Big Data

Auch Kleine können Gross!

Dr. Daniel Benninger
Sawubona GmbH Bern & Hochschule Luzern

13.11.2019

Blick News Sport Meinung Politik Wirtschaft People Leben Green

21.10.2019 08:41 Uhr

Mit App auf Stimmenfang

Die FDP setzt im Wahlkampf auf Big Data

Was man bereits aus den USA kennt, hält
Die FDP setzt im Wahlkampf auf daten
heisst, sie besucht potentielle FDP-S
spricht mit ihnen und animiert sie, ih

ME » NEWS » GESELLSCHAFT

Big Data für die Tonne: KI verbessert Müllabfuhr

Künstliche Intelligenz hilft Forschenden, die Routen der Müllabfuhr zu optimieren. Programmiert hat dieses sogenannte neuronale Netzwerk ein Informatik-Absolvent der HSLU.

von pd/jst 08.10.2019

«Güselsäcke» zählen klingt im ersten Augenblick nicht nach einem Forschungsgegenstand. In David Jennis Fall sieht die Sache etwas anders aus. Für seine Master-Arbeit «trainierte» der Informatik-Absolvent der Hochschule Luzern eine Künstliche Intelligenz (KI), Abfallsäcke und -container zu erkennen und zu zählen. Ziel des Projekts: die automatisierte Erfassung von Abfallmengen.

Abfall-Lasters bei der Arbeit: Eine KI könnte dereinst die Route der Müllabfuhr

imann AG

WIRTSCHAFT BERNER OBERLAND 26. OKTOBER 2019

Big Data und Schweizer Holz für das Berner Oberland

Was verrät Big Data über das Berner Oberland? Warum soll zukünftig vermehrt auf den Baustoff Holz gesetzt werden? Was können Gemeinden zu wirtschaftsfreundlichen Rahmenbedingungen beitragen? Zu diesen Fragen informierten sich Vertreterinnen und Vertreter der Gemeindebehörden am Workshop in Spiez.

Sawubona

Spielend reich werden: Im Datenmeer soll das la
möglich werden. (Bild: Wallace Woon / EPA)

Wer soll mit meinen Daten Geld verdienen?

Selfies posten, Likes verteilen und dafür ein paar tausend Dollar erhalten – das klingt verlockend und wäre auch logisch: Informationen über unser Privatleben sind für Firmen wertvoll, also sollten sie uns dafür bezahlen. Trotzdem ist dieses Modell problematisch.

Roberto Simanowski
12.11.2019, 05:30 Uhr

Daten sammeln und die richtigen Schlüsse ziehen: Das ist zentral in der Medizin. (Bild: SPL / Keystone)

KOMMENTAR

Patienten brauchen Ärzte mit menschlicher Intelligenz

Big Data, Algorithmen und künstliche Intelligenz: Das sind auch in der Medizin die Themen der Stunde. Vor lauter Hype droht dabei in Vergessenheit zu geraten, worum es bei der ärztlichen Tätigkeit geht.

Alan Niederer
12.11.2019, 05:30 Uhr

Illustration: Peter Gut

GASTKOMMENTAR

Im Rausch des Sammelns von Einsen und Nullen

Der Nutzwert von Daten wird fetischisiert: Alle wissen, dass sie wertvoll sind – aber niemand weiss genau, wie wertvoll eigentlich. Daten werden heute allüberall in der Hoffnung gehortet, sie dereinst zu Geld machen zu können.

Adrian Daub
11.11.2019

Typologie
Charakterisierung

Ausblick

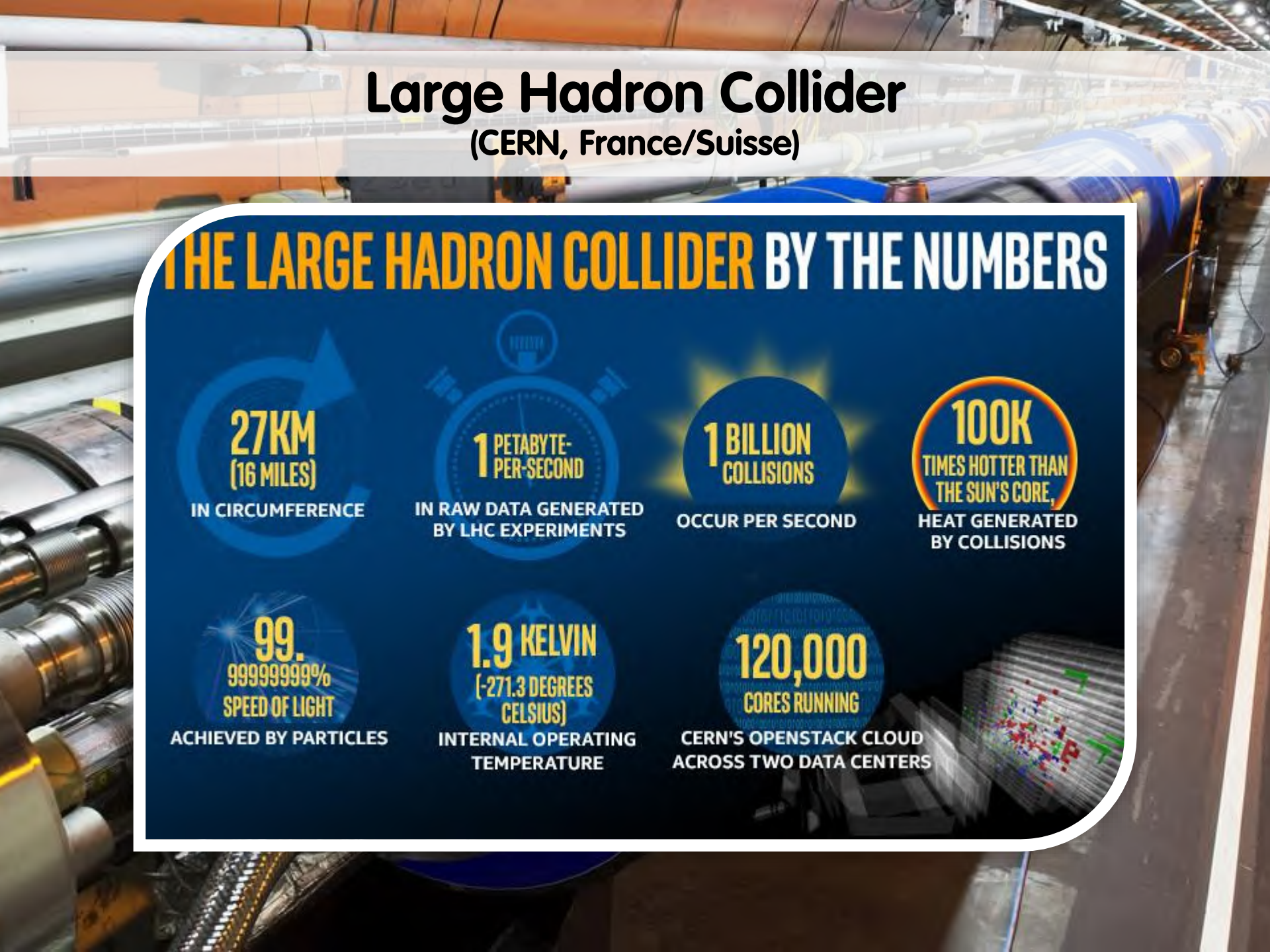
Historie
Wurzeln

Beispiele
Use Cases

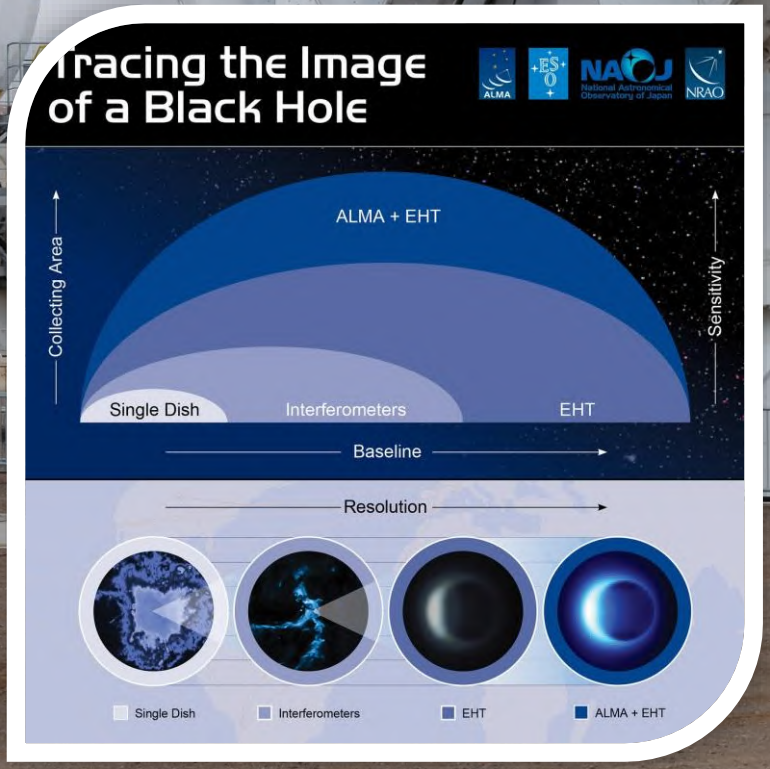
Historie

Ursprung und Entwicklung von Big Data

Large Hadron Collider
(CERN, France/Suisse)



ALMA Observatory
(Atacama Desert, Chile)



FIFA World Cup 2018
(SRG/Swiss TXT,Switzerland)

FIFA WC 2018 Streaming Report

Webtrekk & Akamai Network Data

SWISS TXT

18'400'000
Views (Starts)

3'900'000
Viewers

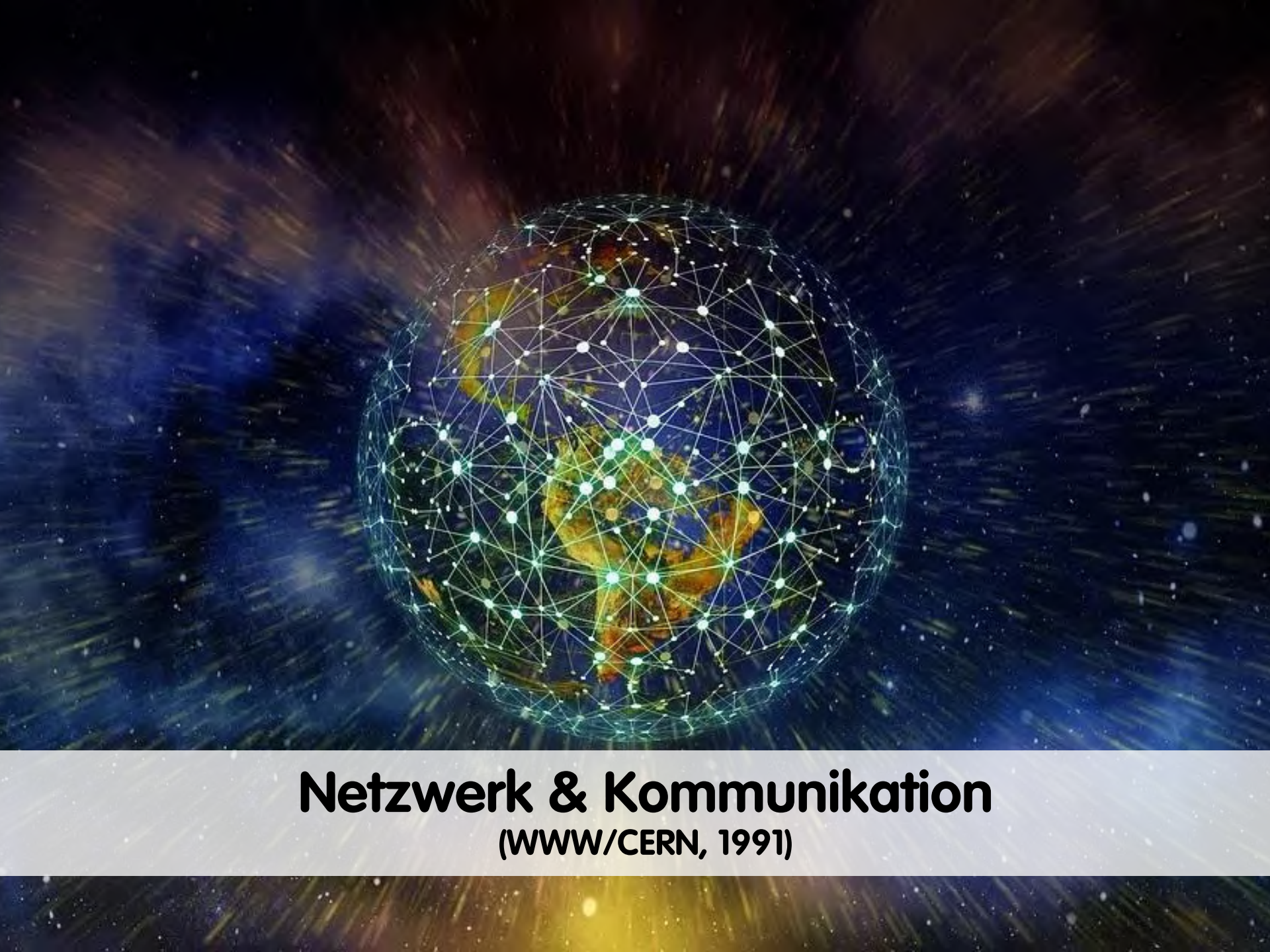
Total amount of Plays and Viewers on FIFA Special on Web and App solutions over all days and events. Simulcast stream plays are not included.

Concurrent Streaming Peak: 320'000

SUI - SWE 8th final game 3.7. at 16:40

7'186'788 Hrs
Total Streaming Time

10,7 PB
Live Video Content delivered



Netzwerk & Kommunikation

(WWW/CERN, 1991)

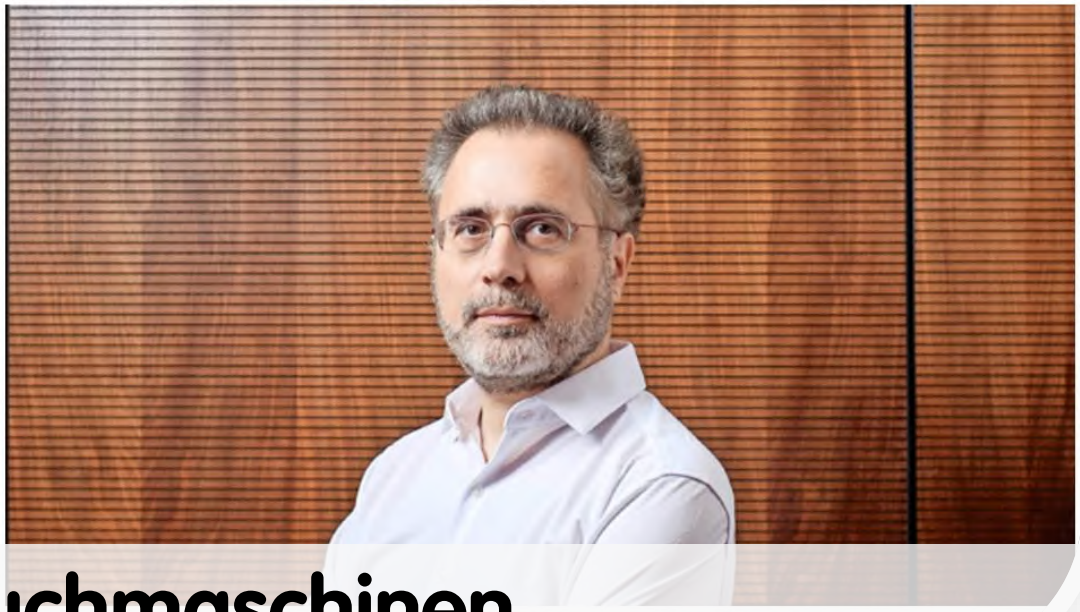
Smartphone & Apps

(Apple, 2007)

«Unser Ziel war nur, irgendwie den nächsten Montag zu überleben»

Googles Mitarbeiter Nummer 8, der Schweizer Urs Hölzle, spricht über die Anfänge der Firma.

