kiezer kennen door alle relevante social media te gebruiken**



# **Big Data definitie met Big Data gemaakt**

**“opslag en analyse van grote of complexe data verzamelingen mogelijk gebruikmakend van NoSQL, MapReduce en machine learning”**

Date: 04-10-2013  
Source: Technology Review  
Subject: The Big Data Conundrum: How to Define It?

**Jonathan Stuart Ward  
and Adam Barker at  
the University of St  
Andrews in Scotland**

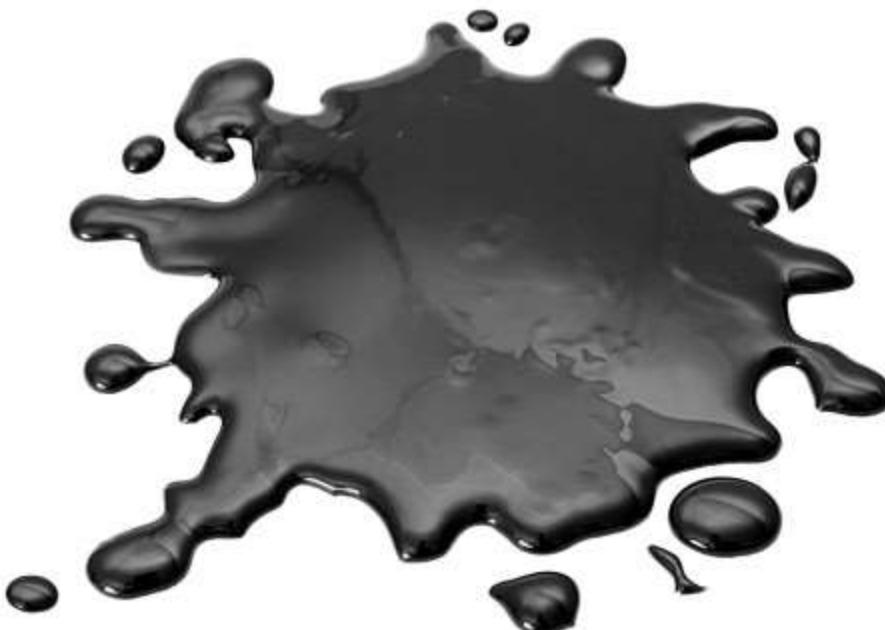


# Waarom nu een “Big Data” breekpunt?

- Logisch gevolg van digitalisering  
**(data communicatie en rekenkracht is goedkoop)**
- Algoritmen beter door veel data
- **Economische druk?**
- “kritische massa” aan data beschikbaar

A large, powerful ocean wave is shown crashing, with white spray and foam at the base. The water is a deep teal color, transitioning to darker blue at the top of the wave. The overall scene conveys a sense of energy and movement.

**Big Money**



**olie**

**ANDARR**



# synthetische drugs



# **Big Data Voorbeelden**



**25 items bepalen  
zwangerschap**

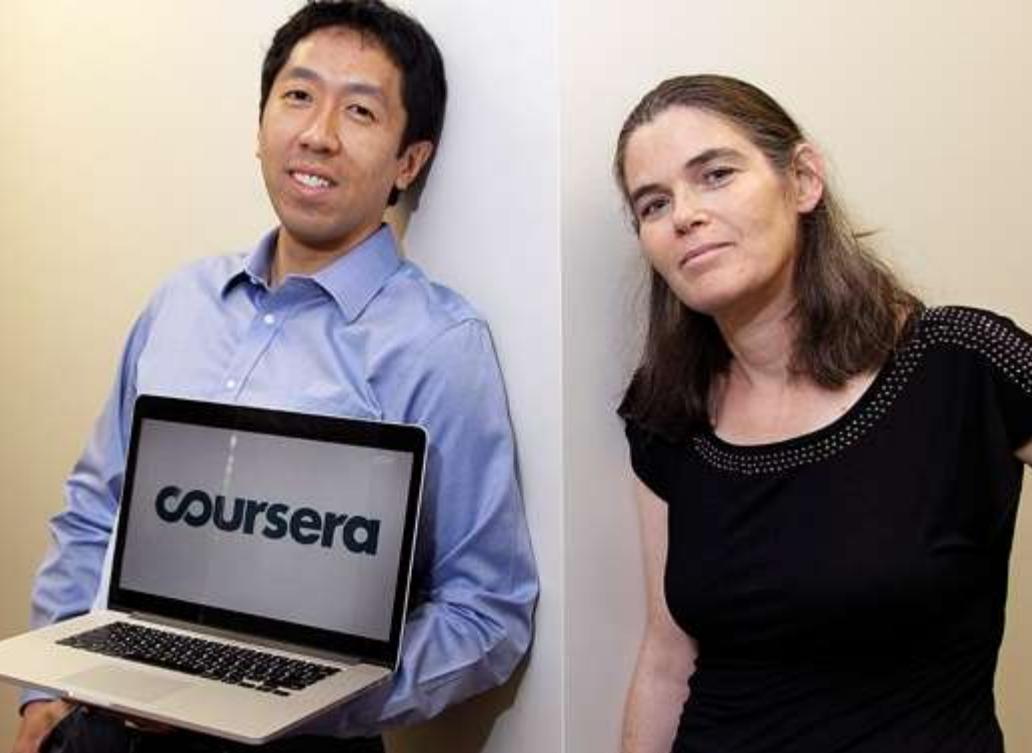
**Het voorspellen van  
koopgedrag of  
uitval onderdeel**

**10 terabytes elke  
30 minuten**



**...T...Systems**





Date: 05-11-2012  
Source: Technology Review  
Subject: The Technology of  
Massive Open Online Courses

**Vaststellen dat van de 2 miljoen studenten die een online cursus volgen er 200 studenten een denkfout maken**



**Voorspellen met waarschijnlijkheid van  
95% is dat een tweet binnen enkele  
uren trending topic wordt**



Massachusetts  
Institute of  
Technology

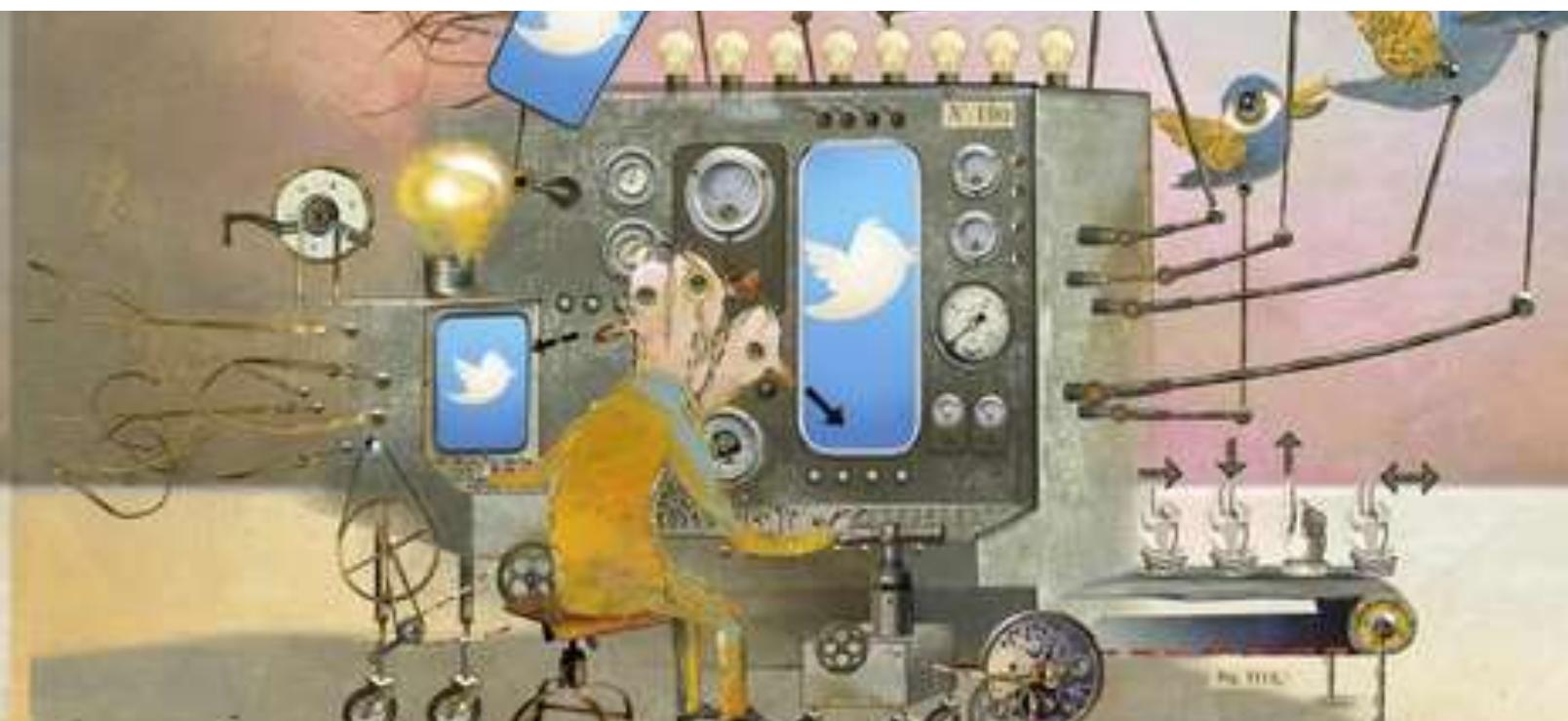
Date: 25-11-2013

Source: The New York Times

Subject: Inside a Twitter Robot Factory



**20 miljoen valse Twitter accounts, bijna 9% van de Twitter's actieve gebruikers  
Stroppa en De Micheli**



**Mr. Vidmar's robots have helped make his clients "trending topics" on Twitter**



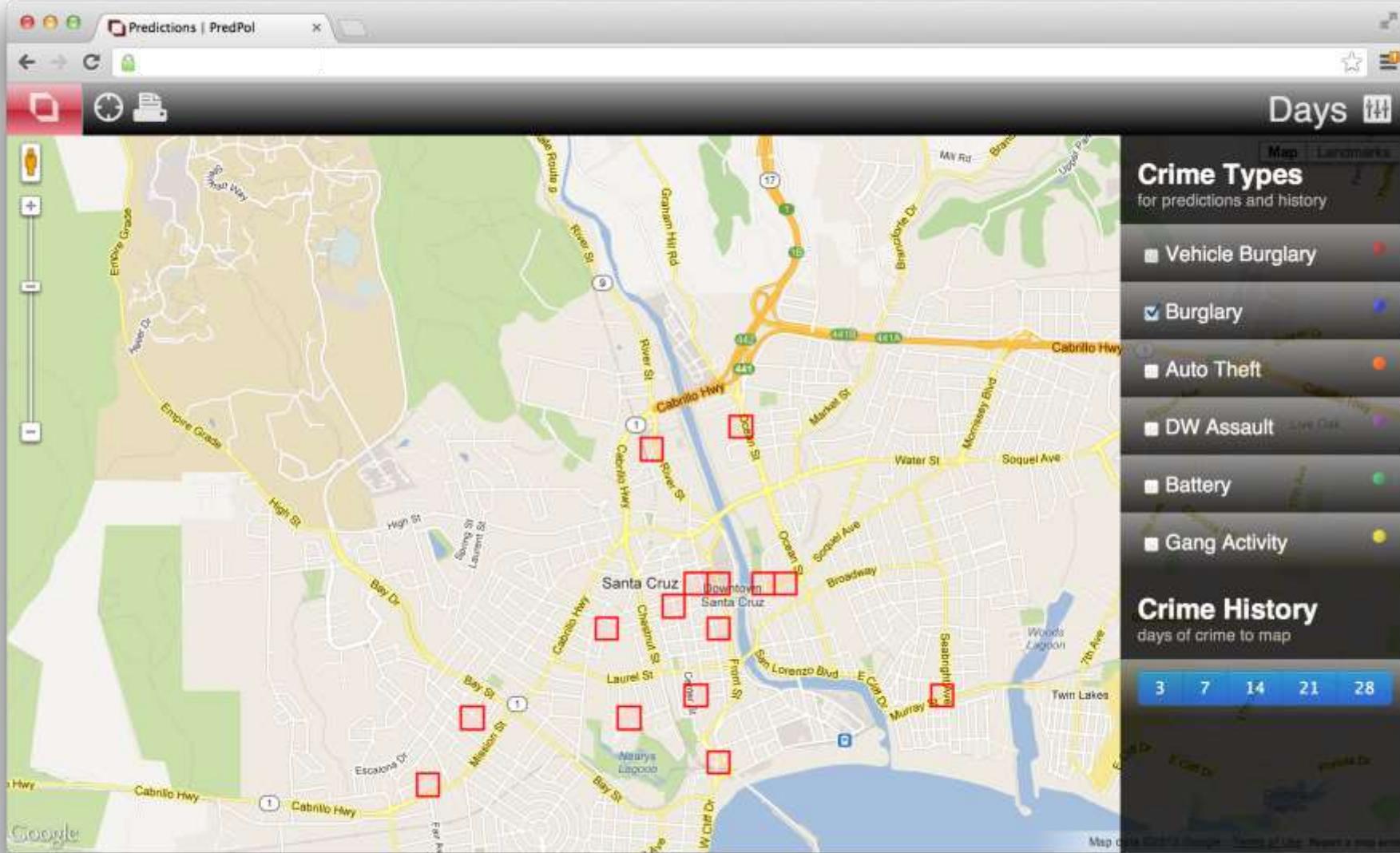


## Doorrekenen van 60 jaar van meteorologische gegevens om de kansen van extreme koude of hittegolven tot 40 dagen vooruit te voorspellen