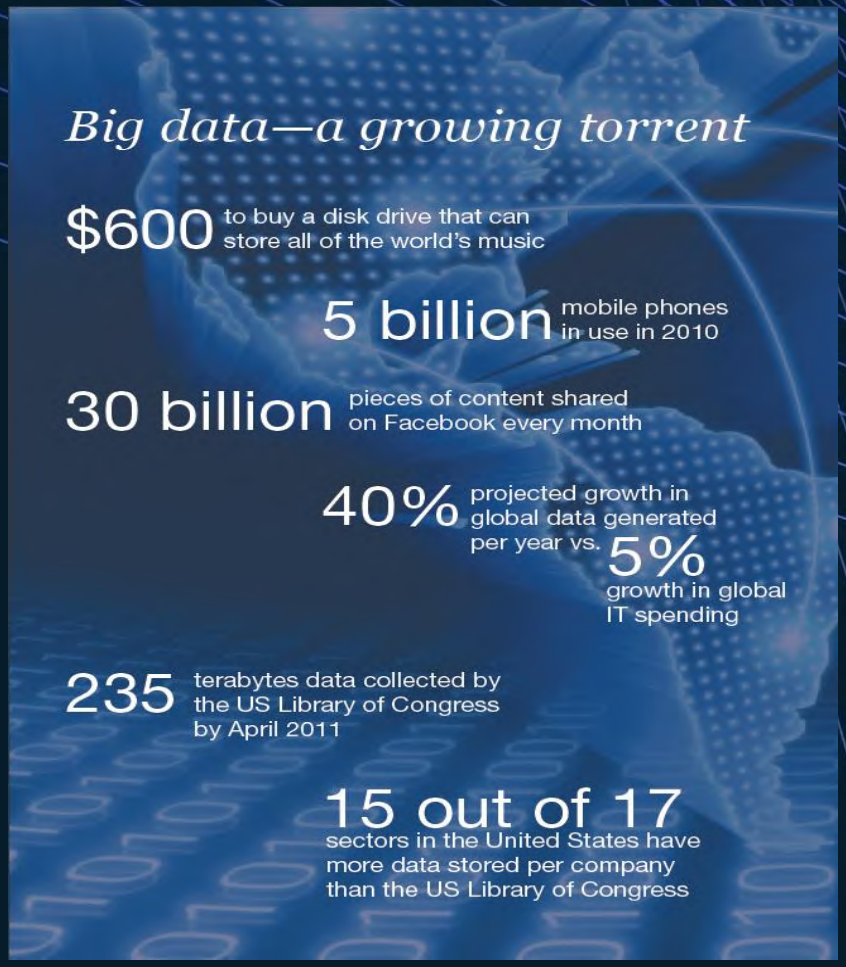
Freitag 8. November 2019 14:01 von Malte Conradi

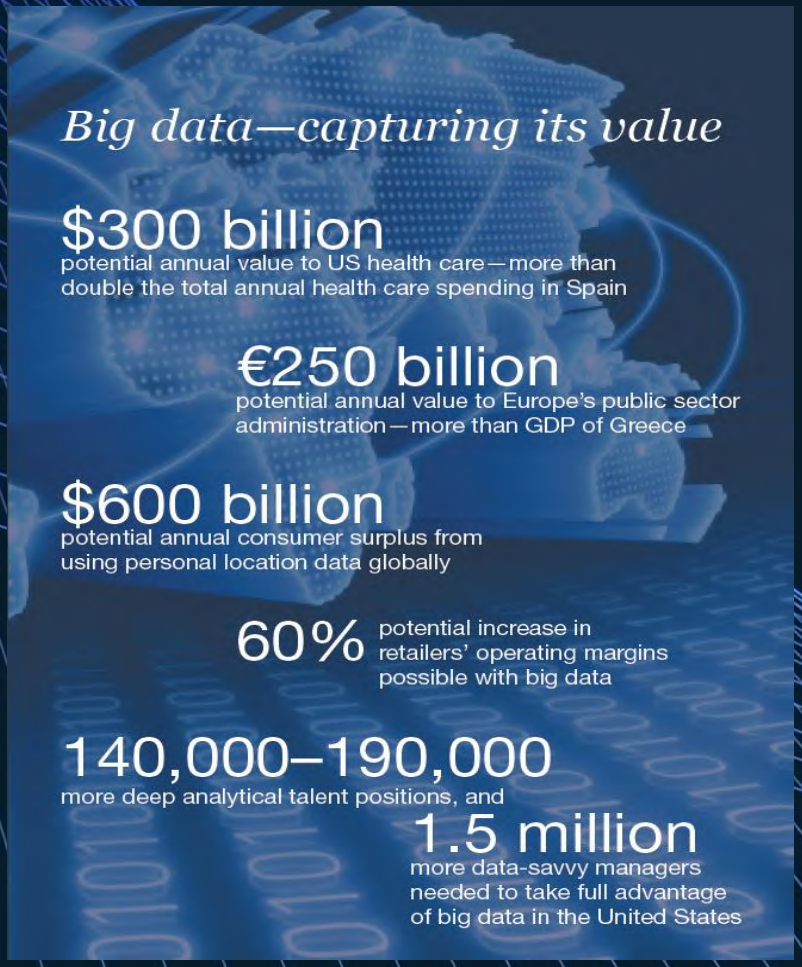


Suchmaschinen

(Google, 1997)

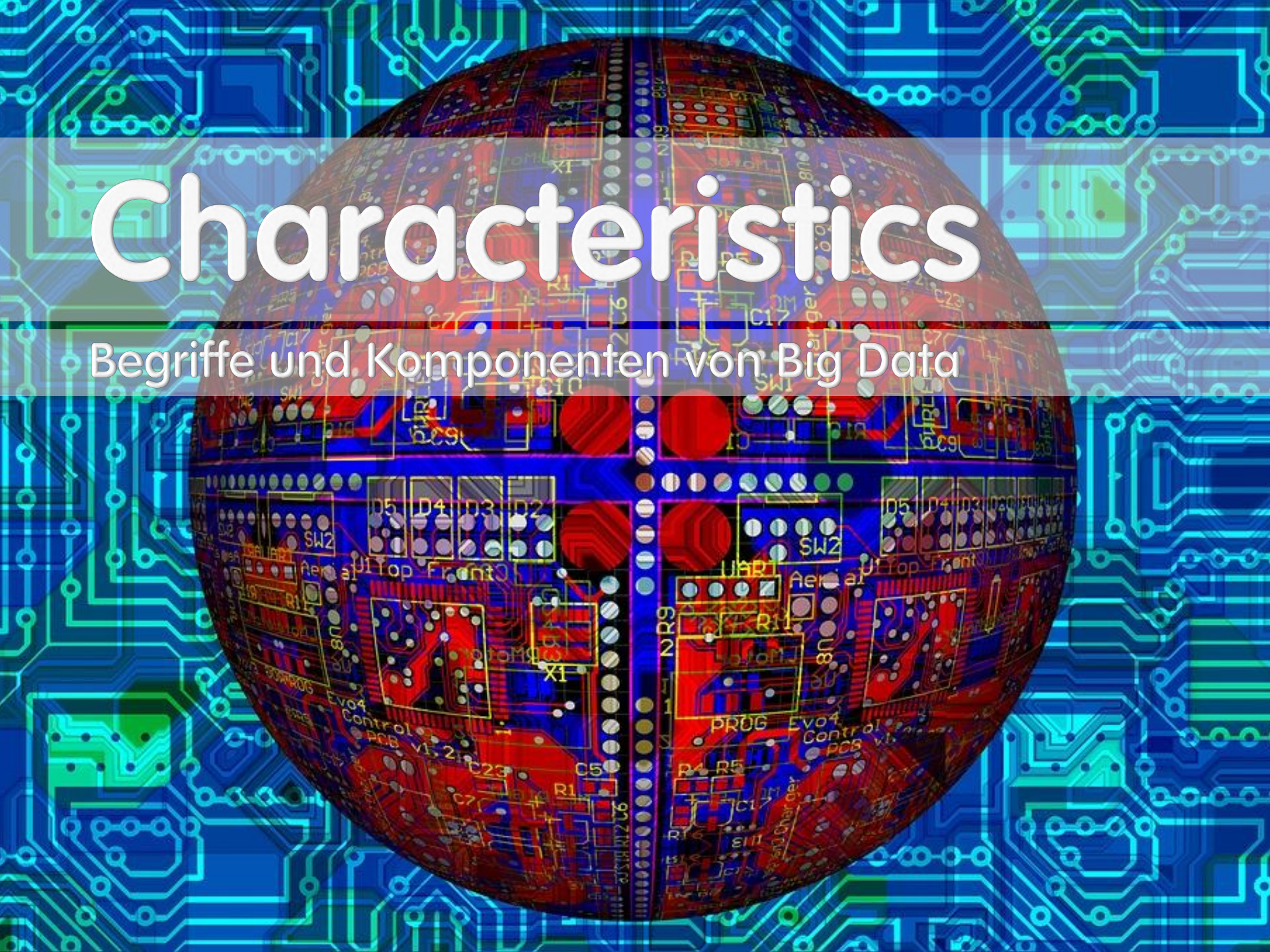
«Big Data Analytics: The next Fontier for Innovation, Competition, and Productivity»





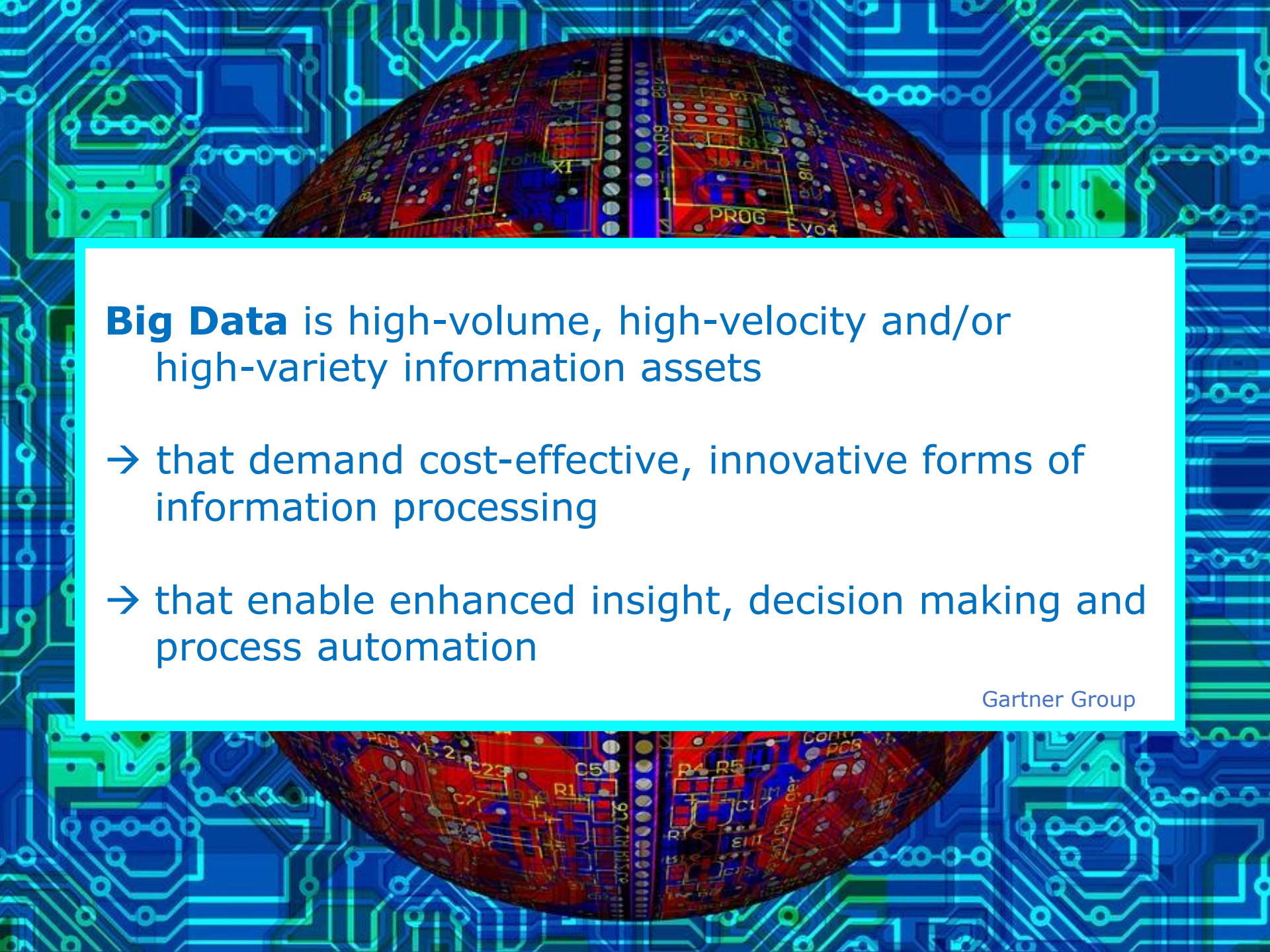
Berater

(McKinsey, 2011)



Characteristics

Begriffe und Komponenten von Big Data



Big Data is high-volume, high-velocity and/or high-variety information assets

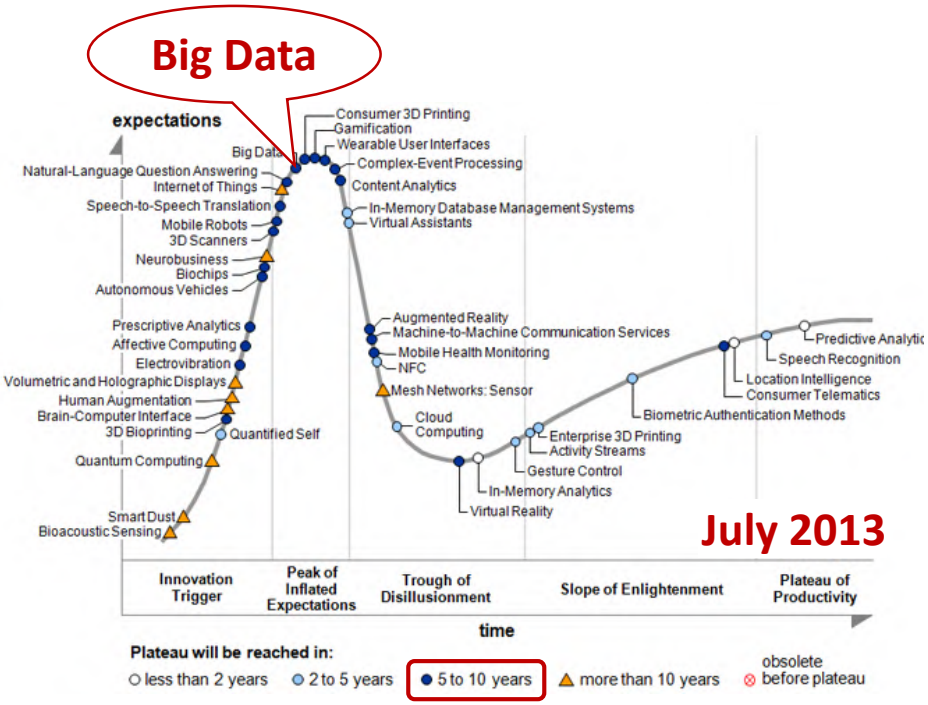
→ that demand cost-effective, innovative forms of information processing

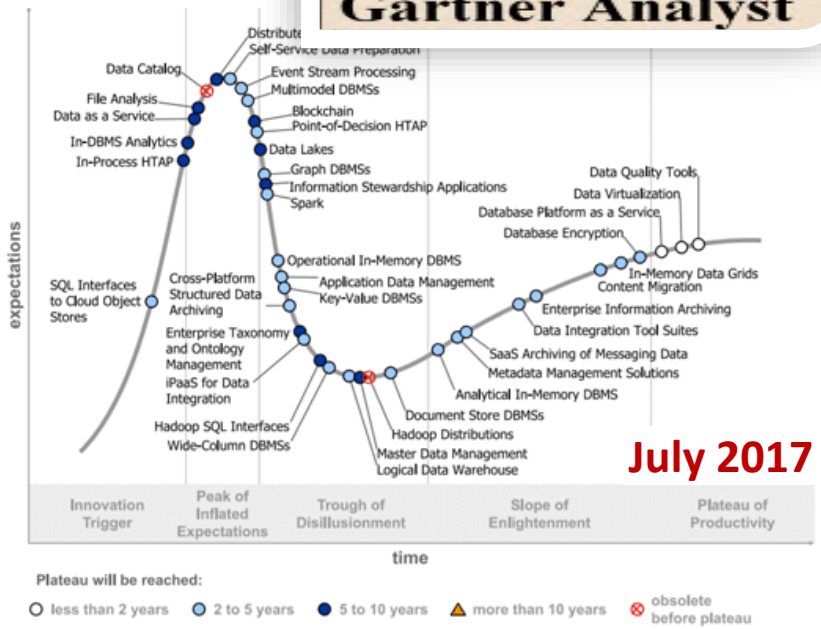
→ that enable enhanced insight, decision making and process automation

Gartner Group

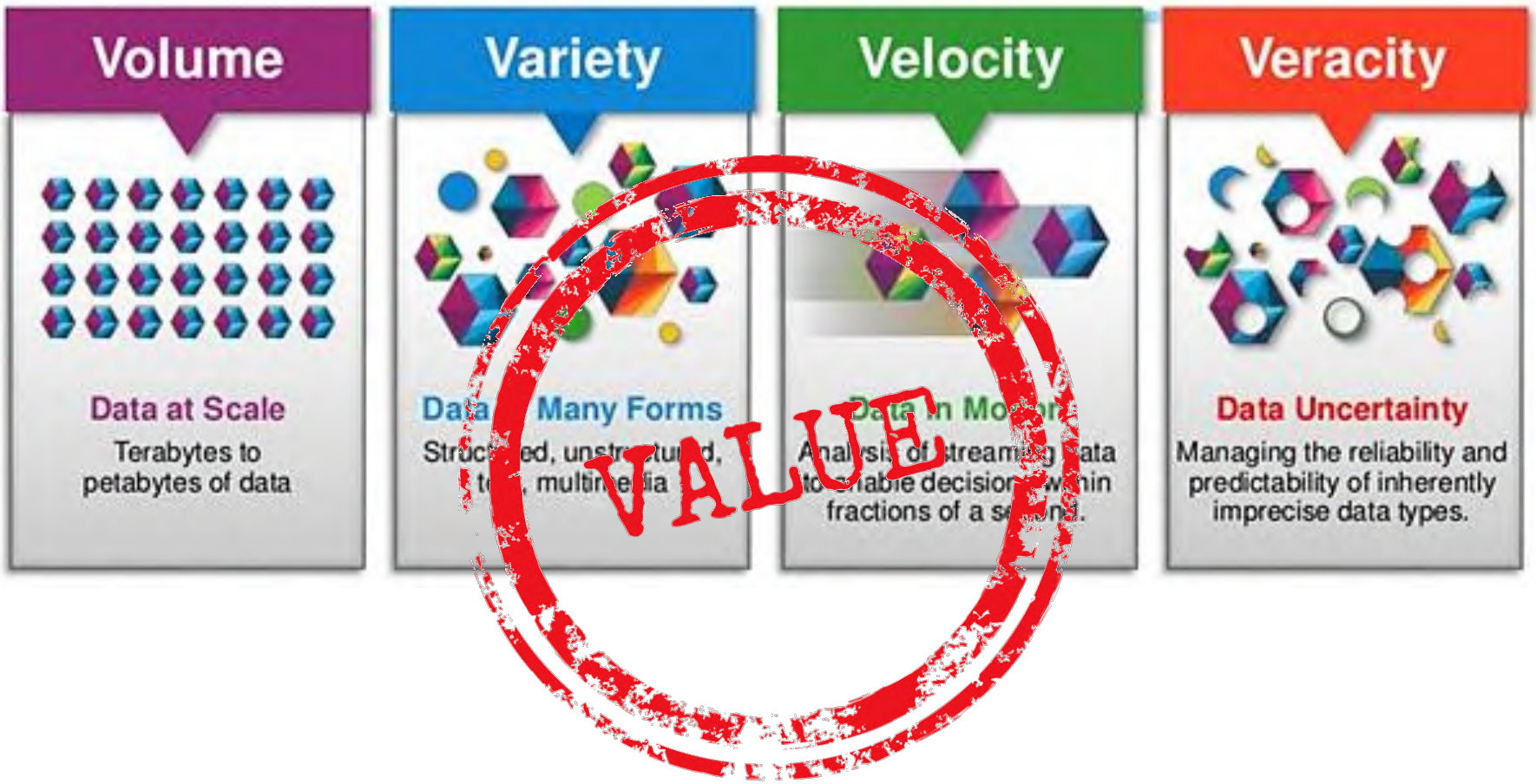
Big Data by Gartner Group

"Big data...has become prevalent in our lives"
-- Betsy Burton, Gartner Analyst

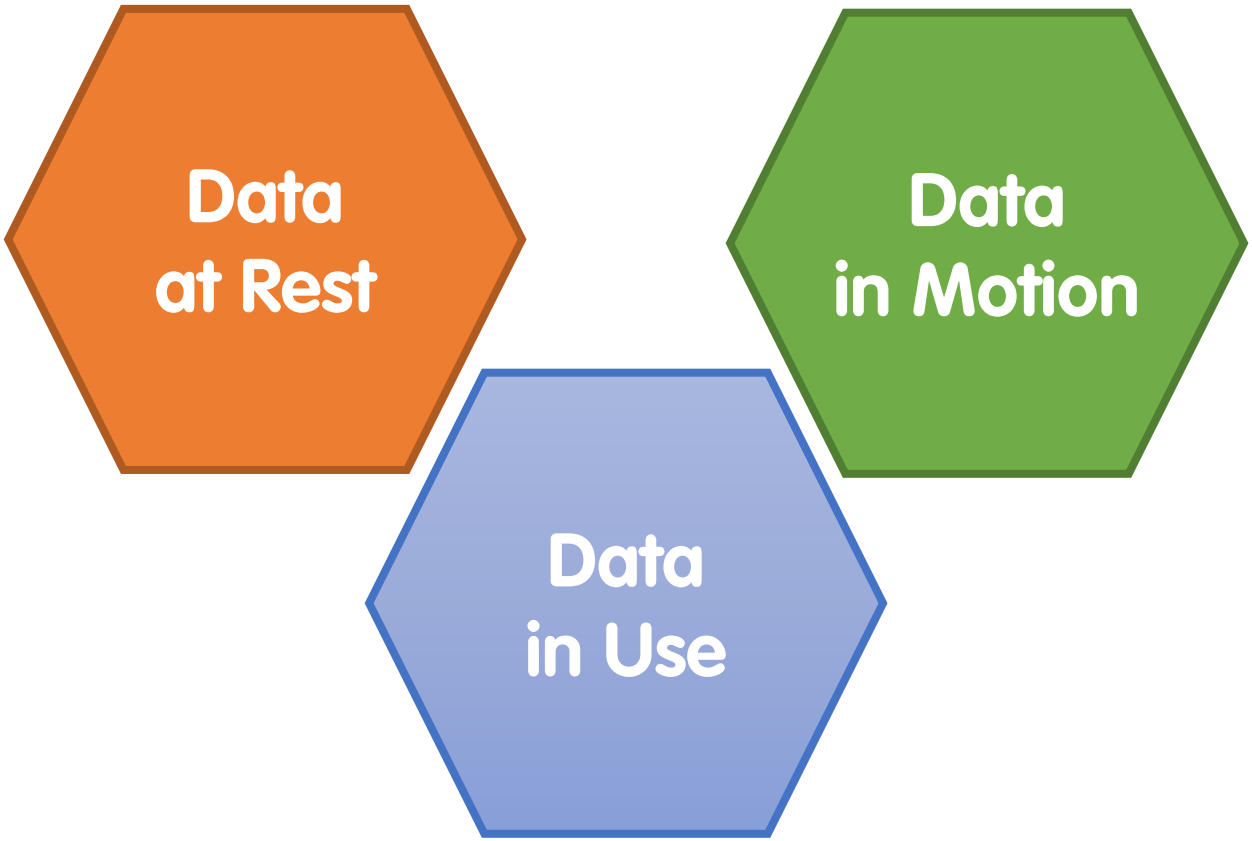




Big Data Characteristics - The 5 V's

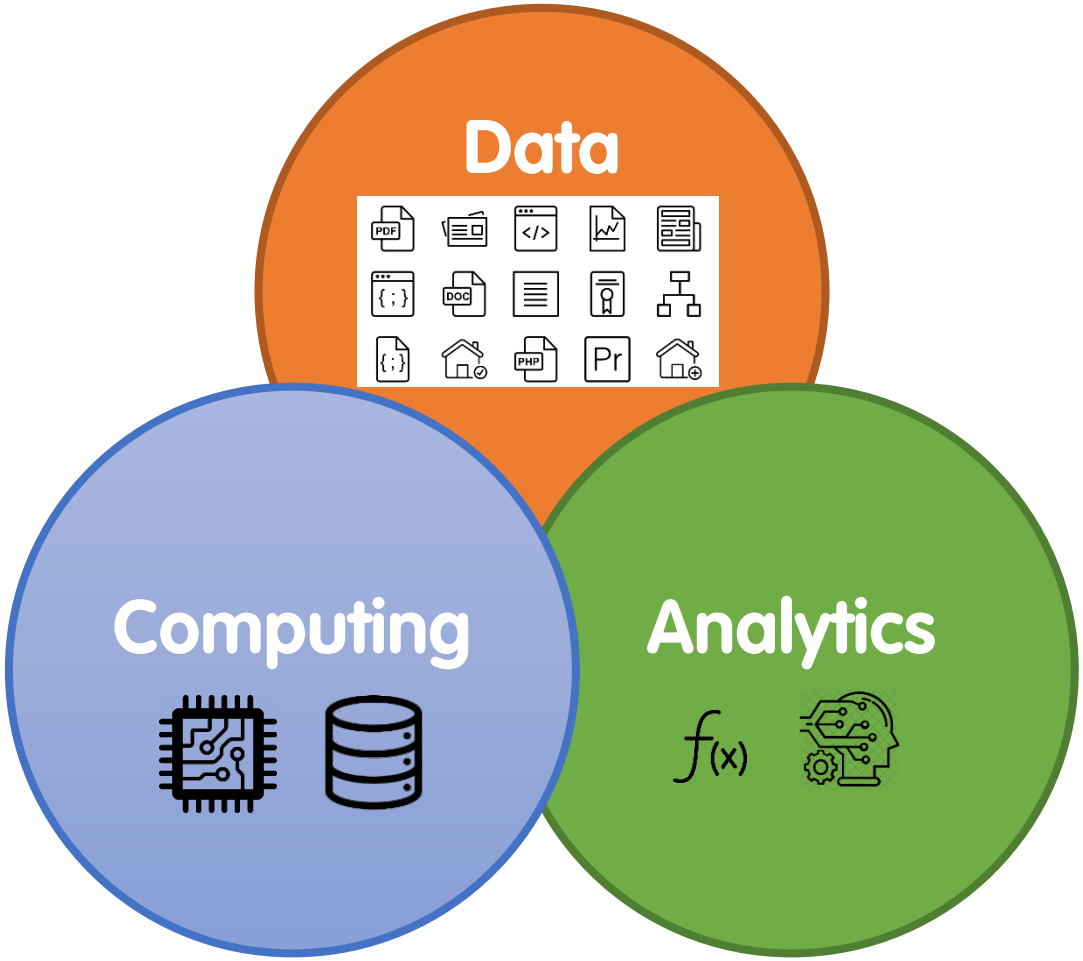


Stages of Digital Data

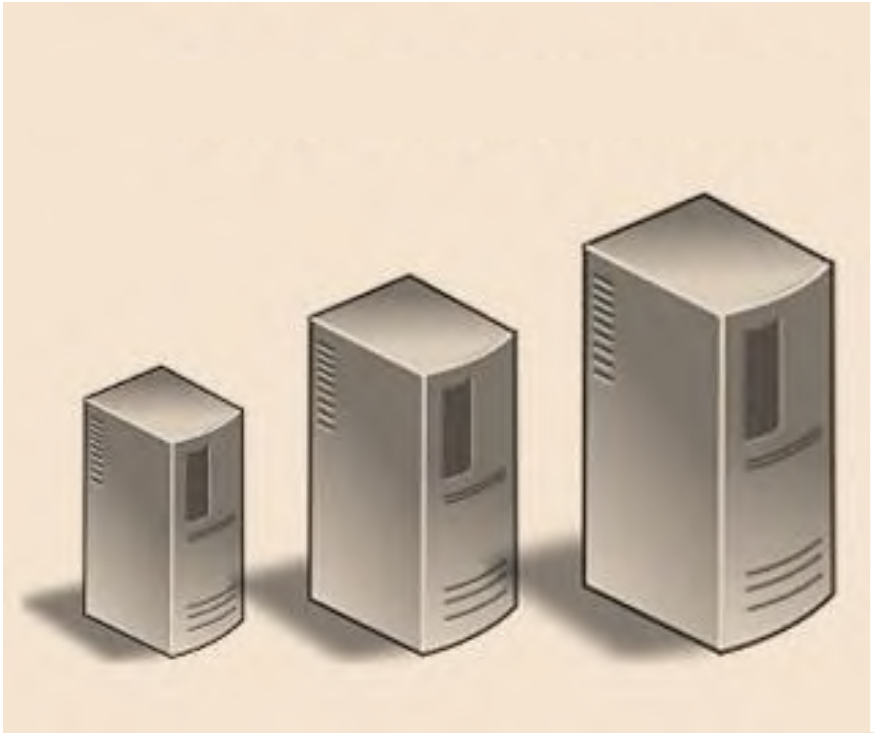


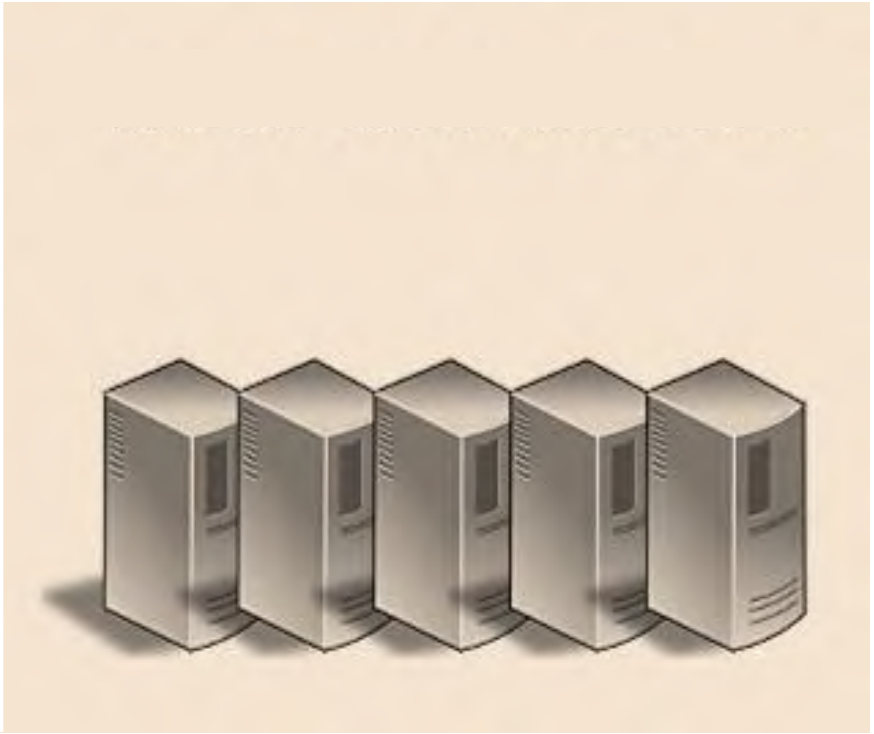
Data in Motion (e.g. Twitter)