# PREDICTIVE POLICING



# Customer Predictive Analytics

## Cloud based Analyse software

Bepaalt de kans dat een ‘lead’ een deal wordt

	Infer Score	Company	Name
<input type="checkbox"/>	88	7-Eleven , Inc.	Dremmel, Rick
<input type="checkbox"/>	87	Millipore	Flannery, David
<input type="checkbox"/>	81	Reebok International	Hearon, Fadel
<input type="checkbox"/>	79	Diageo North America	Walsh, Alicen
<input type="checkbox"/>	78	Allegis Group	Francis, Mollie
<input type="checkbox"/>	75	Arrowpoint Capital Corp.	Clinton, Gerardo
<input type="checkbox"/>	71	Erie Insurance Group	Lemmon, James
<input type="checkbox"/>	70	United Rentals	Spencer, Will
<input type="checkbox"/>	69	Airgas	Chbihna, Lori
<input type="checkbox"/>	68	Harman International	Whelans, Eduardo



Intensiteit klantcontact  
Interne medewerkers mobiliseren

CEO Vik Singh

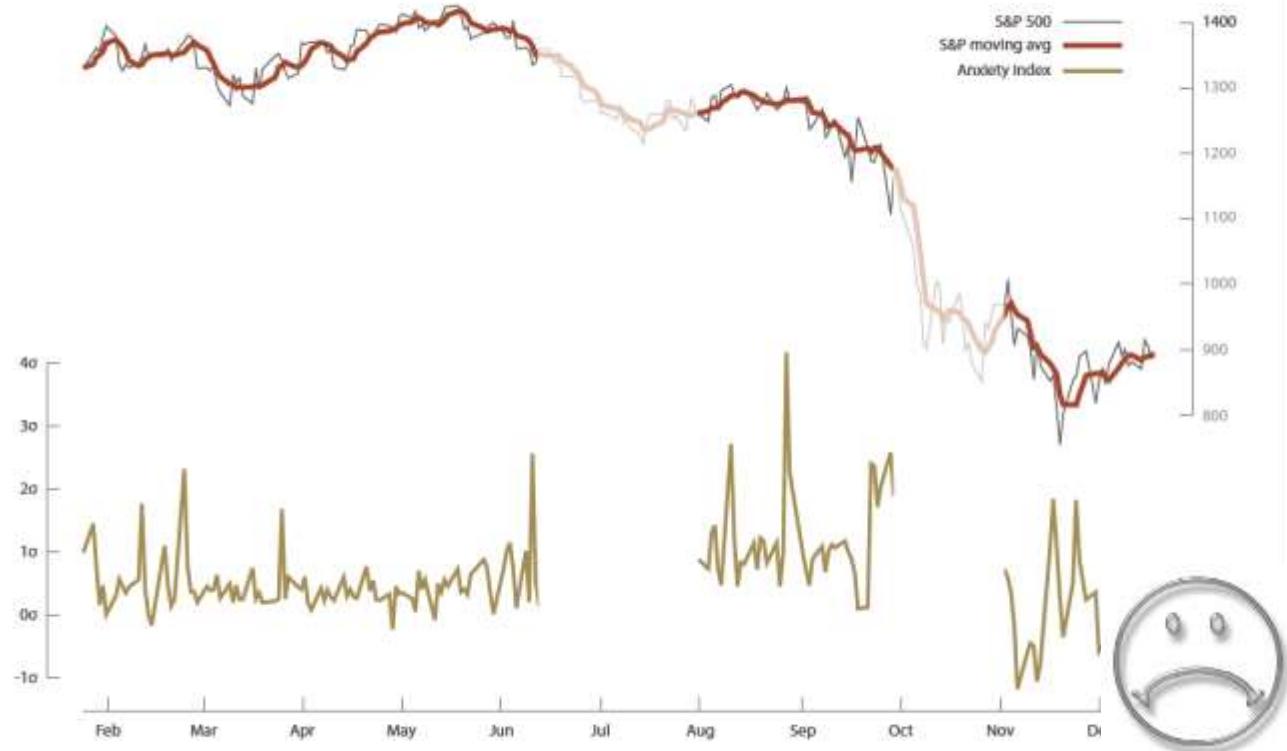
# Gilbert and Karahalios Finds Blogs Can Predict Future

06/21/2010 *New Scientist* magazine

## Zeitgeist meter – Anxiety Index

De S&P 500-index iets lager dan de voorspellingen van het einde van de driedaagse aandelen koers als angst index piekt.

20 miljoen blogs uit de LiveJournal website om een index van de Amerikaanse nationale gemoedtoestand te creëren.

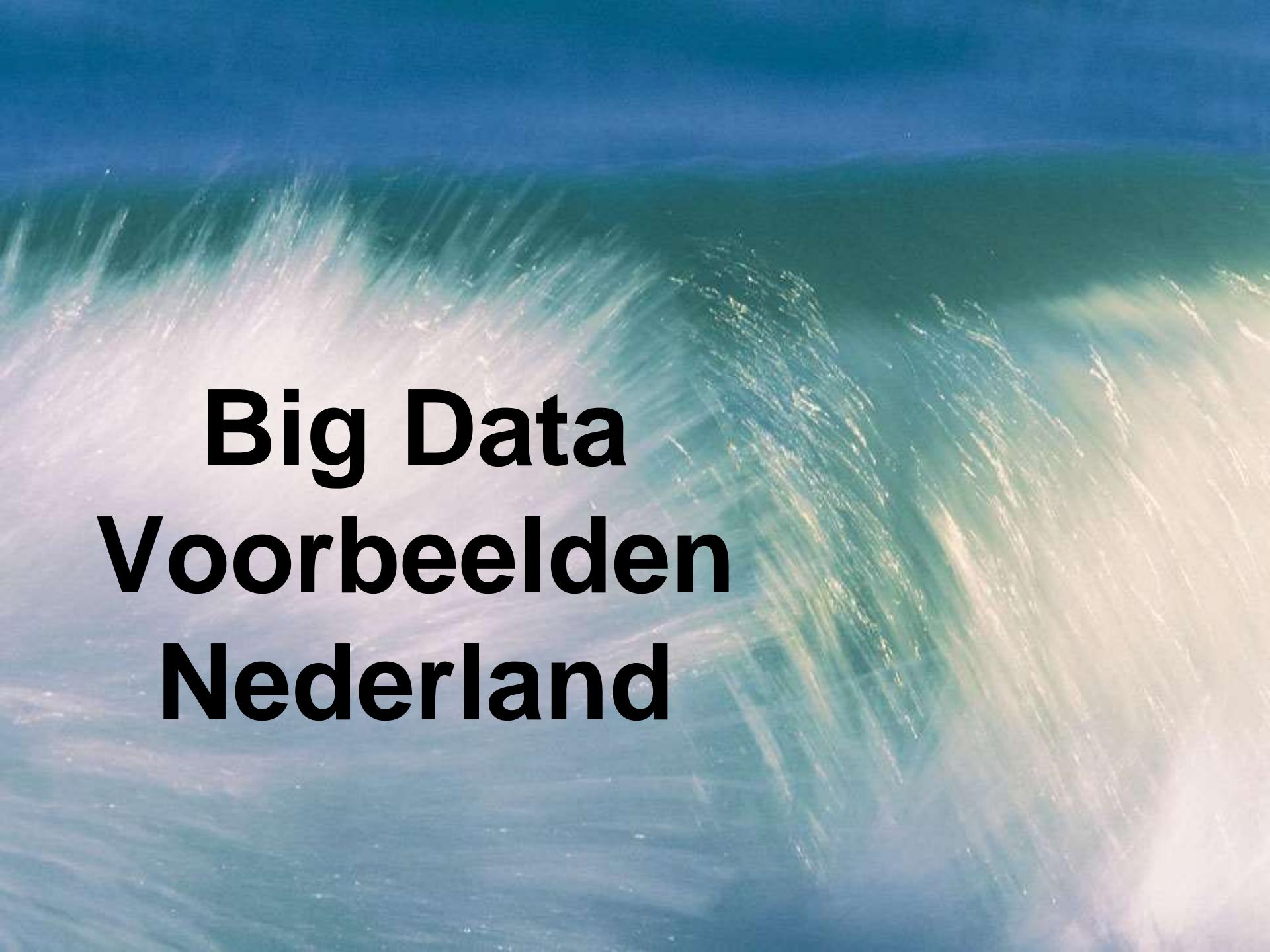




## Vincent van Leeuwen

12  
ersvoorspeller wint  
pitchwedstrijd





# **Big Data Voorbeelden Nederland**



# French campings near water : 689 records

data set contains websites of French campings near a beach or a lake.

## Marc Noët

[www.camping-le-mas-de-la-plage.fr](http://www.camping-le-mas-de-la-plage.fr)

[www.palmirabeach.fr](http://www.palmirabeach.fr)

[www.mairie-telgruc.fr](http://www.mairie-telgruc.fr)

[www.camping-lesdunes.fr](http://www.camping-lesdunes.fr)

[www.campingdespins.fr](http://www.campingdespins.fr)



## dataprovder





Richard Verhoeff



**Nederlands bedrijf gebruikt de nieuwe  
online datadienst van Google BigQuery  
voor Center Parcs**

**crystalloids**  
crystallizing your predictive enterprise

Google bigquery

# Digitale Delta: Slimmere informatie-uitwisseling binnen de watersector moet leiden tot lagere kosten, innovatie en sterkere internationale positie.

25-06-2013 | Rijkswaterstaat Water, Verkeer en Leefomgeving | persbericht

**real-time data** van waterafvoersystemen en waterstanden, te combineren met meteorologische data en **simulatiemodellen**, maakt dat we wateropslag beter kunnen monitoren

Er zijn gegevens over regenval, waterstanden, waterkwaliteit, dijksensoren, radardata, **voorspellingsmodellen** en gegevens van sluizen, gemalen en dammen

Ongelimiteerd werken met data in de watersector

Een Top sector partnership met IBM, TU Delft, Deltares, Delfland en Rijkswaterstaat





## Data Platform

### Data Warehouse



- >700GB Compressed / >2.5TB Uncompressed
- Largest tables load >50M records/day

### MapReduce

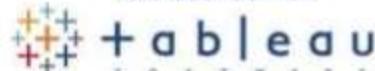


massive data - minimal code

- >150M events/day
- >500M events/day by Q1 2013

## Reporting

### Tableau Server



- Daily, Weekly, and Monthly Push Reports
- >192 views, 5.5 GB of extracts and data sources
- 100+ sessions per day

## Analytics

### Tableau Desktop plus R



- Five desktop users of Tableau
- Driving company forecast process and planning
- Analysis forms backbone of new strategies

**Rob Winters**

**ANDARR**



**Laatste 2 jaar ontvingen  
119 database software  
leveranciers \$1.17 miljard  
venture-capital**

Dow Jones VentureSource



# **Big Data Voorbeelden**



Date: 29-06-2012

Source: The Wall Street Journal

Subject: Your E-Book Is  
Reading You

Digital-book publishers and  
retailers now know more about  
their readers than ever before.



Can identify which passages of  
digital books are popular with  
readers

Acquiring Goodreads give insight into  
the behavior of book influencers



# Patrick Meier



Japanese earthquake and tsunami in 2011, more than 300,000 tweets every minute

developing Twitter “classifiers,” algorithms identify relevant tweets during crises, with accuracy 70 to 90 %

The background of the image is a close-up, high-angle shot of a massive, curling ocean wave. The wave's face is a vibrant greenish-blue, with white spray and foam visible where it is breaking. The water behind the wave is a darker, deeper blue. The overall texture is dynamic and suggests a strong current or wind.

# Datameesters



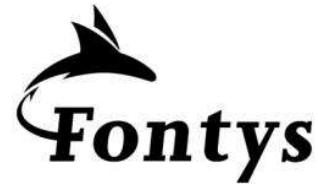
**2015**

**4.4 miljoen experts nodig**

**Te kort**

# Archetype

## Hadoop Hacker & Data Detective



**Bedrijfsprocessen**

**Applicaties (hive, pig,..)**

**Hadoop (data OS)**

**Hardware**



# Data Detective

Bedrijfsprocessen

Applicaties (hive, pig,..)

Hadoop (data OS)

Hardware

# Data Detective

Bedrijfsprocessen

Applicaties (hive, pig,..)