Key Elements



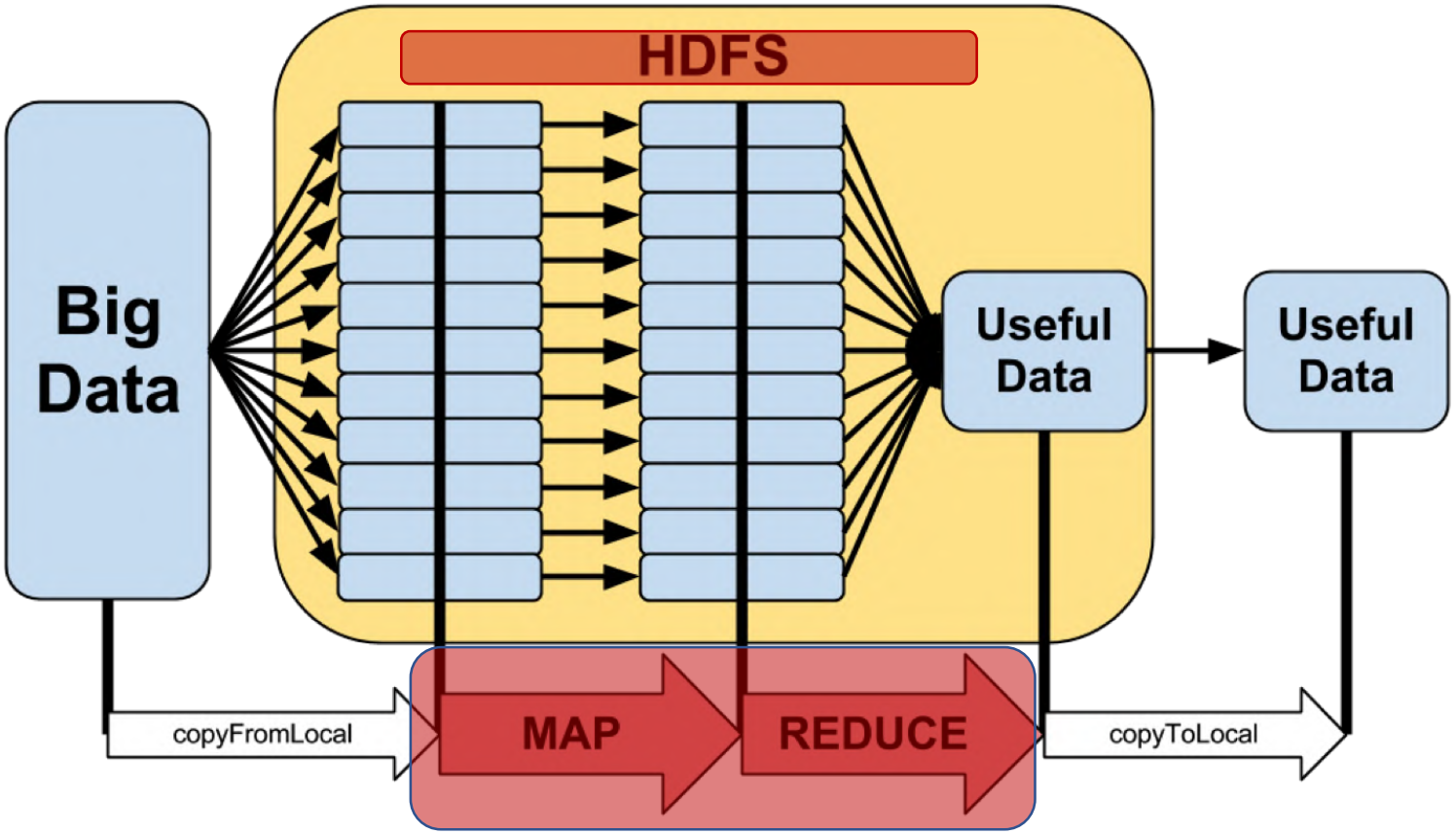
Scale Up vs. Scale Out



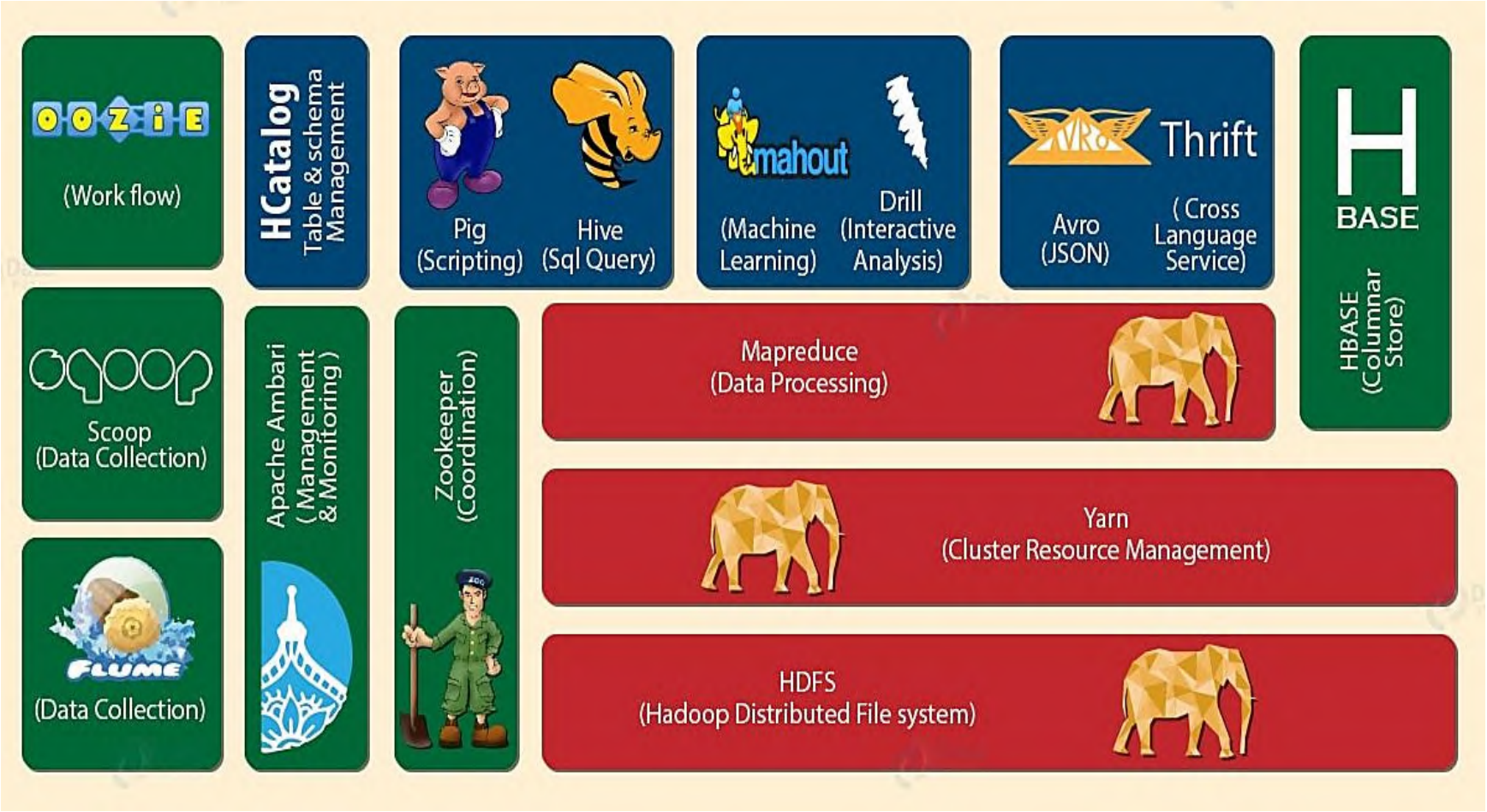



hadoop
Map Reduce

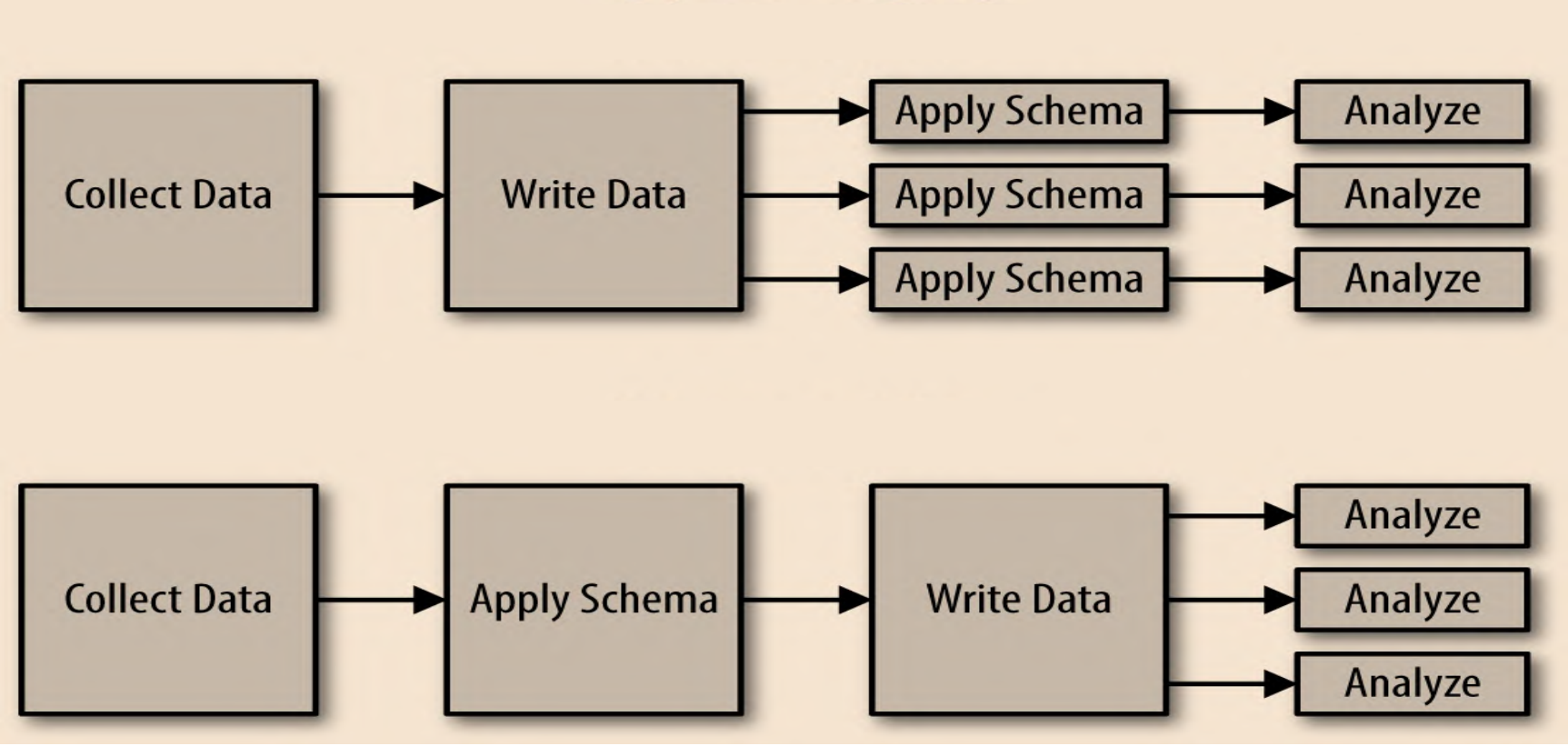
The Hadoop & MapReduce Framework



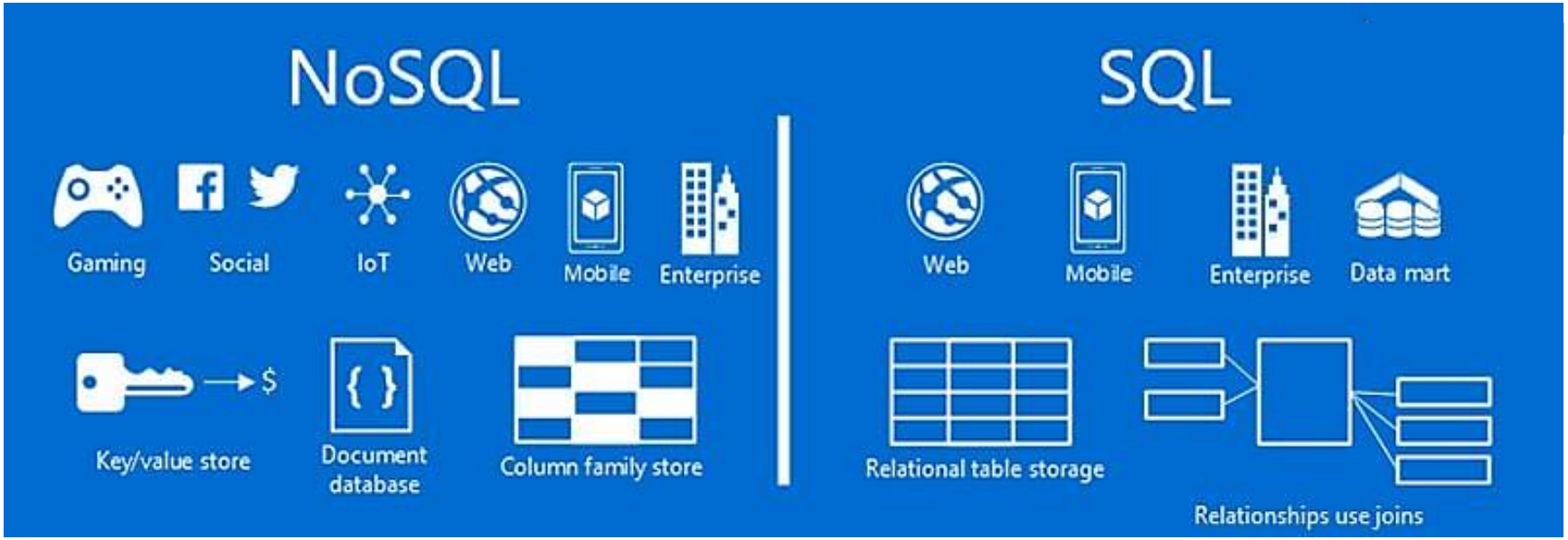
Hadoop & MapReduce – A Whole Ecosystem



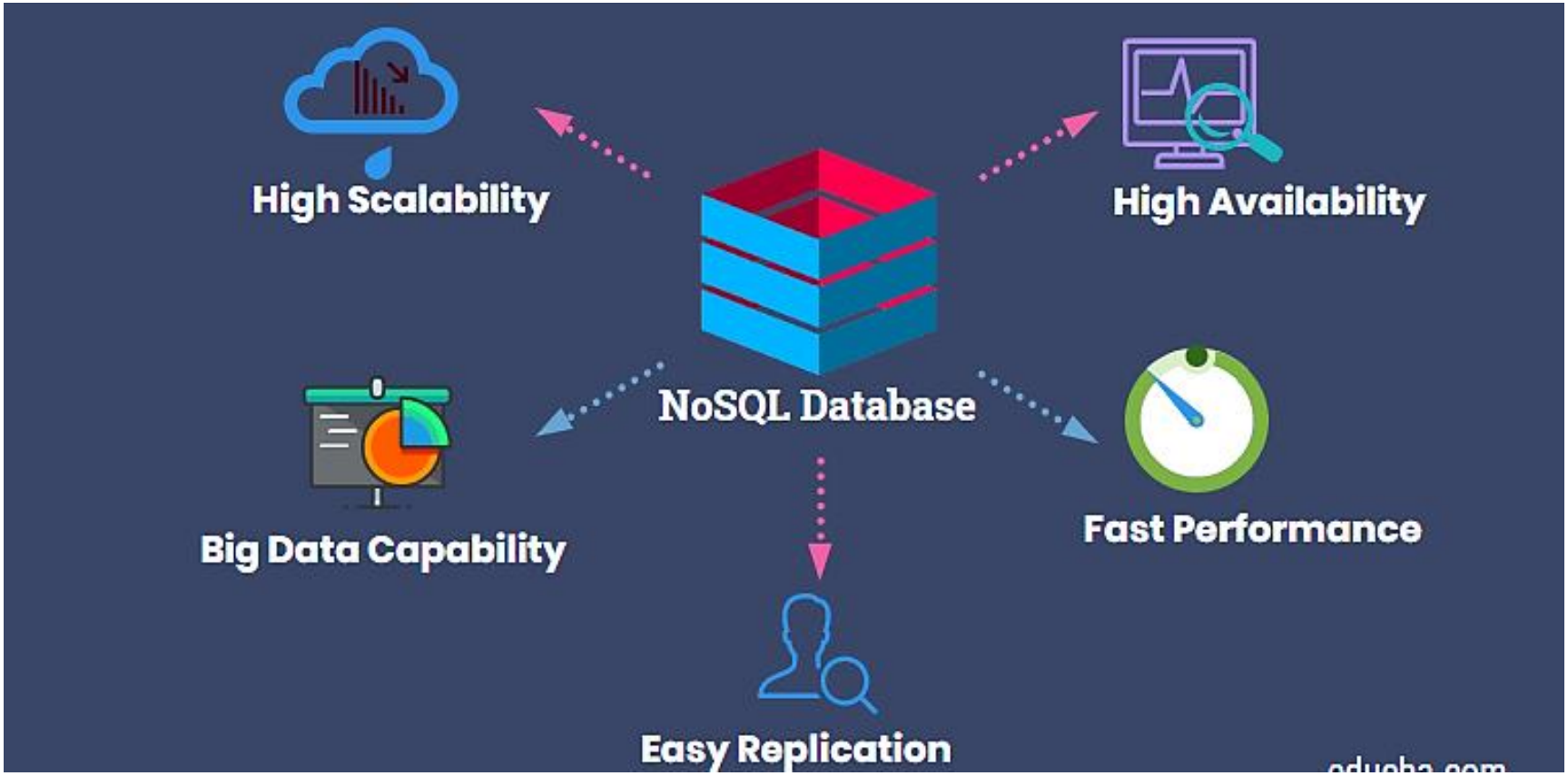
Schema on Read vs. Schema on Write



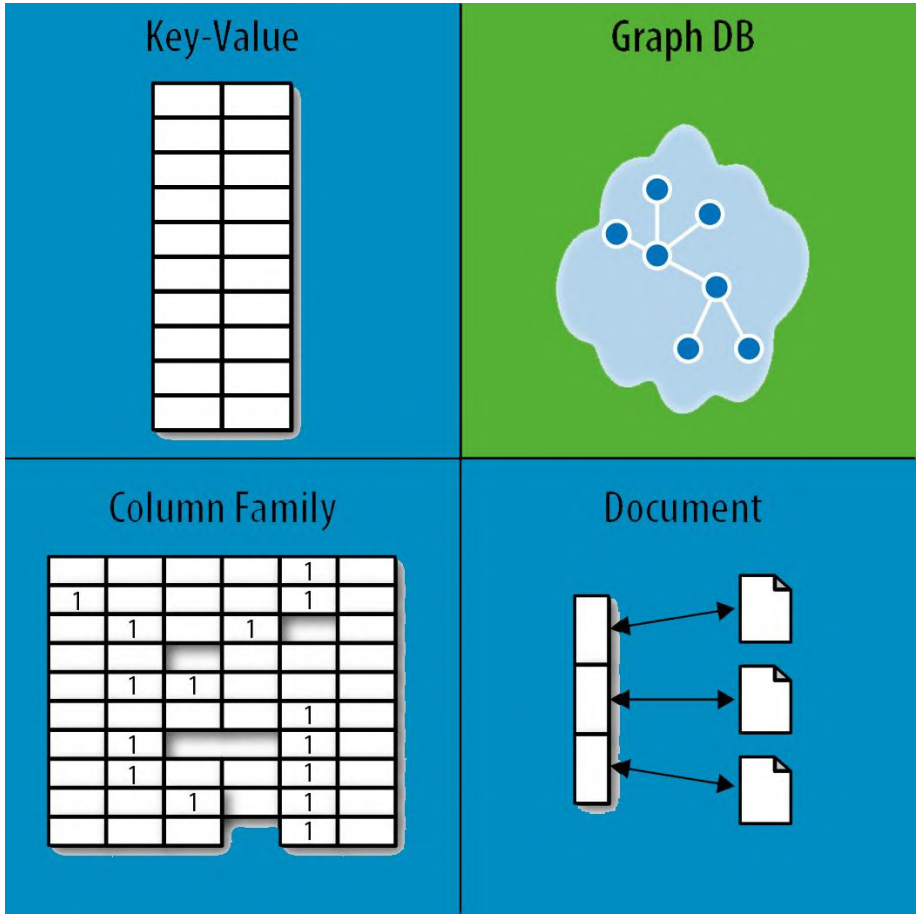
NoSQL vs. SQL


NoSQL
Gaming
Social
IoT
Web
Mobile
Enterprise
Key/value store
Document database
Column family store
SQL
Web
Mobile
Enterprise
Data mart
Relational table storage
Relationships use joins

The Evolution of NoSQL Databases



The Evolution of NoSQL Databases



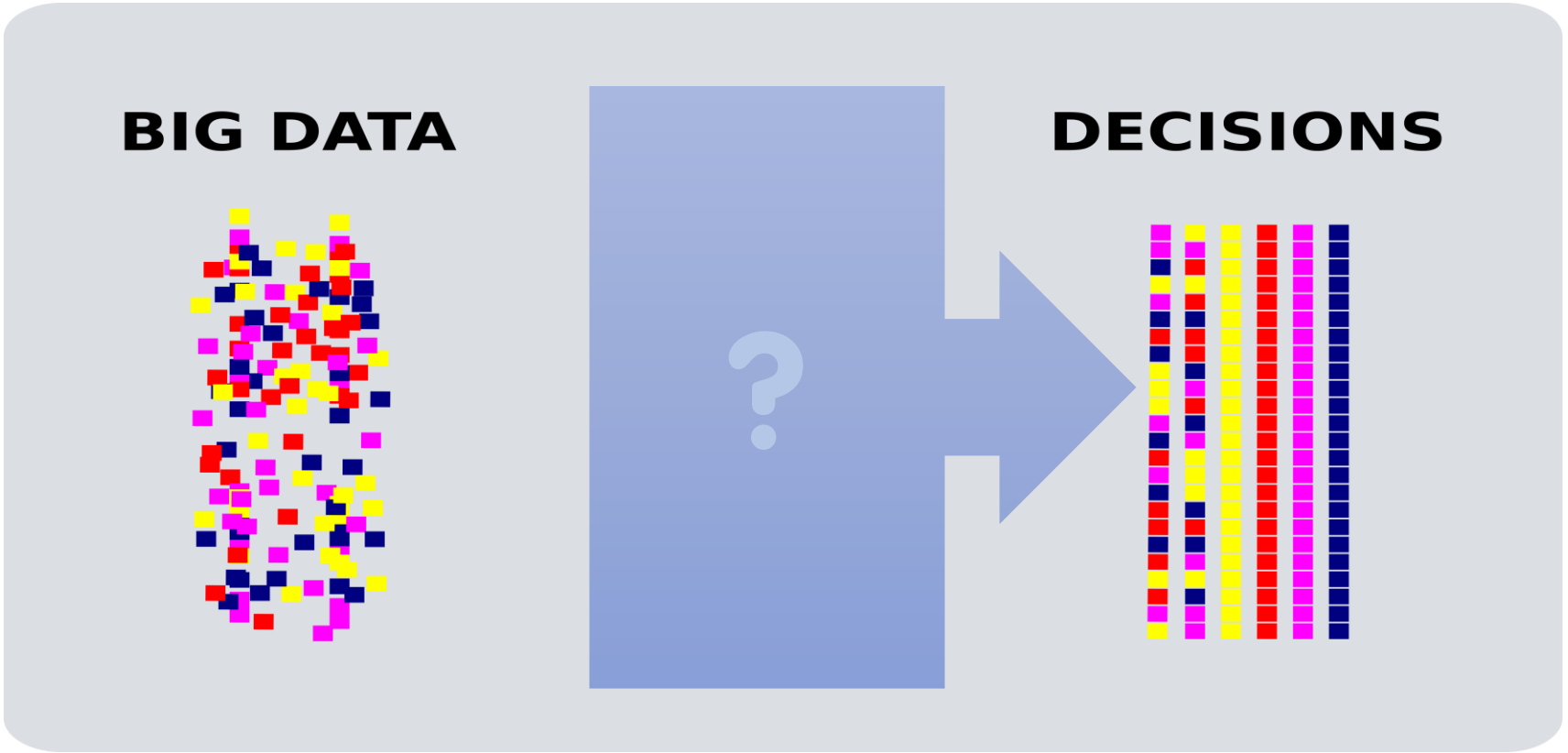
The Evolution of Vendors, Services and Products