Hadoop (data OS)

Hardware

**Hadoop Hacker**

# Hadoop Hacker

Ontwikkelaar en engineer

Bouwen en onderhouden met Big Data gereedschappen en methoden

*Onderzoeken , Ontdekken en Onderhouden*

# Data Detective

Domein expert en analist

Vertalen business vraag naar Big Data vraag

*Ongearticuleerde behoefte omzetten in specifieke vragen...*

# Hadoop Hacker

- Gedegen ervaring met Hadoop omgeving en enige No SQL systemen
- Algoritme-kennis data mining, machine learning, data visualisatie, en relevante software engineering
- Kennis van Java en Linux
- Kennis van virtualisatie en gedistribueerde systemen (ICT clusters)

- Gedegen begrip van dienstontwikkeling processen in specifiek marktdomein
- Data journalistiek
- Ervaring en kennis van statistiek en statistische software (zoals SPSS, SAS of R).
- Redelijk begrip data mining, machine learning, data visualisatie en relevante software engineering
- Conceptuele kennis van Hadoop en andere Big Data technologieën en methoden (No SQL)



# Data Detective

# Hadoop Hacker & Data Detective

## balans in functies:



**Data scientist**

**Data architect**

**Data visualiseerder**

**Data organisatie veranderaar**

**Data engineer/operator**

**Data steward**

**Data virtualization/cloud specialist**

**Data analist**

**Data .....**

**Data .....**

**Data .....**



**Big Data  
Schuldigen**





Date: 01-11-2012

Source: The Economist

Subject: Deus ex machina

## Voters are being targeted in new and powerful ways



Voter-registration files have been merged with vast quantities of bought consumer data, on top of which come bought or acquired e-mails, mobile and landline numbers, as well as data gathered through canvassing, phone banks and social-media pages. The campaigns are also making use of cookies, the crumbs of data people leave behind when they browse the net.

Date: 07-06-2013

Source: Associated Press

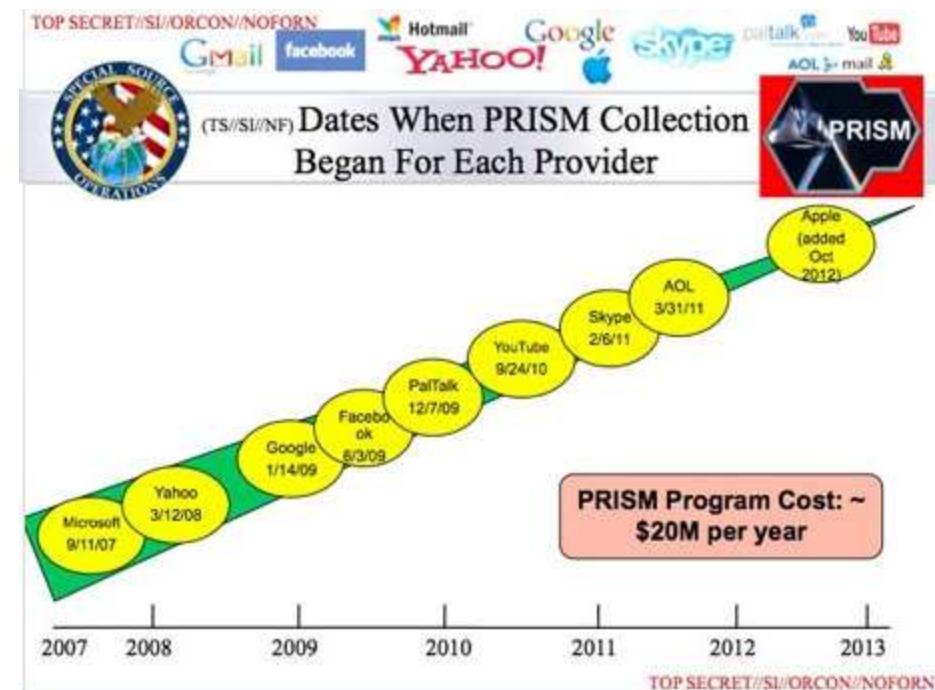
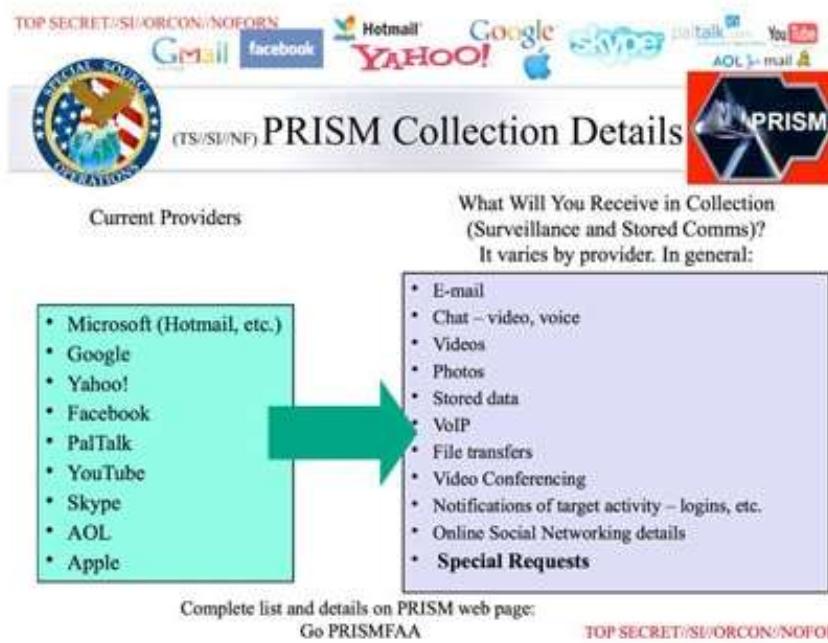
Subject: Is Big Data Turning Government Into Big Brother?

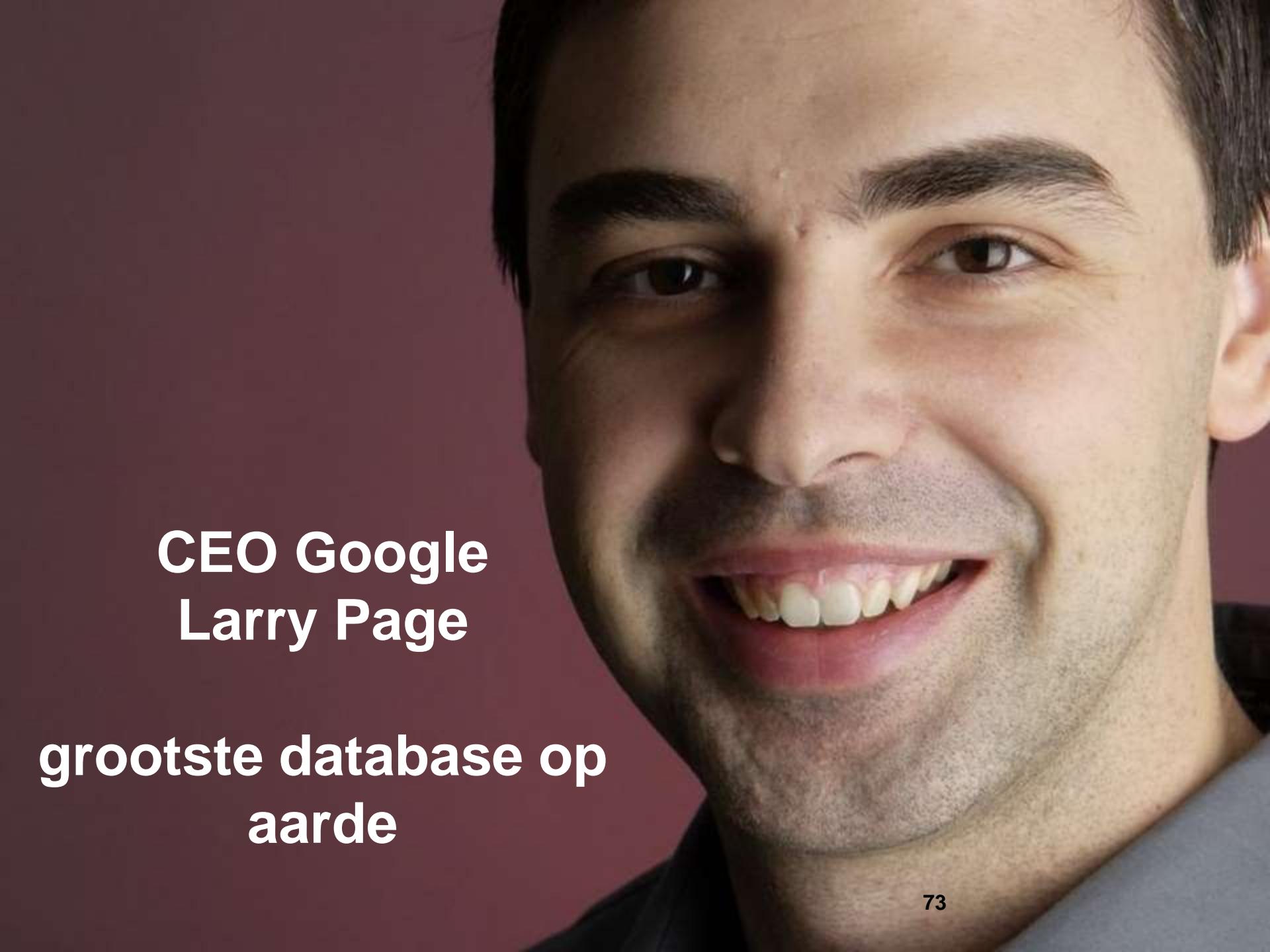


Date: 07-06-2013

Source: The Guardian

Subject: NSA PRISM program taps in to user data of Facebook, Yahoo and others



A close-up, slightly blurred portrait of Larry Page, the CEO of Google. He is looking directly at the camera with a warm, friendly smile. His dark hair is visible at the top right. The background is a solid, dark reddish-brown color.

**CEO Google  
Larry Page**

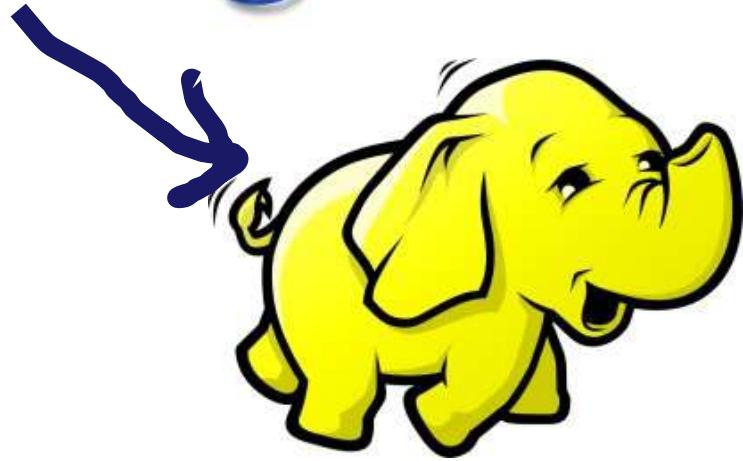
**grootste database op  
aarde**

Google™



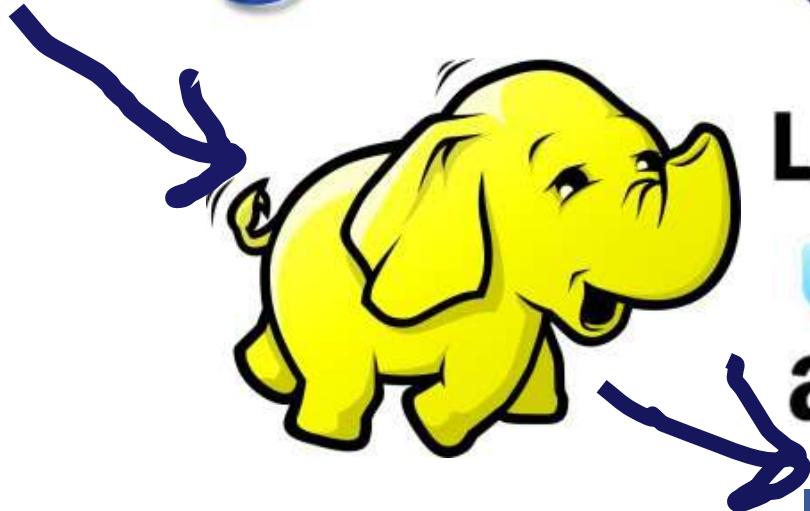
**ANDARR**

Google™



**ANDARR**

Google™



LinkedIn

twitter

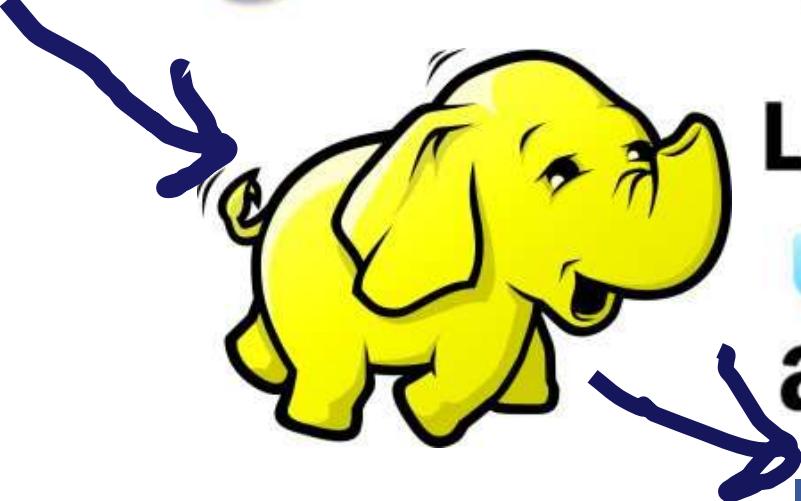
amazon®

facebook

YAHOO!

cloudera

The Hortonworks logo, which consists of three green elephant silhouettes walking in a line, followed by the word 'Hortonworks' in a bold, black, sans-serif font.