V2 – Last updated 5/3/2017

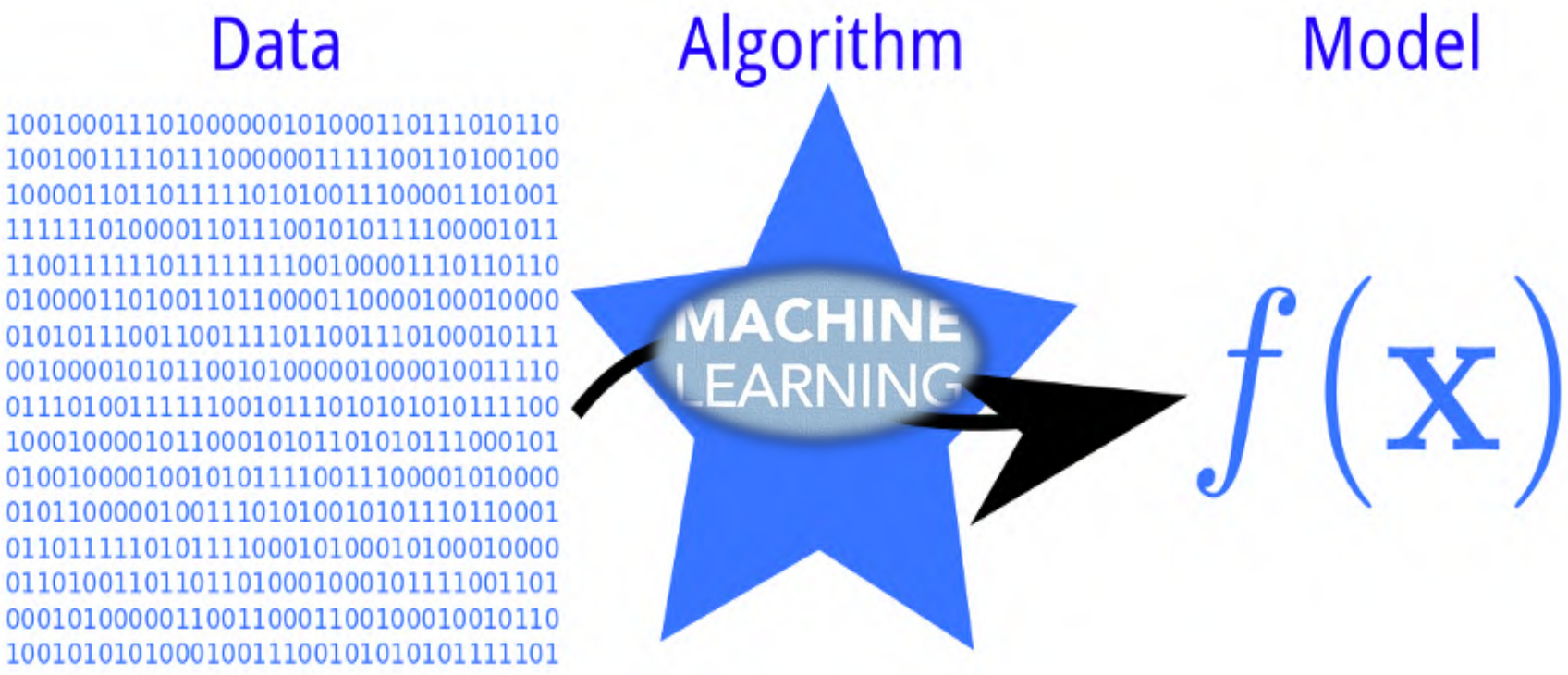
 mattturck.com/bigdata2017

FIRSTMARK EARLY STAGE VENTURE CAPITAL

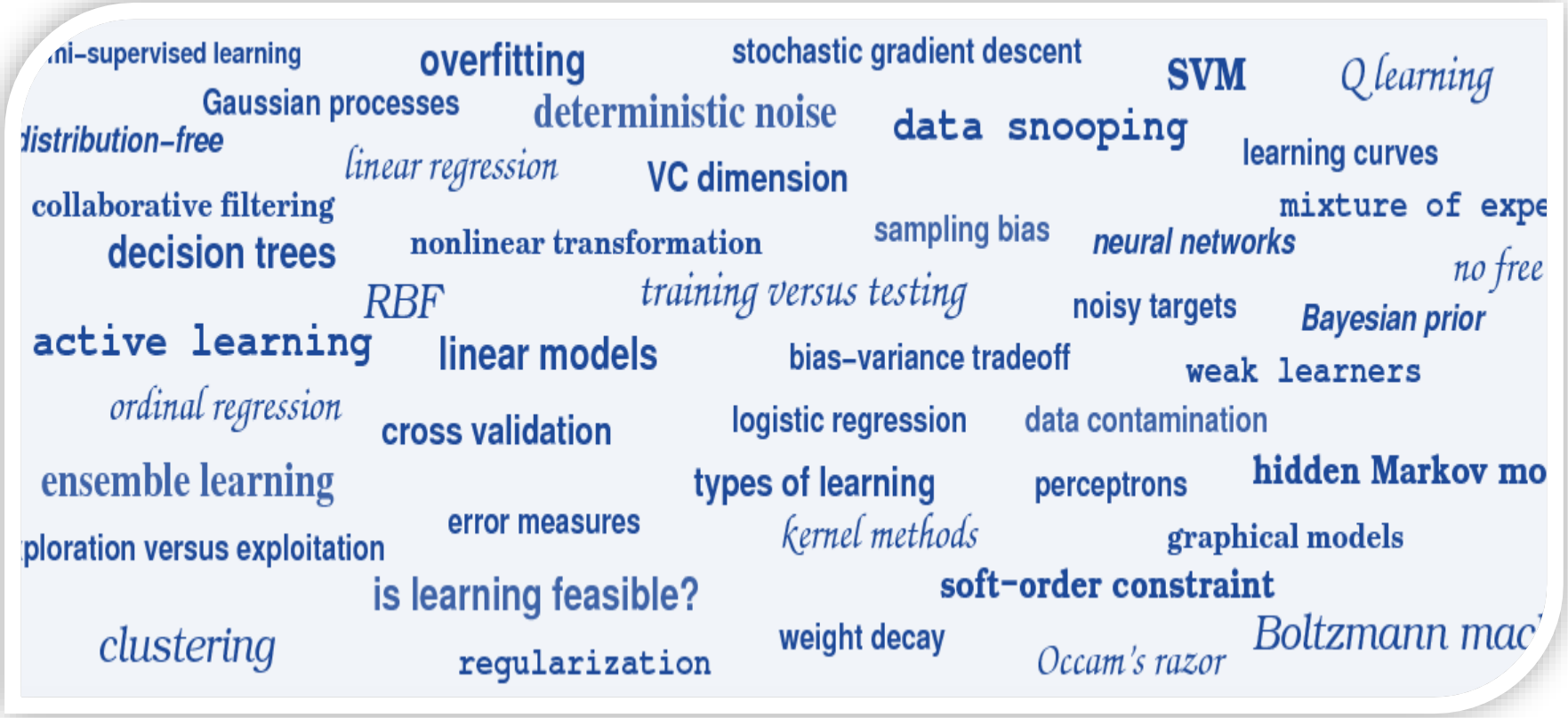

idea


BIG DATA
?
DECISIONS

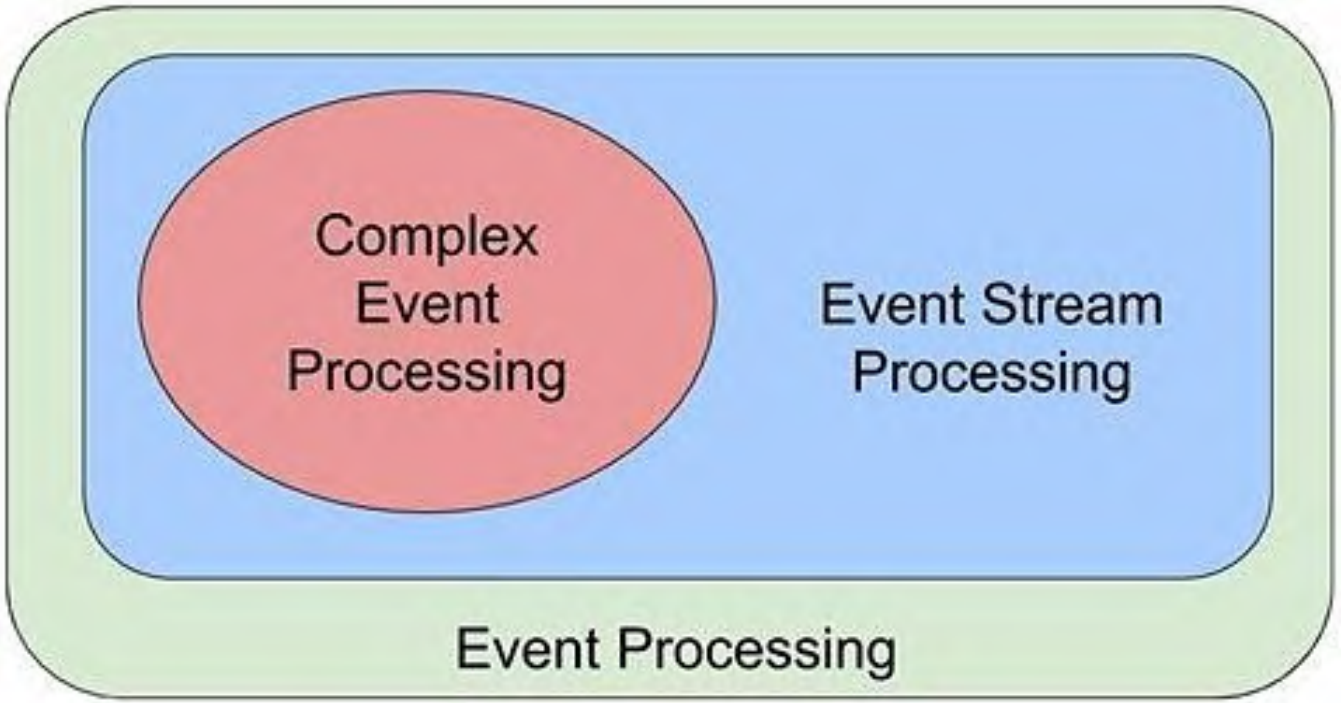
From Big Data to Decisions



From Big Data to Analytics to Decisions

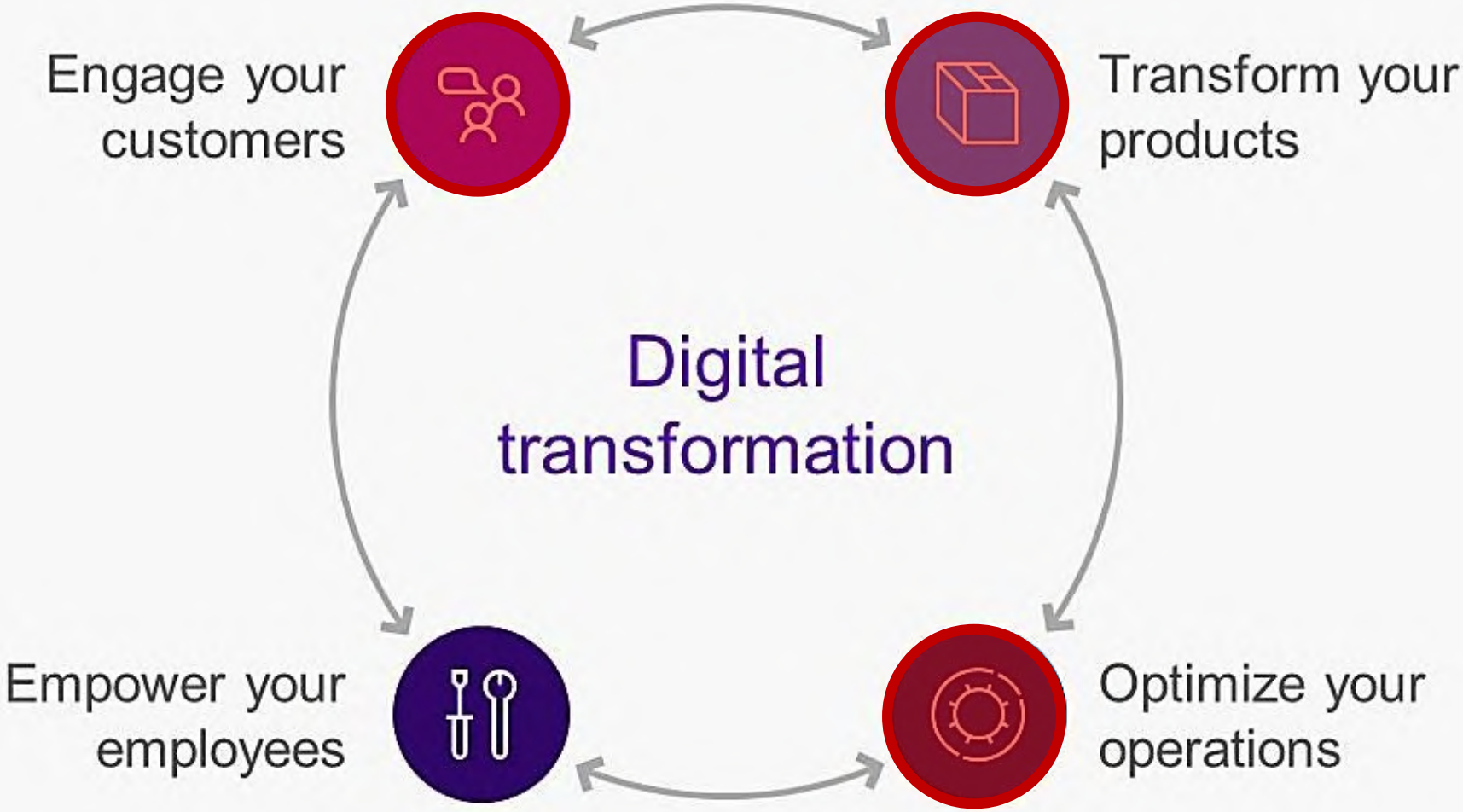


From (Big Data) Streams to Actions



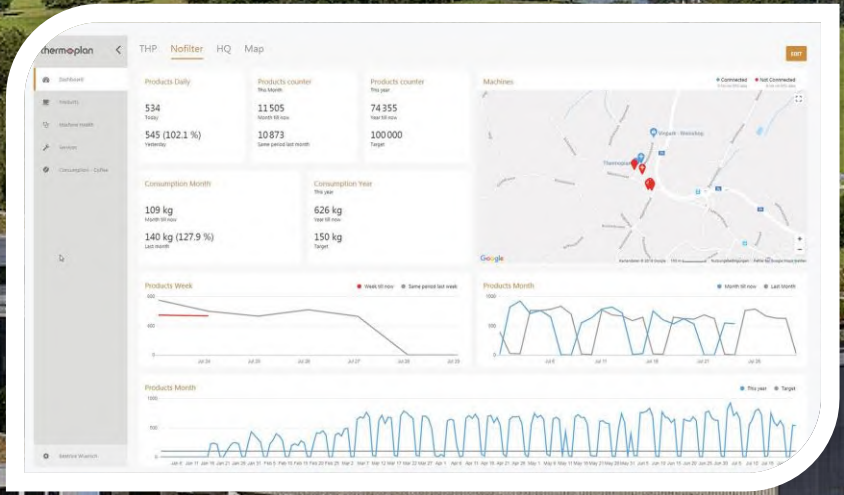
Beispiele

Big Data Use Cases


Engage your customers
Transform your products
Digital transformation
Empower your employees
Optimize your operations

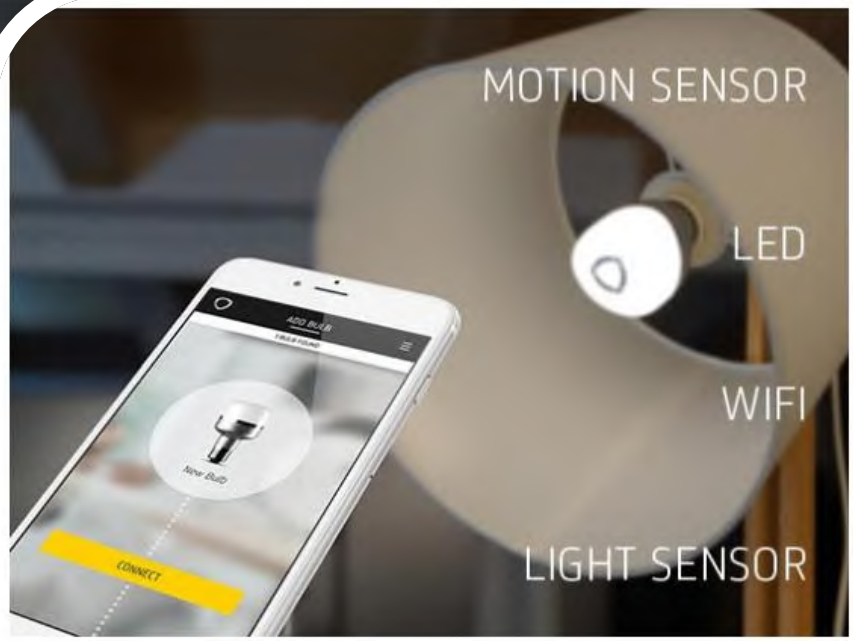

Transform Your Product
coffee

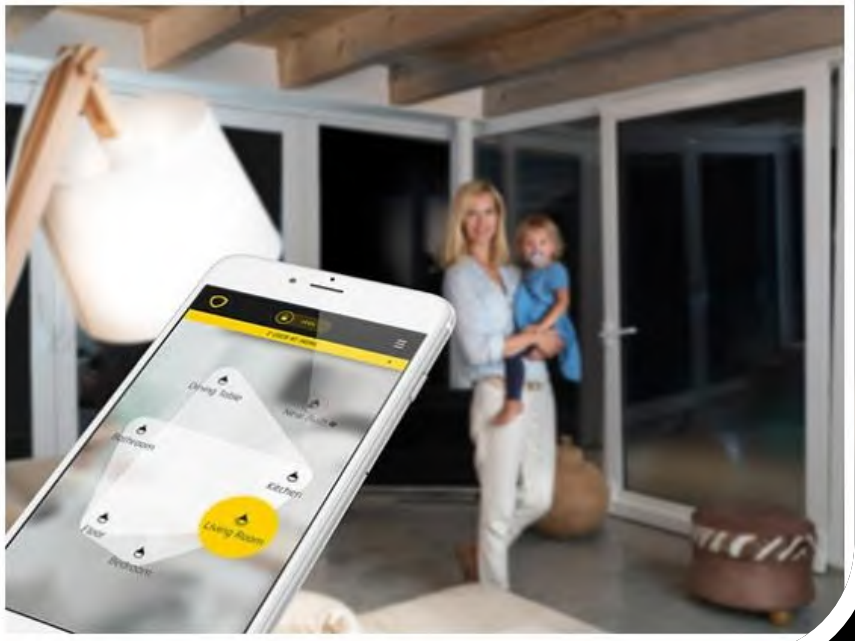
Internet of Things / Support Optimization & Recommendations




Transform Your Product

Internet of Things / Machine Learning





COMFYLIGHT


Optimize Your Operations
kWh
AEG
kWh
AEG
EW Minden-Ravensberg GmbH

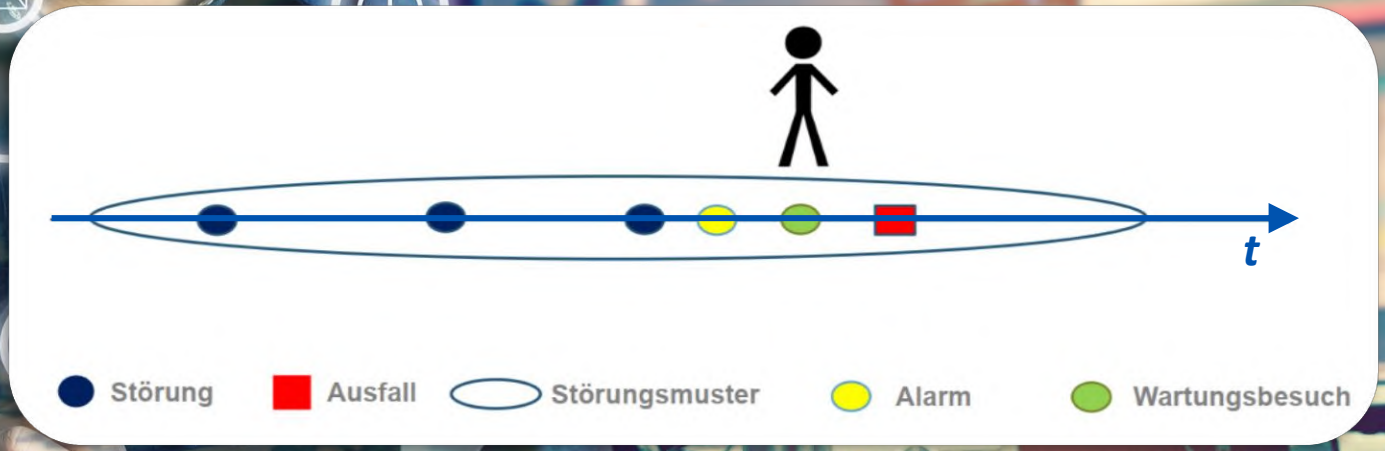
Smart Metering / Time Series Analysis & Recommendations

web.smart-me.com

smart-me

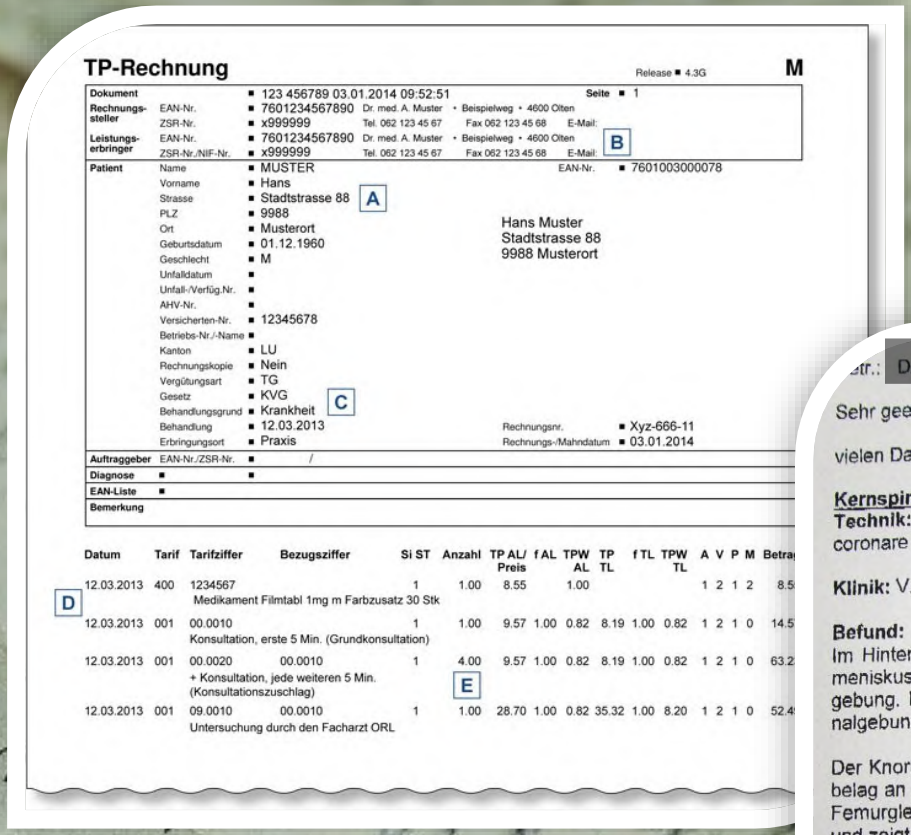
Optimize Your Operations

Predictive Maintenance / Complex Event Processing (CEP)



Optimize Your Operations

Text Analytics / Natural Language Processing (NLP)

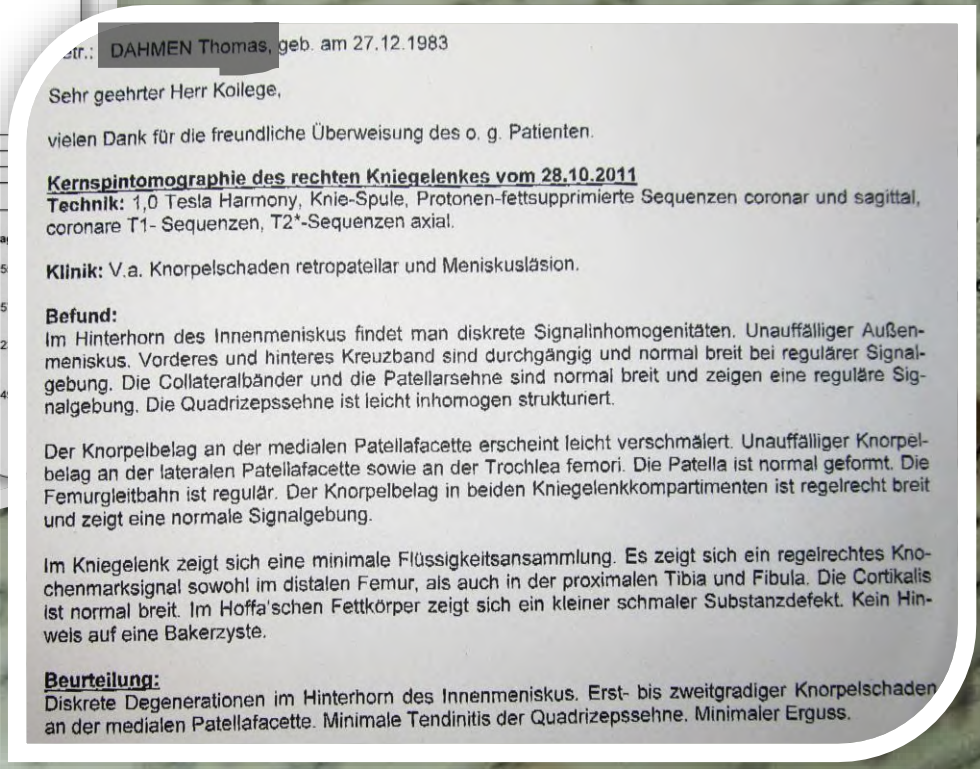


TP-Rechnung Release ■ 4.3G **M**

Dokument		■ 123 456789 03.01.2014 09:52:51		Seite ■ 1	
Rechnungs-steller	EAN-Nr.	■ 7601234567890	Dr. med. A. Muster • Beispielweg • 4600 Olten		
	ZSR-Nr.	■ x999999	Tel. 062 123 45 67	Fax 062 123 45 68	E-Mail:
Leistungs-erbringer	EAN-Nr.	■ 7601234567890	Dr. med. A. Muster • Beispielweg • 4600 Olten		B
	ZSR-Nr./NIF-Nr.	■ x999999	Tel. 062 123 45 67	Fax 062 123 45 68	E-Mail:
Patient	Name	■ MUSTER		EAN-Nr.	■ 7601003000078
	Vorname	■ Hans			
	Strasse	■ Stadtstrasse 88 A			
	PLZ	■ 9988			
	Ort	■ Musterort			
	Geburtsdatum	■ 01.12.1960			
	Geschlecht	■ M			
	Unfalldatum	■			
	Unfall-/Verfüg.Nr.	■			
	AHV-Nr.	■			
	Versicherten-Nr.	■ 12345678			
	Betriebs-Nr./-Name	■			
	Kanton	■ LU			
	Rechnungskopie	■ Nein			
	Vergütungsart	■ TG			
	Gesetz	■ KVG			
	Behandlungsgrund	■ Krankheit C			
	Behandlung	■ 12.03.2013		Rechnungsnr.	■ Xyz-666-11
	Erbringungsort	■ Praxis		Rechnungs-/Mahndatum	■ 03.01.2014
Auftraggeber	EAN-Nr./ZSR-Nr.	■ /			
Diagnose		■			
EAN-Liste		■			
Bemerkung					

Hans Muster
Stadtstrasse 88
9988 Musterort

Datum	Tarif	Tarifziffer	Bezugsziffer	Si ST	Anzahl	TP AL/ Preis	f AL	TPW AL	TP TL	f TL	TPW TL	A	V	P	M	Betra
12.03.2013	400	1234567		1	1.00	8.55		1.00				1	2	1	2	8.5
D Medikament Filmtabl 1mg m Farbzusatz 30 Stk																
12.03.2013	001	00.0010		1	1.00	9.57	1.00	0.82	8.19	1.00	0.82	1	2	1	0	14.5
Konsultation, erste 5 Min. (Grundkonsultation)																
12.03.2013	001	00.0020	00.0010	1	4.00	9.57	1.00	0.82	8.19	1.00	0.82	1	2	1	0	63.2
+ Konsultation, jede weiteren 5 Min. (Konsultationszuschlag) E																
12.03.2013	001	09.0010	00.0010	1	1.00	28.70	1.00	0.82	35.32	1.00	8.20	1	2	1	0	52.4
Untersuchung durch den Facharzt ORL																



tr.: DAHMEN Thomas, geb. am 27.12.1983

Sehr geehrter Herr Kollege,

vielen Dank für die freundliche Überweisung des o. g. Patienten.

Kernspintomographie des rechten Kniegelenkes vom 28.10.2011
Technik: 1,0 Tesla Harmony, Knie-Spule, Protonen-fettsupprimierte Sequenzen coronar und sagittal, coronare T1- Sequenzen, T2*-Sequenzen axial.

Klinik: V.a. Knorpelschaden retropatellar und Meniskusläsion.

Befund:
Im Hinterhorn des Innenmeniskus findet man diskrete Signalinhomogenitäten. Unauffälliger Außenmeniskus. Vorderes und hinteres Kreuzband sind durchgängig und normal breit bei regulärer Signalgebung. Die Collateralbänder und die Patellarsehne sind normal breit und zeigen eine reguläre Signalgebung. Die Quadrizepssehne ist leicht inhomogen strukturiert.

Der Knorpelbelag an der medialen Patellafacette erscheint leicht verschmälert. Unauffälliger Knorpelbelag an der lateralen Patellafacette sowie an der Trochlea femori. Die Patella ist normal geformt. Die Femurgleitbahn ist regulär. Der Knorpelbelag in beiden Kniegelenkkompartimenten ist regelrecht breit und zeigt eine normale Signalgebung.

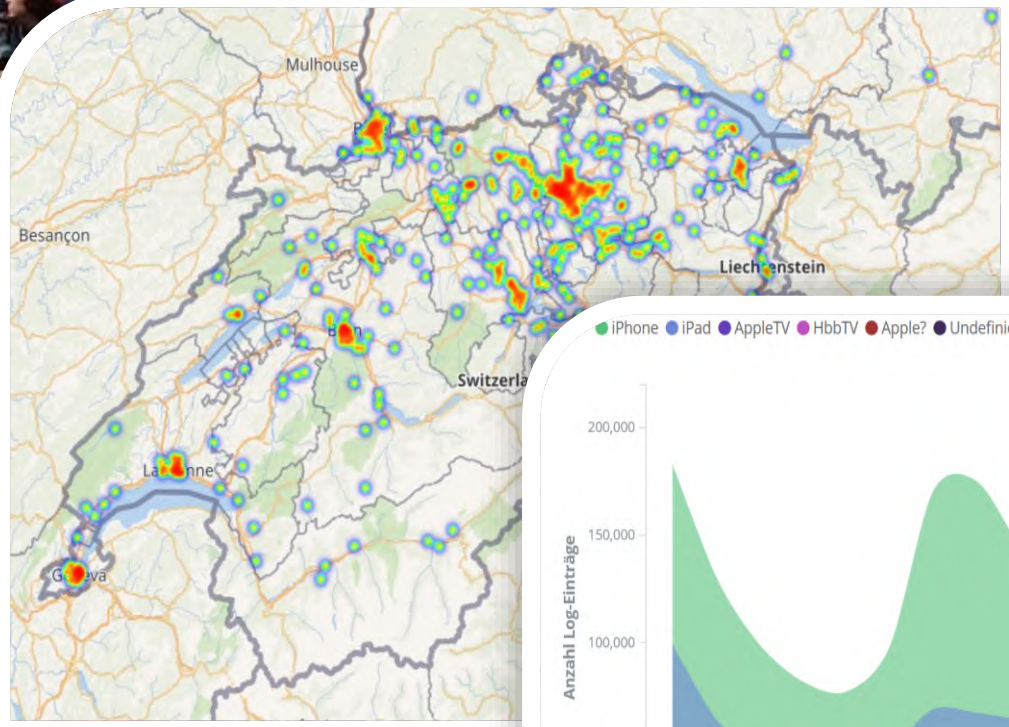
Im Kniegelenk zeigt sich eine minimale Flüssigkeitsansammlung. Es zeigt sich ein regelrechtes Knochenmarksignal sowohl im distalen Femur, als auch in der proximalen Tibia und Fibula. Die Cortikalis ist normal breit. Im Hoffa'schen Fettkörper zeigt sich ein kleiner schmaler Substanzdefekt. Kein Hinweis auf eine Bakerzyste.

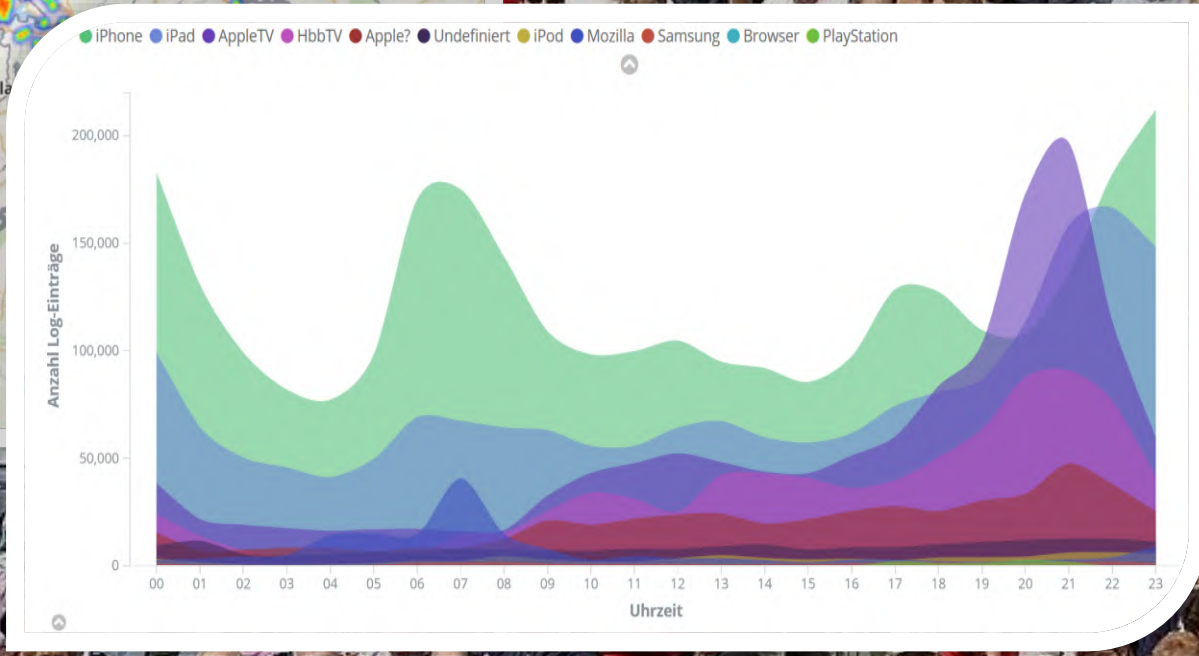
Beurteilung:
Diskrete Degenerationen im Hinterhorn des Innenmeniskus. Erst- bis zweitgradiger Knorpelschaden an der medialen Patellafacette. Minimale Tendinitis der Quadrizepssehne. Minimaler Erguss.

Engage Your Customers

Podcast Raw Data (Logs)

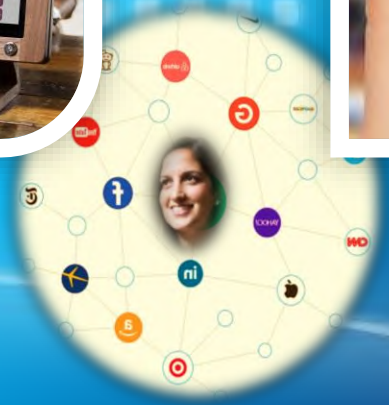
Podcast Usage Analysis





Engage Your Customers

Offline versus Online Shopping



Clickstream Analytics to Know Your Customer

Ausblick

Billionaires Innovation Leadership Money Business Small Busi

2,802 views | Nov 4, 2019, 10:33pm

Questioning The Long-Term Importance Of Big Data In AI

Rob Toews Contributor ⓘ
AI
I write about the big picture of artificial intelligence.