Google™



LinkedIn

twitter

amazon®

facebook.

YAHOO!

cloudera



Hortonworks

TERADATA

EMC<sup>2</sup>

IBM

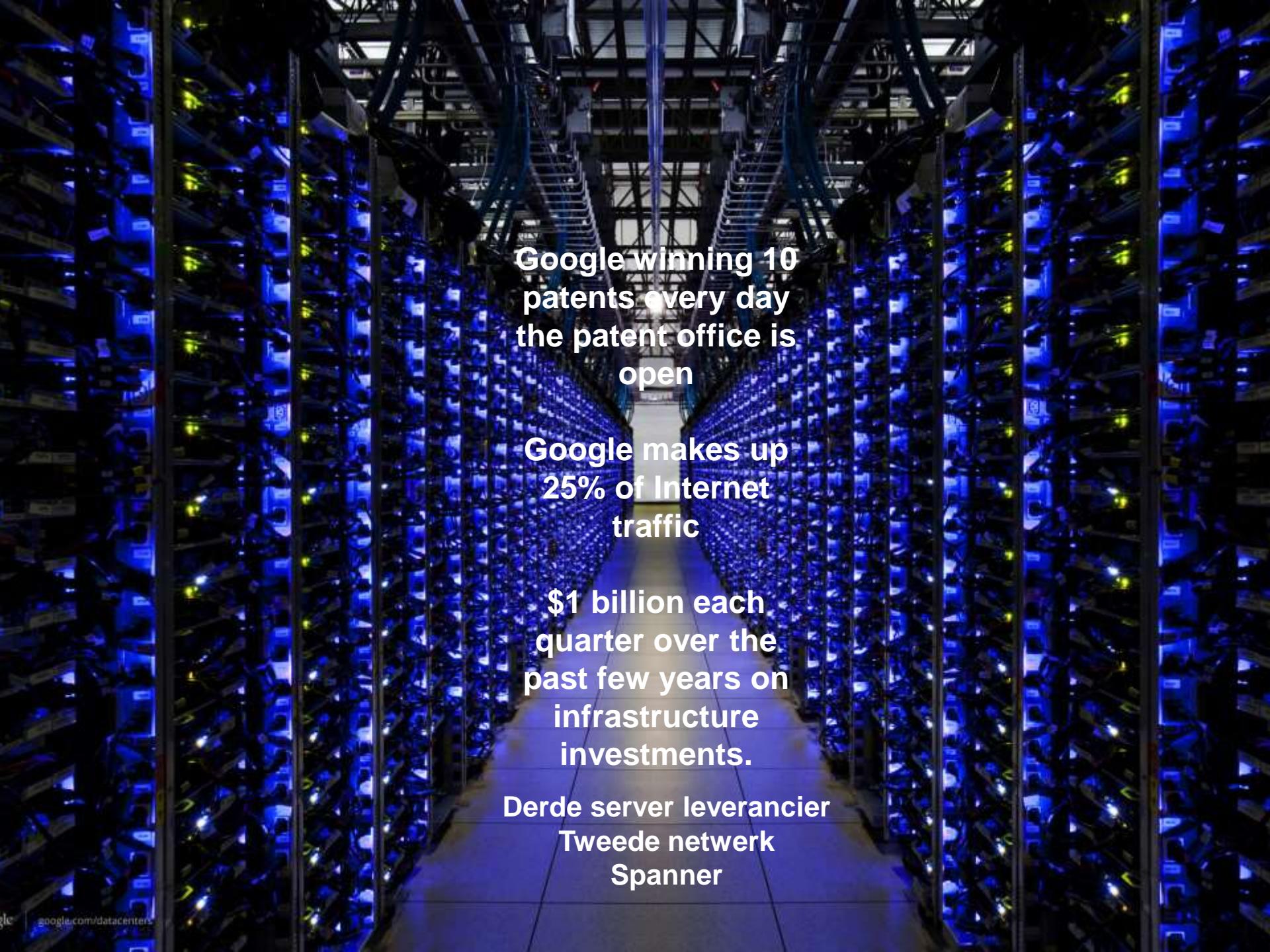
ORACLE®



Microsoft

A large, powerful ocean wave crashing, symbolizing the volume and power of big data.

# **Big Data gereedschappen**



**Google winning 10  
patents every day  
the patent office is  
open**

**Google makes up  
25% of Internet  
traffic**

**\$1 billion each  
quarter over the  
past few years on  
infrastructure  
investments.**

**Derde server leverancier  
Tweede netwerk  
Spanner**

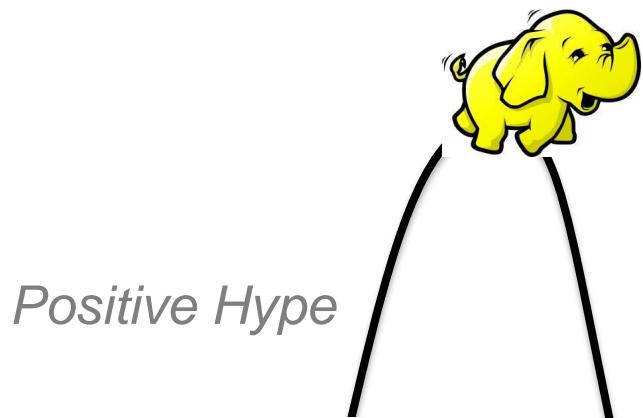


## Twee smaken

**Open Source omgeving gebaseerd op Hadoop**  
(omgeving meer voor Hadoop Hacker)

**Gesloten Cloud Omgeving: Google Big Query**  
(omgeving meer voor Data Detective)

**Peak of Inflated  
Expectations**



**Visualisatie**

**Slope of  
Enlightenment**

**Plateau of  
Productivity**

**Trough of  
Disillusionment**

**Innovation  
Trigger**

**ANDARR**

**Gartner**



# **Big Data Visualisatie**



# **“Big Data” visualization advanced data visualization Large Scale Visual Analytics**

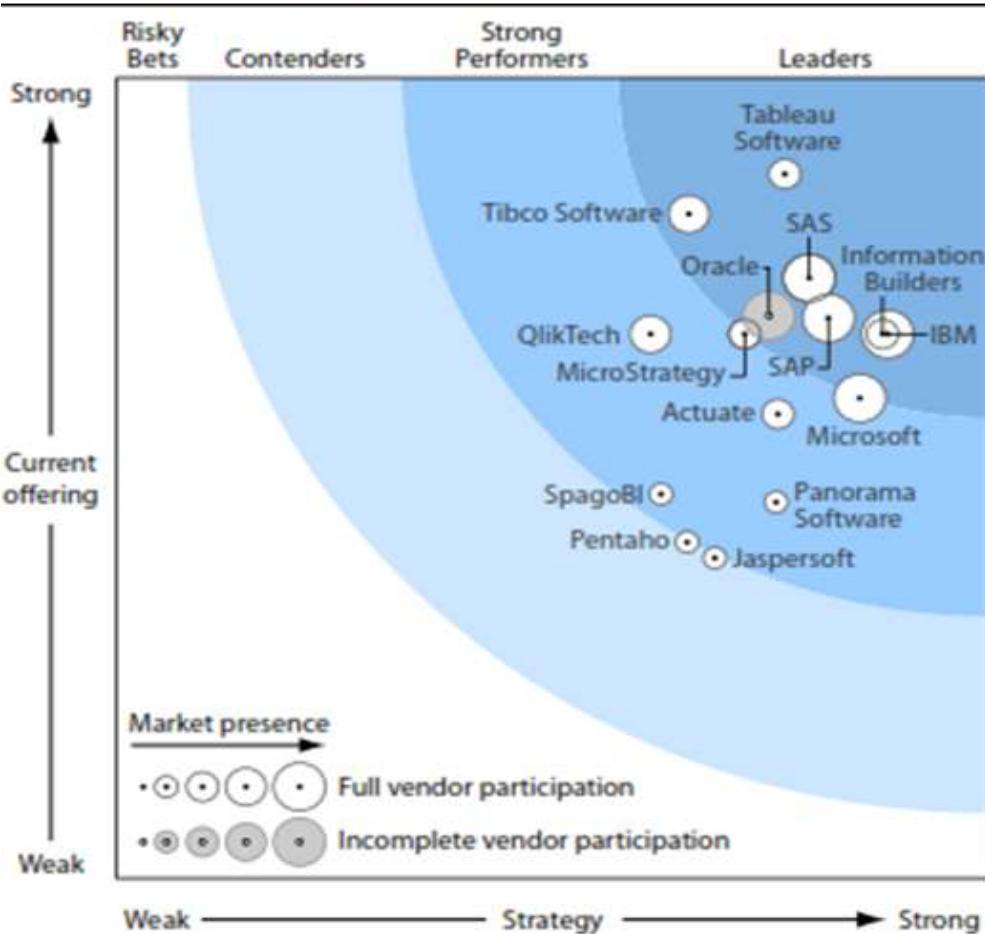
**Verdeel en heers ...**

**Meer..**

- interactief
- doorlopen en inzoomen
- visual querying
- aggregatie
- crowd sourcing
- ...

## Advanced Data Visualization Platforms, Q3 2012

by Boris Evelson and Noel Yuhanna





# ICT & Media Design

## Big Data visualisatie

### Waarom?

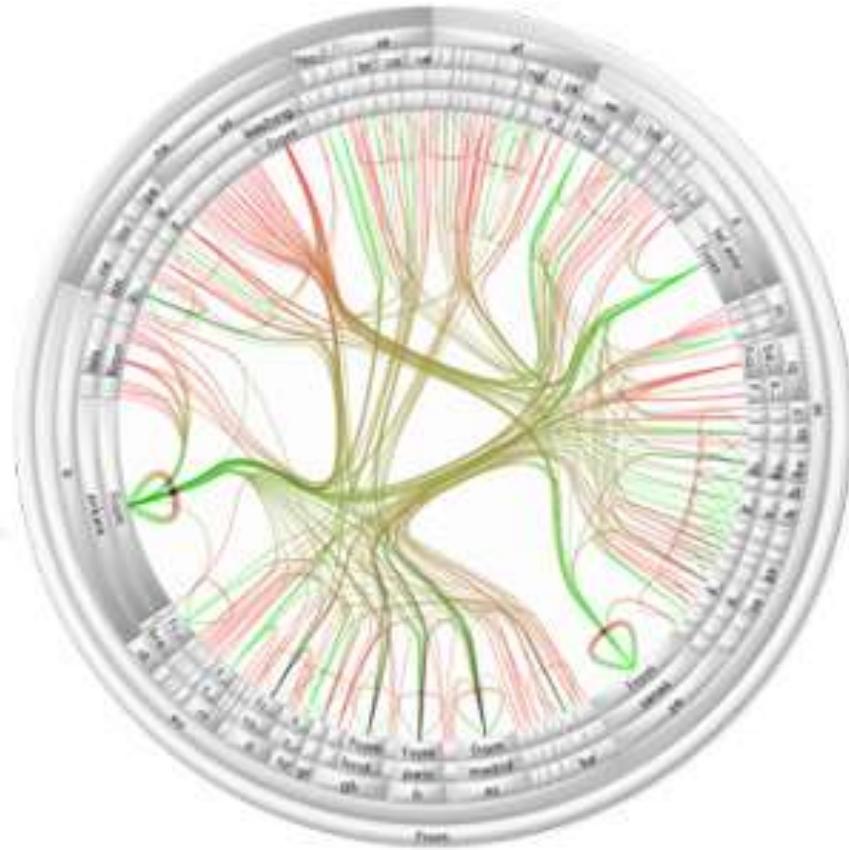
- Communicatie
- Analyse en probleem oplossen

### Wat?

- Informatie en inzicht halen uit grote hoeveelheden dynamische en onzuivere data
- Lever een tijdige, correcte en begrijpbare interpretatie
- De communicatie van de interpretatie leidt tot een juiste actie
- Kunst (invloed van esthetiek op beslissing)



**synerscope**  
connecting the dots



**Schermgrootte te beperkt voor  
zeer grote data sets**



## Interaction Mode

- Auto-Expand
- Highlight
- Selection

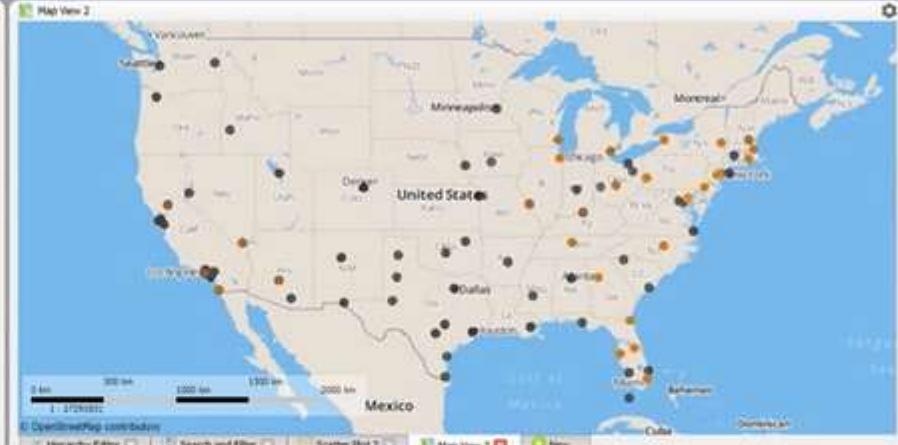
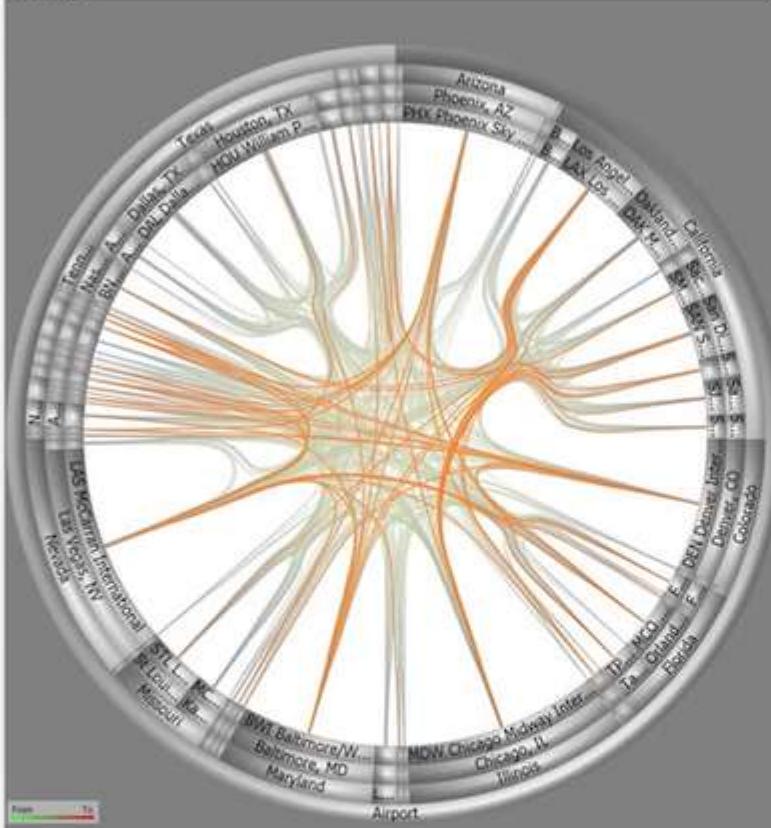
## Selection

- 3C q Clear Selection
- MB + E Invert Selection
- Expand Links to Nodes
- ↔ Expand Nodes to Links

## Drilling

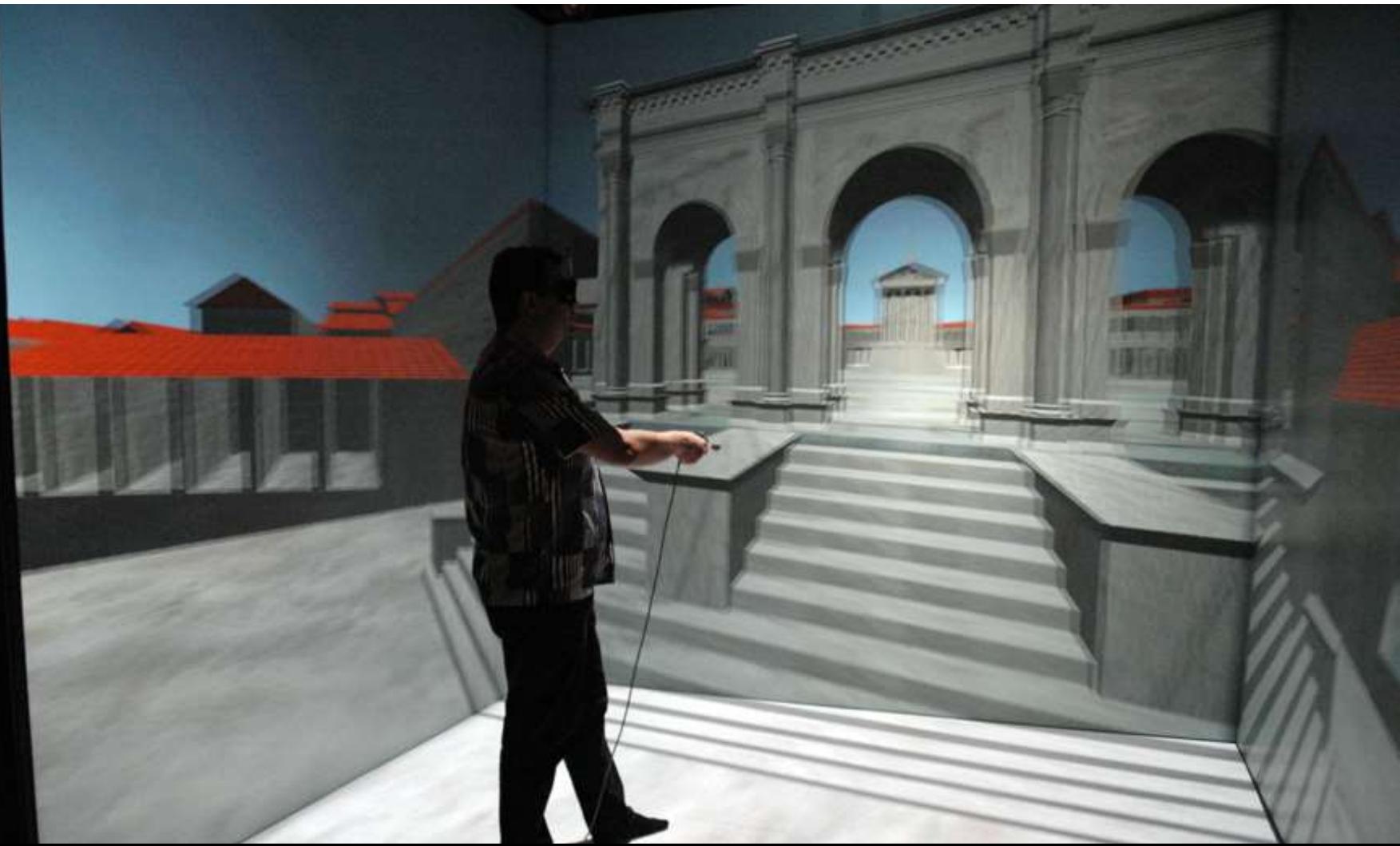
- Drill up to Overview
- Drill down to Selection

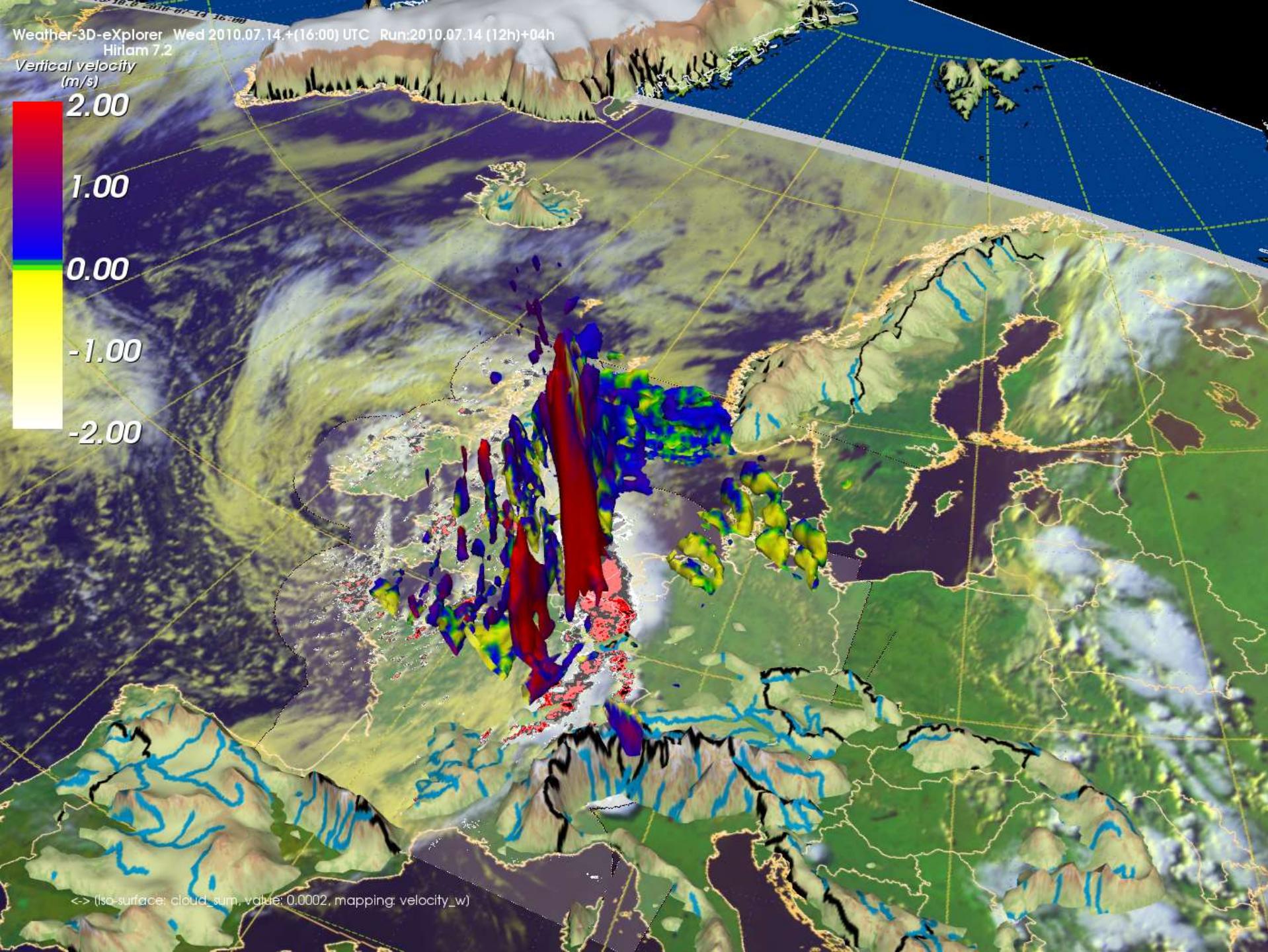
- Drill down to Constraints
- Reset Constraints



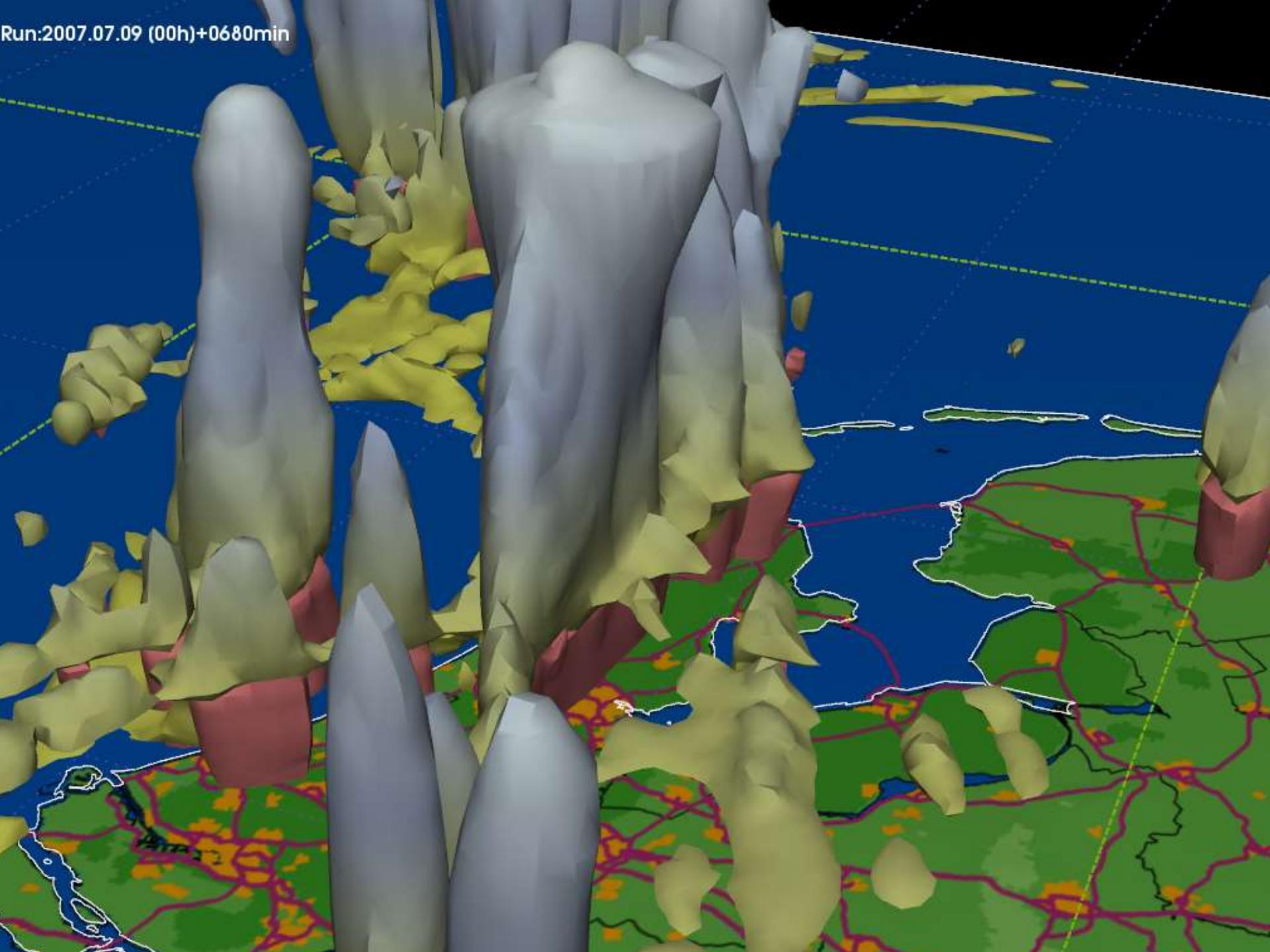
03:47

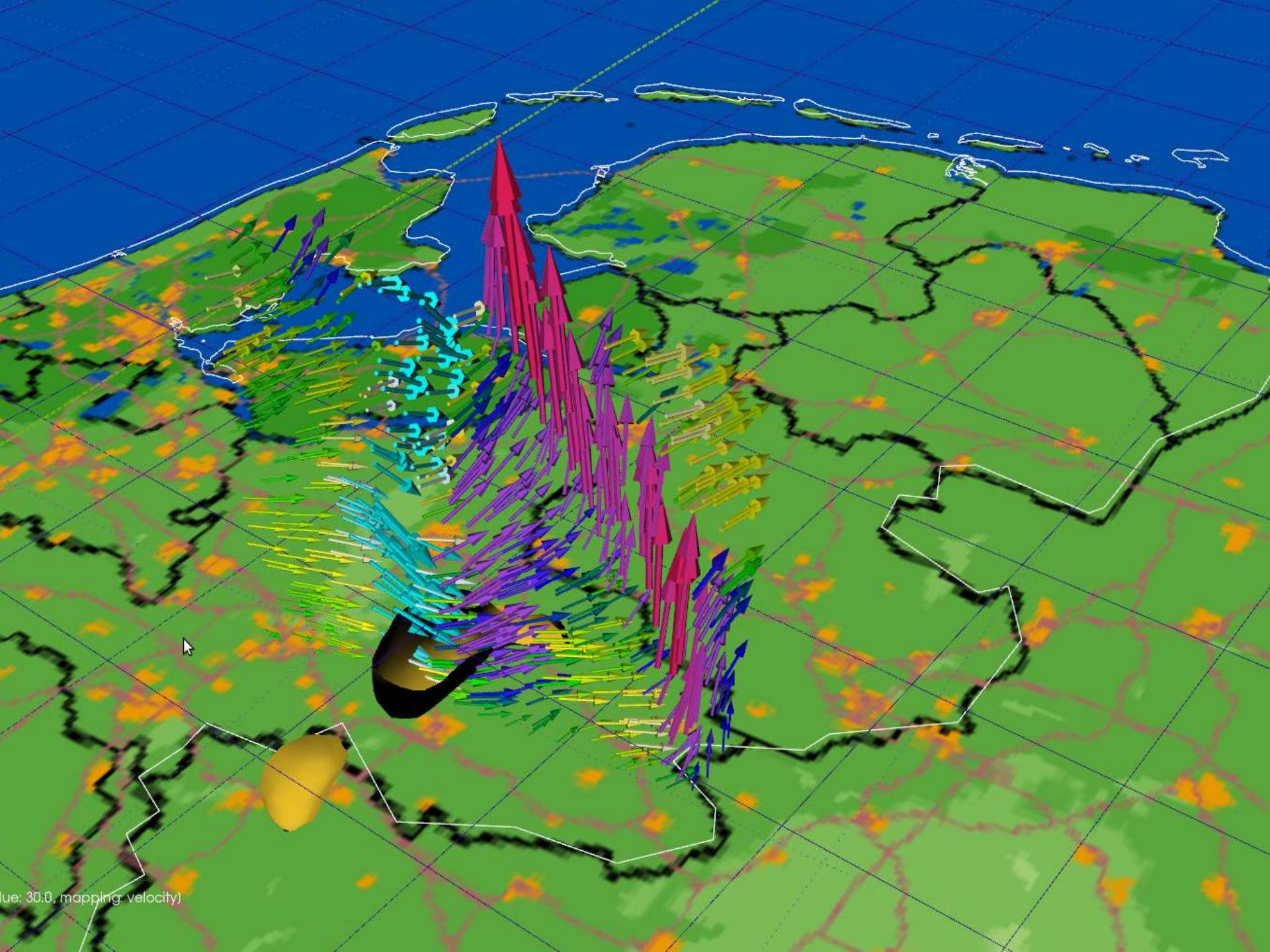
HD





Run:2007.07.09 (00h)+0680min





Weather-3D-eXplorer Tue 2010.08.03.+(10:00) UTC Run:2010.08.03 (06h)+04h  
Hirlam 7.2

Velocity (m/s)

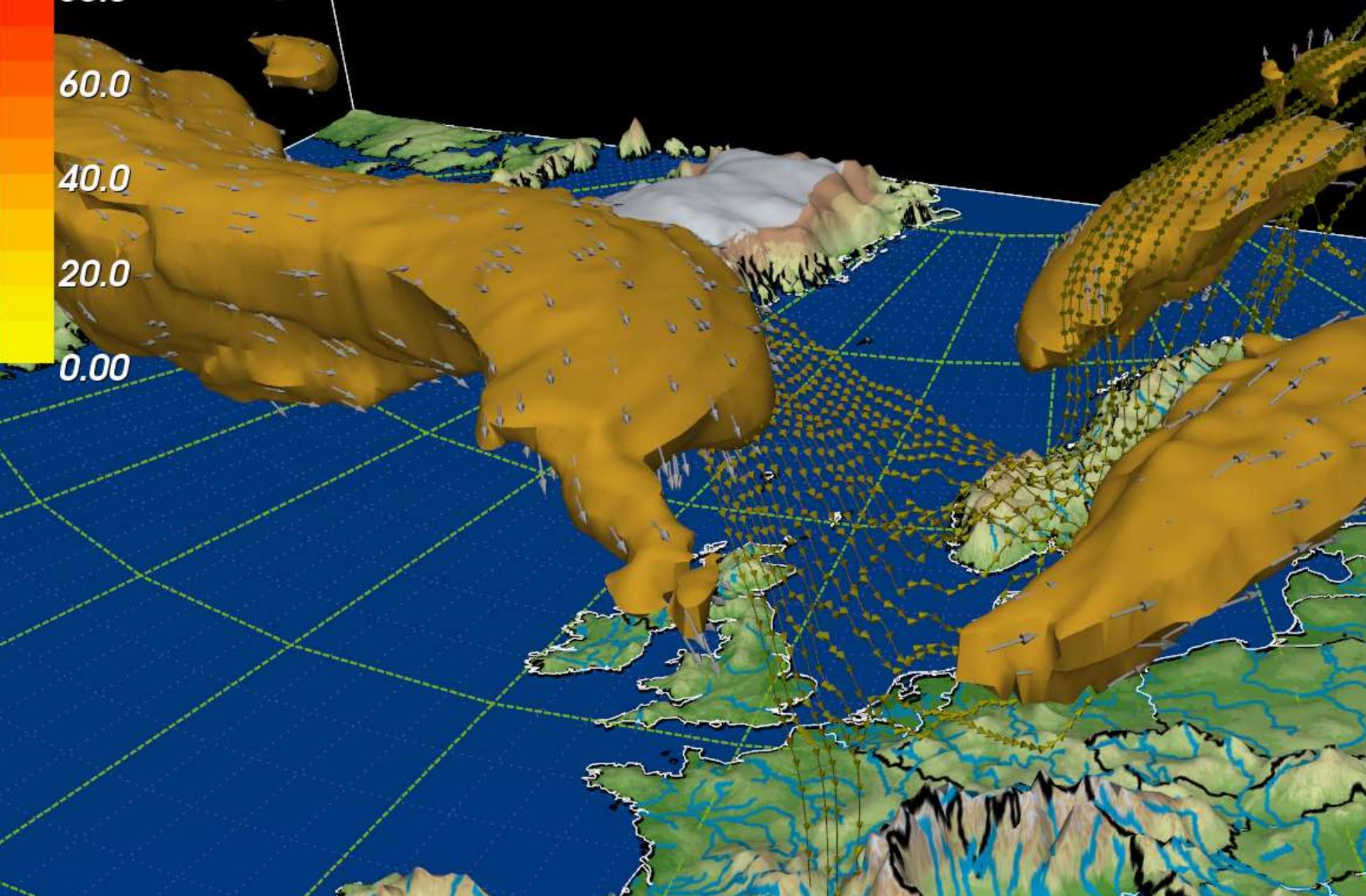
80.0

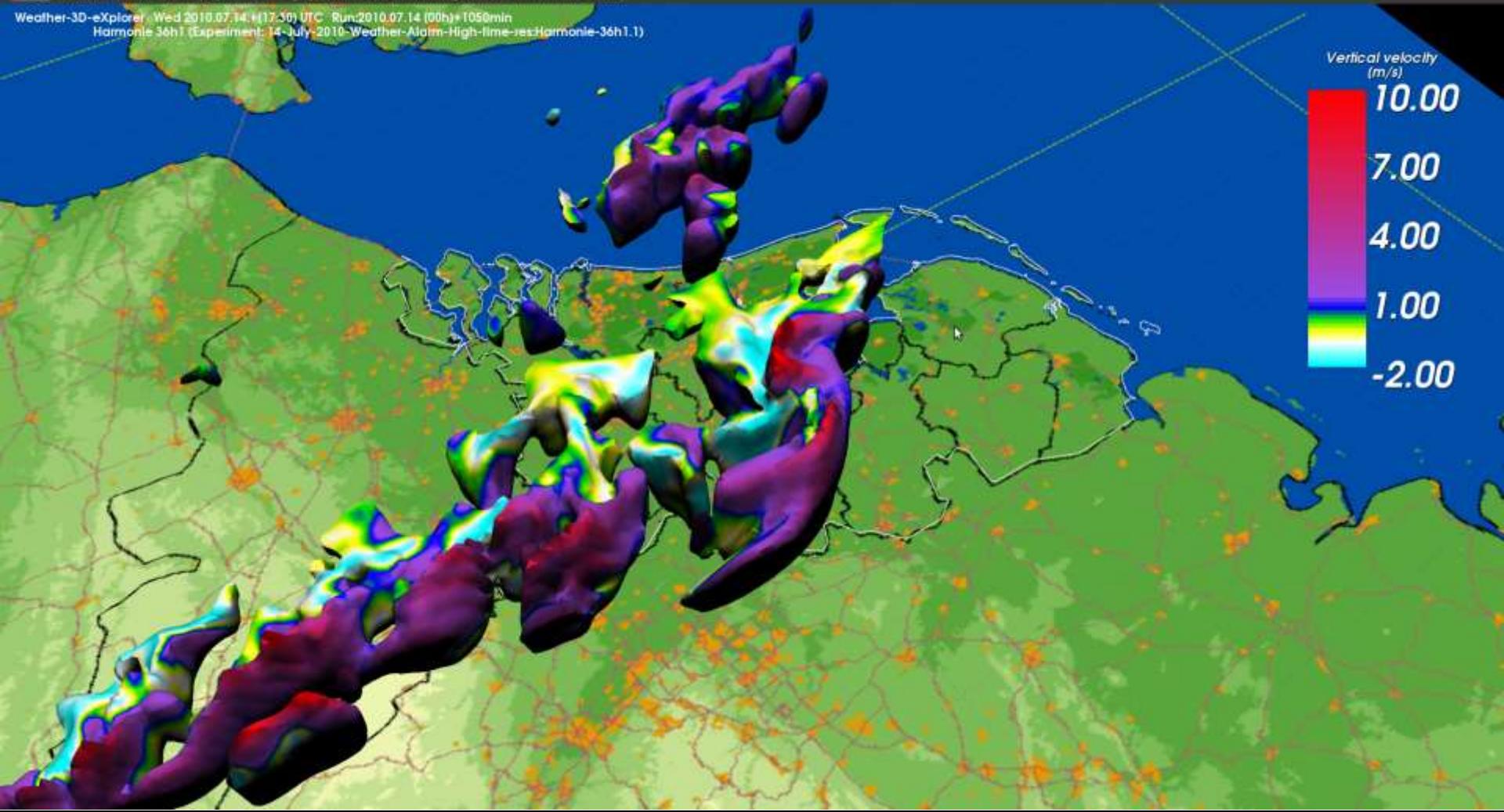
60.0

40.0

20.0

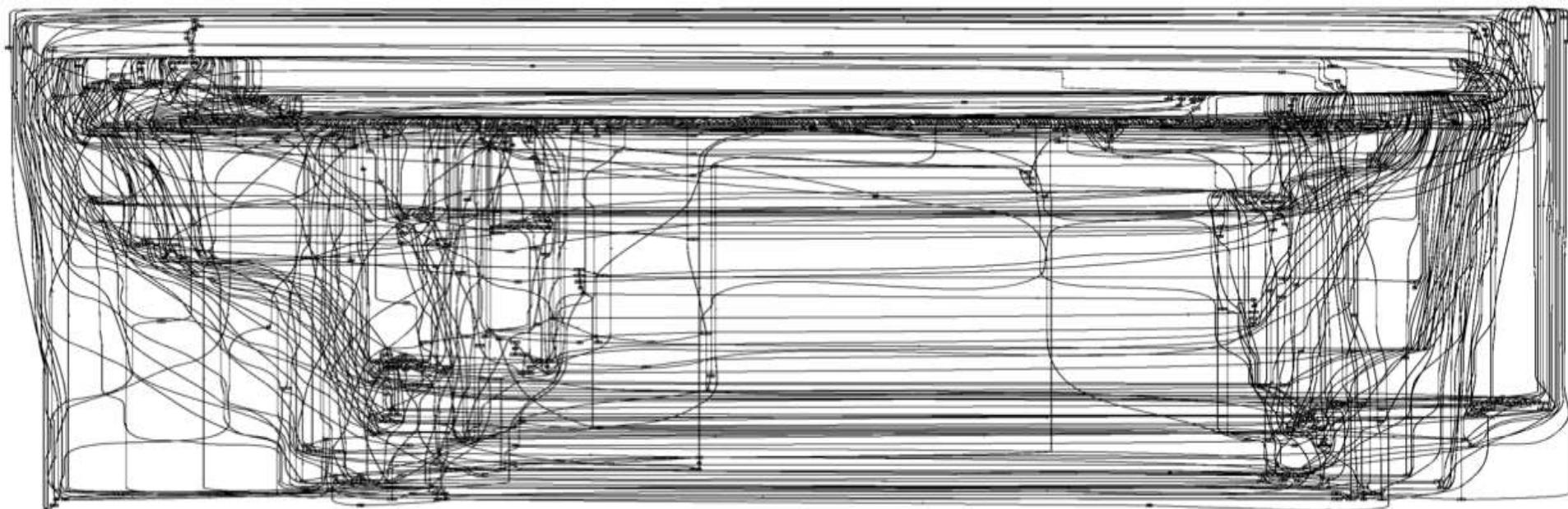
0.00





3D visualisatie weermodel M.Koutek KNMI  
In de CAVE van Sjaak Verwaaijen

# PhD working on Process Mining in SAP

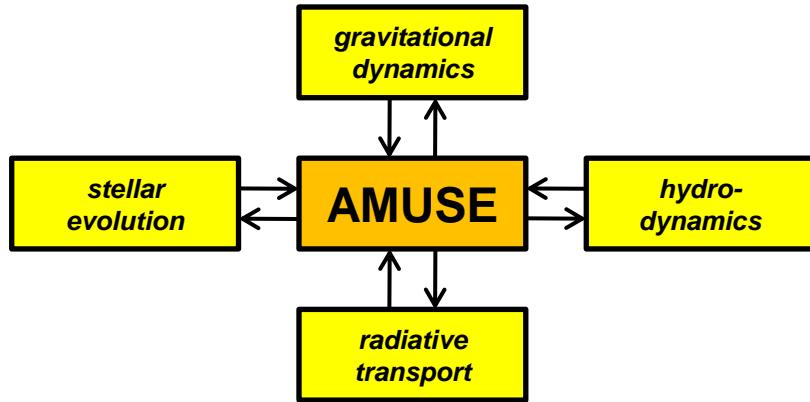


Prof.dr.ir. Wil van der Aalst

ANDARR

**TU/e**

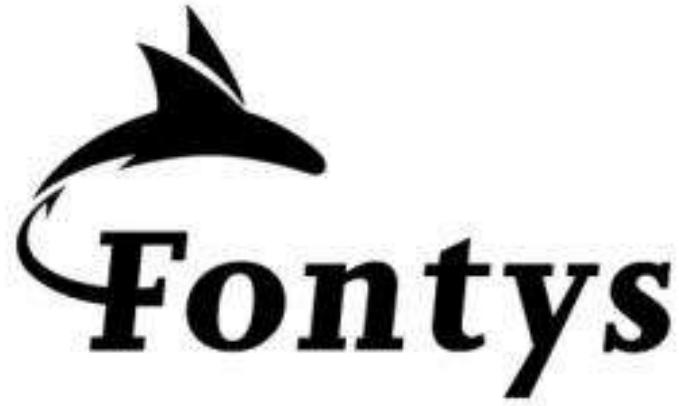
# Big Data Processing: Jungle Computing



## Computational Astrophysics

The background of the image is a close-up, high-angle shot of a massive, curling ocean wave. The wave's face is a vibrant greenish-blue, with white spray and foam visible where it is breaking. The water behind the wave is a darker, deeper blue. The overall texture is dynamic and suggests a strong current or wind.

# Datameesters



**Toonaangevend Big Data Expertise Centrum**

**De beste Big Data practitioner**



# De jacht op data sets

**Overwinnen barrière van vertrouwelijkheid  
en concurrentie gevoeligheid**

**Overwinnen ‘data eilanden’**

**Op slot gaan ‘Social Media’ data**

**Zelf initiëren van het ontstaan datasets**

**Data sets uit sensor netwerken**

**Open gaan overheidsdata**

A large, powerful ocean wave is shown crashing, with white spray and foam at the base. The water is a deep teal color, transitioning to lighter shades where it's disturbed by the wave. The overall scene conveys a sense of energy and movement.

**Big Data  
plus & min**



3602 hgtv dream home pictures [2006-03-29 21:49:51]  
3605 tv conference [2006-03-29 21:49:50]  
3605 tv convention [2006-03-29 21:50:21]  
3605 driving directions [2006-04-03 22:18:20]  
3605 driving directions [2006-04-03 22:18:20]  
3605 driving directions [2006-04-03 22:18:20]  
3605 hgtv.com [2006-04-08 10:13:40]  
3605 hgtv dream home pictures [2006-04-08 10:25:02]  
3605 hgtv dream home pictures [2006-04-08 10:25:02]  
3605 hgtv dream home pictures [2006-04-08 10:25:02]  
3605 snakes in mountains of eastern north carolina [2006-04-08 21:46:02]  
3605 snakes in mountains of eastern north carolina [2006-04-08 21:46:02]  
3605 . [2006-04-08 21:57:17]  
3605 pictures of snakes in north carolina [2006-04-08 21:58:07]  
3605 snakes of north carolina [2006-04-08 22:08:56]  
3605 pictures of snakes in north carolina [2006-04-08 22:10:50]  
3605 snakes of north carolina [2006-04-08 22:11:07]  
3605 hgtv dream home pictures [2006-04-08 22:11:17]  
3605 map of area around your home [2006-04-08 22:25:46]  
3605 gps map of area around your home [2006-04-08 22:26:57]  
3605 map question directions [2006-04-12 07:22:08]  
3605 map quest driving directions [2006-04-12 07:23:17]



AOL Main Mail AIM Join AOL Shopping AOL A to Z

**AOL search**  
powered by Google

**Web** **Images** **Video** **Audio** **News** **Shopping**

aol **SEARCH**

**Cash Loan.**  
Cash The Same Day, No Hassle Quick Cash Loan \$100 - \$2,000.  
[Cash4U.com](#)

**AOL.com - Welcome to AOL.**  
Start your Internet experience at AOL.com and find what you are looking for; including AOL download software at [www.aol.com/](#)

**Search the Web**  
The AOL Search engine delivers great search results, enhanced by Google, plus relevant multimedia results at [search.aol.com/](#)

**AOL Instant Messenger**  
The official Web site of AOL Instant Messenger. Exchange messages with friends, family, and colleagues at [www.aim.com/](#)

**Music Videos, Songs, Internet Radio, Free Music Downloads - AOL Music.**  
Listen to music, including AOL music radio, watch free music videos online, download free music, and more at [music.aol.com/](#)

**Welcome to AOL UK**  
AOL UK is a web portal offering free email, instant messaging, search, news, shopping, travel, music and more at [www.aol.co.uk/](#)



**Buried in a list of 20 million Web search queries  
collected by AOL and released in 2006 on the  
Internet is user No. 4417749**

**No. 4417749 hundreds of searches 3  
months “numb fingers”, “60 single men”,  
“dog that urinates on everything”,...**





No. 4417749

**Thelma Arnold, a 62-year-old widow who  
lives in Lilburn**

**22817** seems to look up every word in a dictionary

what does acute mean

what does accompany mean

what does adrenaline mean

what does alternative mean

what does acute mean

what does ample mean

what does abundant mean

what does ambition mean

what does ambiguous mean

what does agony mean

what does achieve mean

what does apprehend mean

what does annoy mean

**22817**'s gives up after just two hours. A while later, 22817 searches for "summer activities".

Maybe there's something more interesting to do?

**28963** looks for "porn sites". 28963 quickly amends the search query to read "freee porn sites".

(Two days later, 28963 shows a sudden interest in genital warts.)

**Hip Hop fan 29076** likes AntiStudy.com. His searches include "disney chanal", "emty lots", "michael jordon timeline" and "goolge".

**1133** is looking for "Google grass". (What's Google grass?)

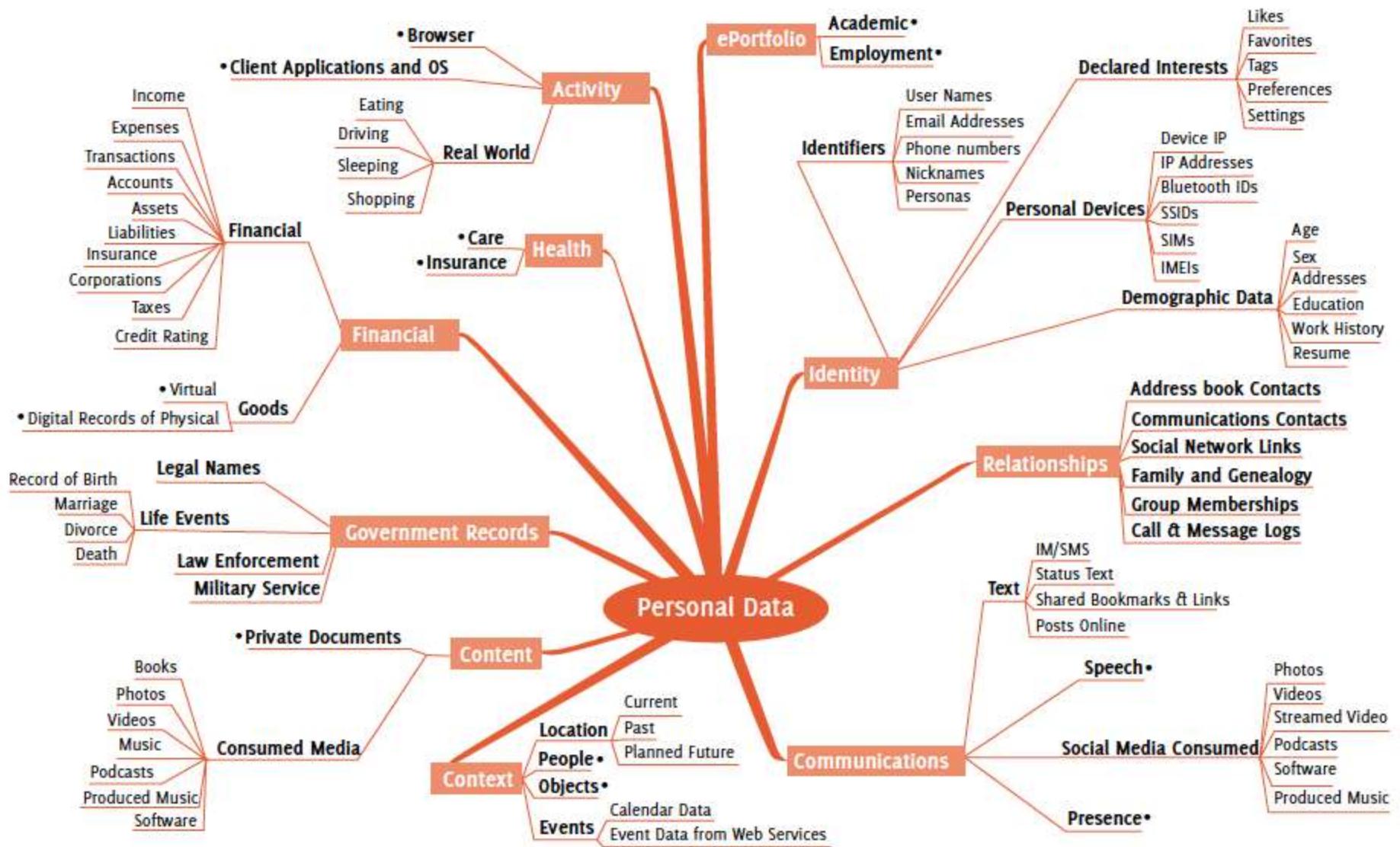
**2761** wants to acquire a box of lobster tails. Might come in handy for the trip to Amsterdam...



**AOL's chief technology officer left the company,**

**and two other workers were fired in the aftermath of a privacy breach**





Kaliya Hamlin, maakte de volgende mindmap *personally identifiable information*



## **1 Information Collection (use, storage, and manipulation of collected data)**

**Harms: Surveillance, Interrogation**

## **2 Information Processing**

**Harms: Aggregation, Identification, Insecurity, Secondary Use, Exclusion**

## **3 Information Dissemination**

**Harms: Breach of Confidentiality, Disclosure, Exposure, Increased Accessibility, Blackmail, Appropriation, Distortion**

## **4 Invasion**

**Harms: Intrusion, Decisional Interference**





# Privacy Enhancing Technologies:

Anonymisation

Protect Network Invasion

Identity Management (authentication and authorisation without identification)

Credential Systems

Trust Management

Data Processing

Privacy Preserving Data Mining

Privacy Management in Data Repositories

Policy-Checking



A wide-angle photograph of a rural landscape at dusk or dawn. The foreground is filled with tall, golden-brown grass. In the middle ground, a long, low line of small, colorful structures, possibly yurts or simple houses, stretches across the horizon. Beyond them are rolling hills and mountains under a sky with soft, pastel-colored clouds.

Privacy?

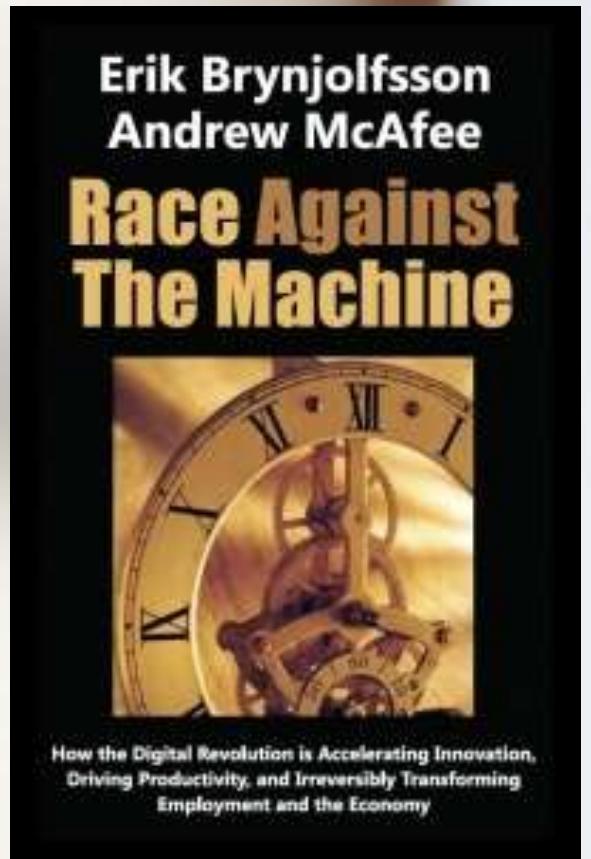


aangeven eigen privacy

# Point Break



**“cookie wet” — heeft  
tegenovergestelde effect**



Erik Brynjolfsson and Andrew McAfee,  
**Race Against the Machine: How the Digital  
Revolution Is Accelerating Innovation,  
Driving Productivity, and Irreversibly  
Transforming Employment and the  
Economy**

The background of the slide features a dramatic, close-up photograph of a massive, curling ocean wave. The water is a deep, dark teal color, with white spray and foam at the crest. The texture of the water is rough and turbulent, suggesting a powerful natural force.

**De donkere kant van data**

**Privacy en vertrouwelijkheid**

**“Race Against the Machine” automatisering kenniswerker**

**Open of Gesloten data**

**Vooroordeel**

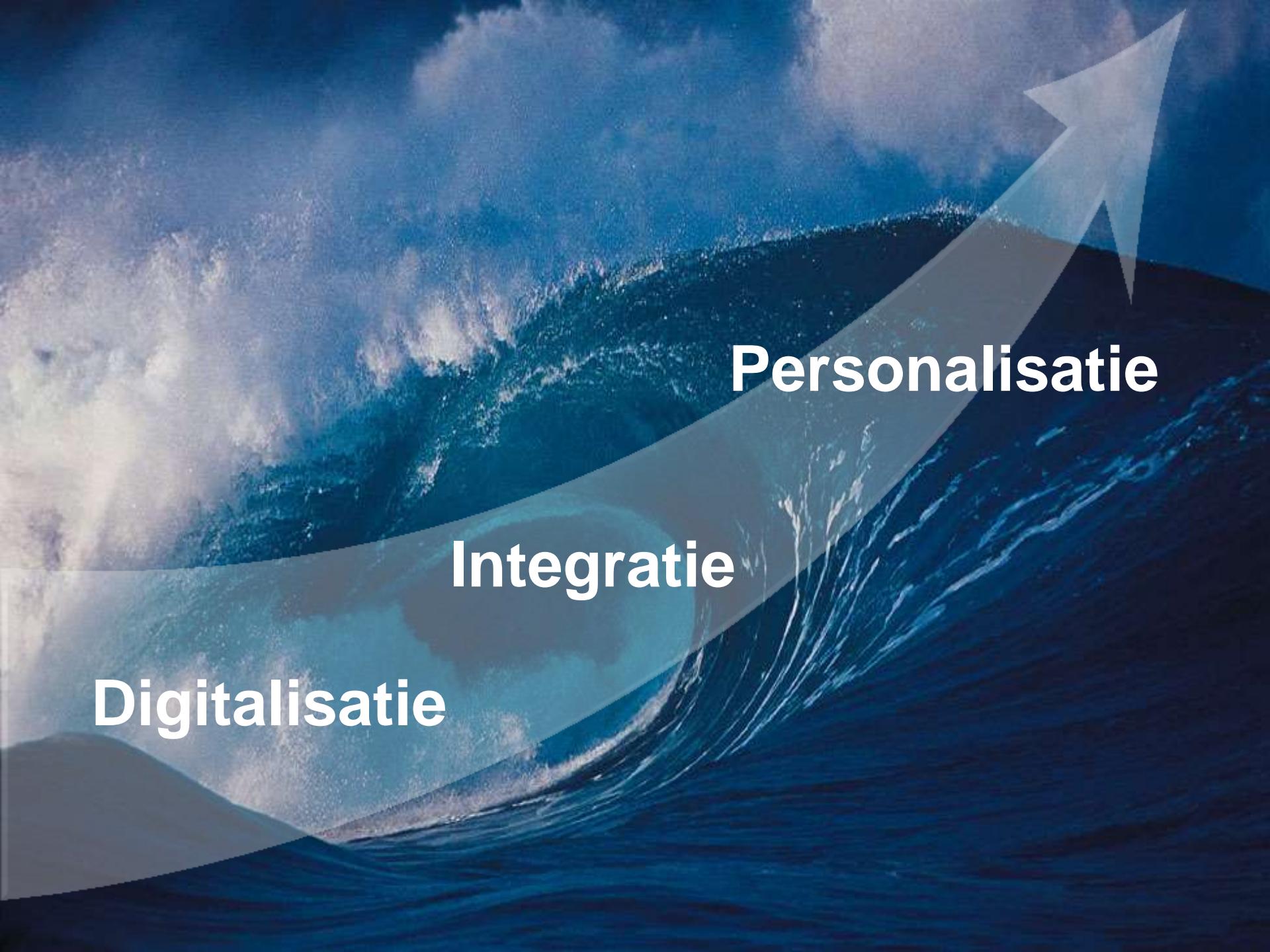
A large, powerful ocean wave is shown crashing, with a massive spray of white foam and golden-yellow water. The wave's base is dark and turbulent, while its crest is bright and luminous. The background consists of more ocean waves under a clear sky.

**Democratisering data**

**Cloud**

**Open Source**

**Open Data**

The background image shows a powerful ocean wave crashing onto a dark, craggy coastline. The water is a deep, vibrant blue, and the spray from the wave is white and frothy. Above the horizon, the sky is filled with heavy, grey clouds. A large, semi-transparent grey diagonal band runs from the bottom left towards the top right, serving as a backdrop for the text.

Digitalisatie

Integratie

Personalisatie

# Consument data

Profiling

Real Time

Voorspellend

Personalisatie

Integratie

Digitalisatie

Consument  
Self service  
Vaststellen sentiment  
Beïnvloeden gedrag  
Meer betrokken



**absorberen effectief maken  
van bestaande Big Data kennis**

# Bronnen:

**“De big data revolutie” van Viktor Mayer-Schonberger en Cukier**

**“The Signal and the Noise – Why So Many Predictions Fail-but Some Don’t “ van Nate Silver**

**“Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses “ van Michael Minelli , Michèle Chambers en Ambiga Dhiraj**

**“Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die” Eric Siegel**

**McKinsey Global Institute 2011:**

**Big Data: The next frontier for innovation, competition and productivity**

**Fraunhofer IAIS 2013: Innovationpotentialanalyse 2013**

# Bronnen:

## Hadoop

*[benodigde basiskennis: SQL, Java, Linux]*

Tom White, Hadoop: the Definitive Guide, 3rd Ed. O'Reilly

Donald Milner, Map Reduce Design Patterns, O'Reilly

Chuck Lam, Hadoop in Action, Manning

Big Data University (IBM) [bigdatauniversity.com](http://bigdatauniversity.com)

Cloudera training courses

2014 Hadoop Summit, 31 maart - 3 april, Amsterdam

O'Reilly Strata Conference

Lector Roel Bakker

# Bronnen:

## Data mining / machine learning

### Statistiek en R:

Coursera, Johns Hopkins, Jeff Leek: [Computing for Data Analysis](#)

P. Dalgaard. [Introductory Statistics with R](#)

B. Yakir. [Introduction to Statistical Thinking](#)

### Statistical Learning

<http://online.stanford.edu/course/statistical-learning-winter-2014>

James, Witten, Hastie and Tibshirani: [An introduction to Statistical Learning](#)

Hastie, Tibshirani, Friedman:[Elements of Statistical Learning](#) (naslagwerk)

Hillary Mason. [An Introduction to Machine Learning with Web Data.](#) O'Reilly.

Hillary Mason. [Advanced Machine Learning.](#) O'Reilly.

Lector Roel Bakker

**“De big data revolutie” van Viktor Mayer-Schonberger en Cukier**

**“The Signal and the Noise – Why So Many Predictions Fail-but Some Don’t “ van Nate Silver**

**“Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses “ van Michael Minelli , Michèle Chambers en Ambiga Dhiraj**

**“Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die” Eric Siegel**

**McKinsey Global Institute 2011:**

**Big Data: The next frontier for innovation, competition and productivity**

**Fraunhofer IAIS 2013: Innovationpotentialanalyse 2013**

**Andarr**      **Emmalaan 21 3581 HN Utrecht**

**Telefoon**    **+31 30 6086188**

**Mobiel**        **+31 6 53606770**

**Email**            **[evtol@andarr.nl](mailto:evtol@andarr.nl)    [e.vantol@fontys.nl](mailto:e.vantol@fontys.nl)**

**Internet**      **[www.andarr.nl](http://www.andarr.nl)**

A dramatic photograph of a massive ocean wave crashing towards the viewer. The wave's face is a deep, dark teal, while the spray and base are a lighter, foamy white. The background consists of turbulent, white-capped waves under a clear blue sky.

Nederland investeert  
in Big Data kennis