AlphaGo, the Go-playing artificial intelligence program developed by Google's DeepMind, defeated ... [+] ASSOCIATED PRESS

No asset is more prized in today's digital economy than data. It has become widespread to the point of cliche to refer to data as "the new oil." As one recent Economist headline put it, data is "the world's most valuable resource."

onaires Innovation Leadership Money Business Small B

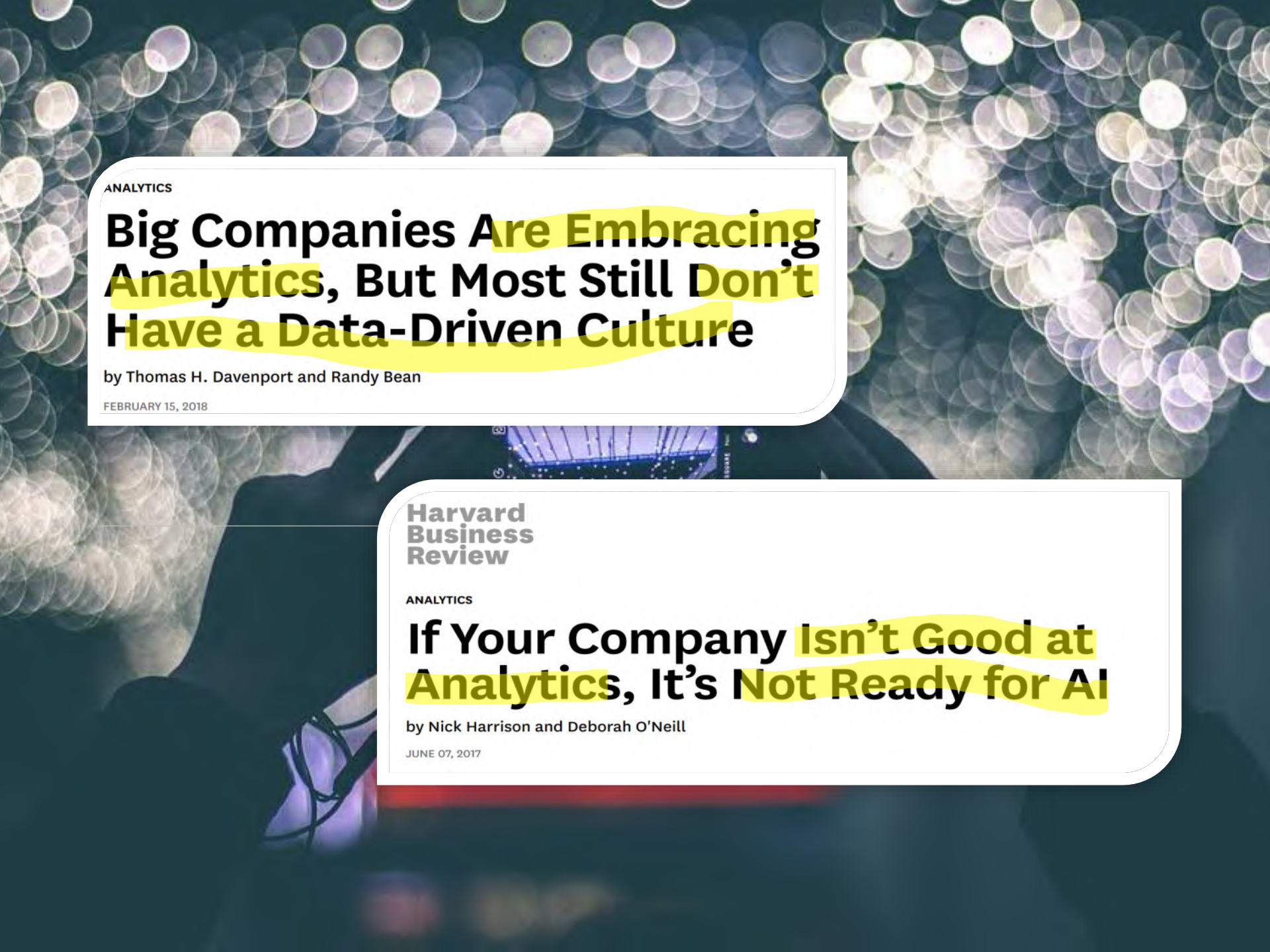
41,829 views | Nov 3, 2019, 01:18pm

Big Data Is Big Trouble For Science

Paul M. Sutter Contributor ⓘ
Science
Astrophysicist | Agent to the Stars

Look at all that data! FREE FOR COMMERCIAL USE (VIA PIXABAY)


SAFE
AI


ANALYTICS
Big Companies Are Embracing Analytics, But Most Still Don't Have a Data-Driven Culture
by Thomas H. Davenport and Randy Bean
FEBRUARY 15, 2018
Harvard Business Review
ANALYTICS
If Your Company Isn't Good at Analytics, It's Not Ready for AI
by Nick Harrison and Deborah O'Neill
JUNE 07, 2017


Big Data?
Transform
Your Thinking!

collaborative
innovative
thinking

Big Data: Auch Kleine können Gross!

Die geschickte Nutzung von Daten ist nicht allein den grossen Unternehmen vorbehalten.

KMU sind darin sogar besser, weil sie flexibler und pragmatischer agieren.

Daniel Benninger zeigt, worauf bei Big Data zu achten ist und wie auch kleine Unternehmen damit am Markt und bei ihren Kunden punkten können.

Dr. Daniel Benninger
Sawubona GmbH Bern & Hochschule Luzern

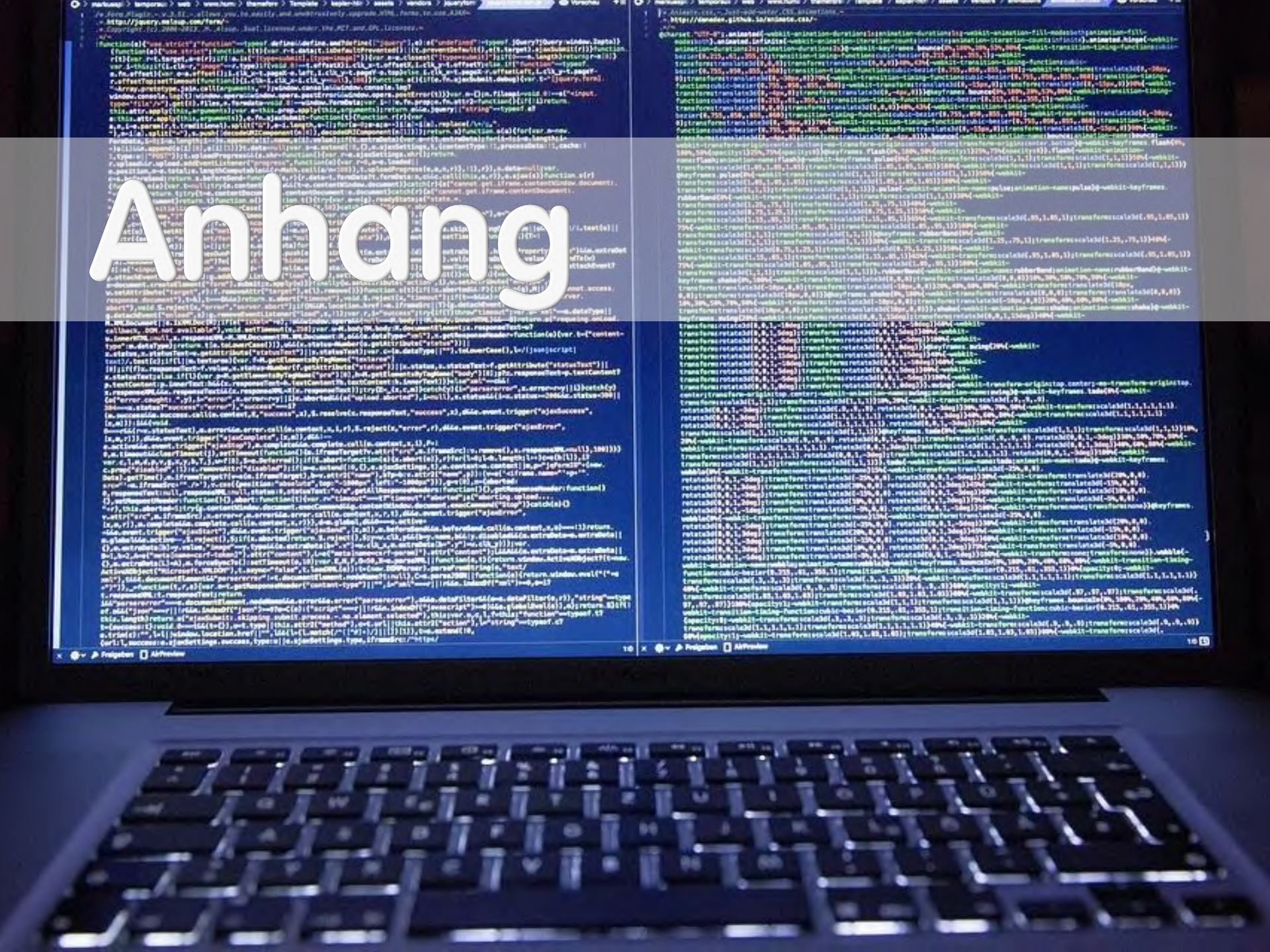
13.November 2019

Daniel Benninger - Kurzporträt

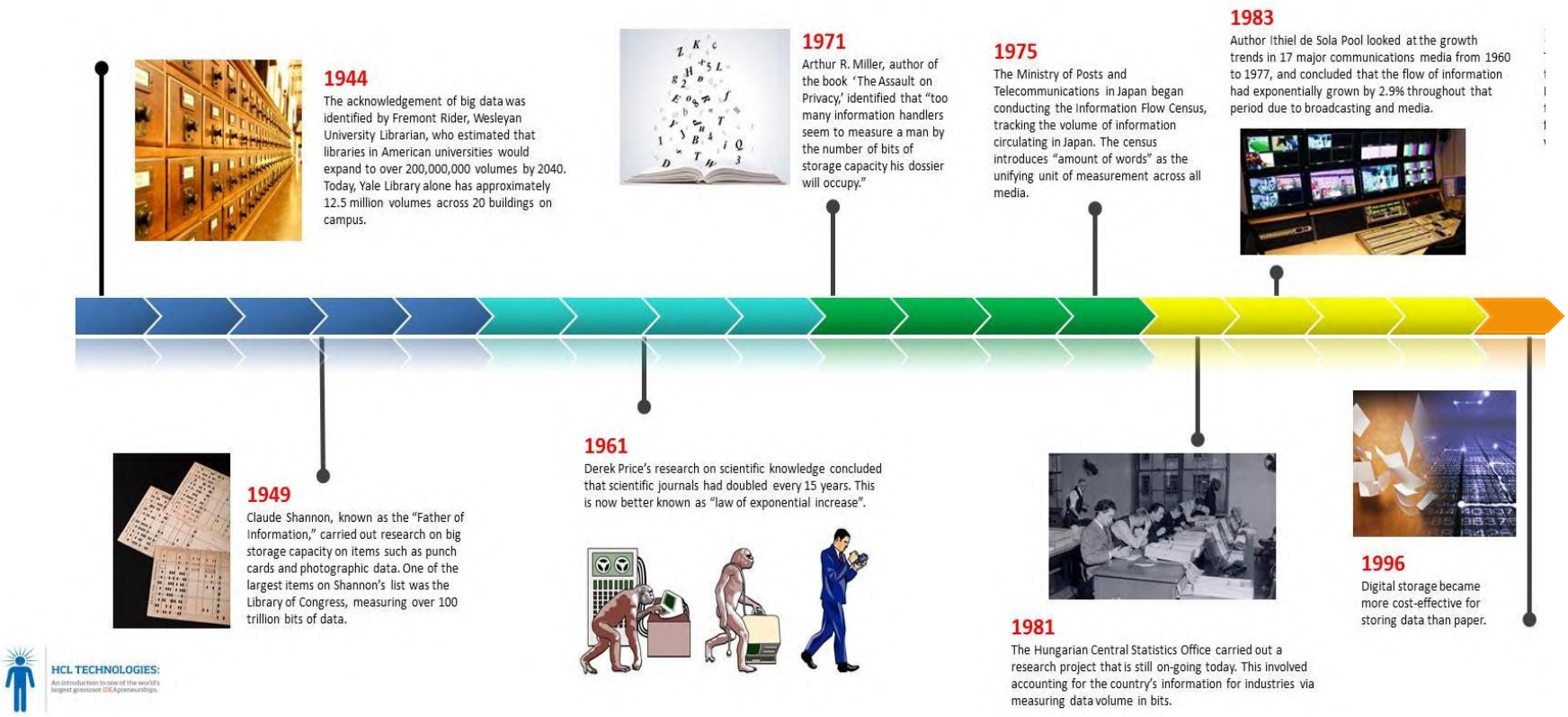
2018-	VR CVCube AG Gründungsmitglied/Vize-Präsident Digital Academy Switzerland
2015-	Managing Partner (Sawubona GmbH) Experte (Berner Fachhochschule), Reviewer MDPI Open Access Journals)
2014-	VR / Direktor ad interim (Sept'2016-Febr'2018) Schweiz. Teletext AG Dozent/Studienleiter (Hochschule Luzern)
2012-2015	Leiter ICT Eidg. Finanzmarktaufsicht FINMA
2008-2012	Head of IT Compliance/ICT Manager Swisscom (Schweiz) AG
2006-2008	Head of IT Management & Architecture Swisscom Fixnet AG
2003-2006	CEO db=ctq
2000-2003	Leiter IT Architektur Schweiz. Mobiliar Versicherungsgesellschaft
1998-2000	Head of Integration Solution Center Swisscom AG, Corporate Information & Technology
1994-1998	Project Manager, Solution Architect Digital Equipment Corporation (DEC)
1988-1994	Leiter Informatik / Partner, Sigmaplan AG
1978-1988	Dozent Software Schule Schweiz / Lehrbeauftragter Universität Bern Universität Bern, Studium Mathematik+Physik, Dr.phil.nat.

daniel.benninger@sawubona.ch
+41 79 334 54 70

Sawubona


Anhang

Timeline – 70 Years of Big Data (1/2)



Quelle: www.hcltech.com/blogs/history-big-data

Timeline – 70 Years of Big Data (2/2)

1997
The term "big data" was used for the first time when researchers M. Cox and D. Ellsworth wrote an article identifying that the rise of data was becoming an issue for current computer systems. In other words, the "problem of big data."

2000
Peter Lyman and Hal R. Varian published the first study that quantified, in computer storage terms, the total amount of new and original information created in the world annually. The study concluded in 1999, a year in which the world had produced approximately 1.5 exabytes of unique information.

2001
Doug Laney published a research note titled "3D Data Management: Controlling Data Volume, Velocity, and Variety." A decade later, these "3 V's" became the defining dimensions of big data.

2005
"What is Web 2.0" was published by writer Tim O'Reilly, in which he stated that "data is the next Intel inside and SQL is the new HTML. Database management is a core competency of Web 2.0 companies, so much so that they have sometimes referred to these applications as 'infoware' rather than merely software."

2008
Economists Bret Swanson and writer George Gilder published "Estimating the Exaflood" which stated U.S. IP traffic could reach one zettabyte by 2015, and that the U.S. Internet of 2015 would be at least 50 times larger than it was in 2006.

2009
Researchers Roger E. Bohn and James E. Short found that modern Americans consumed information for an average of almost 12 hours per day. Consumption totalled 3.6 zettabytes and 10,845 trillion words, corresponding to 100,500 words and 34 gigabytes for an average person on an average day.

2011
Martin Hilbert and Priscila Lopez, estimated that the world's information storage capacity grew at a compound annual growth rate of 25% per year between 1986 and 2007. They also estimated that in 1986, 99.2% of all storage capacity was analog, but by 2007, 94% of all storage capacity was digital.

2013/Present
Businesses are beginning to implement new in-memory technology such as SAP HANA to analyse and optimize mass quantities of data. Companies are becoming ever more reliant on utilizing data as a business asset to gain competitive advantage, with big data leading the charge as arguably the most important new technology to understand and make use of in order to remain relevant in today's rapidly changing market.

Big Data – A Brief History (1/3)

November 1967
B. A. Marron and P. A. D. de Maine publish "**Automatic data compression**" in the Communications of the ACM, stating that "*The 'information explosion' noted in recent years makes it essential that storage requirements for all information be kept to a minimum.*"

1981
The Hungarian Central Statistics Office starts a research project to *account for the country's information industries, including measuring information volume in bits*. The research continues to this day.

September 1990
Peter J. Denning publishes "**Saving All the Bits**" in American Scientist. "*The imperative [for scientists] to save all the bits forces us into an impossible situation: The rate and volume of information flow overwhelm our networks, storage devices and retrieval systems, as well as the human capacity for comprehension*"

1996 - Digital storage becomes more cost-effective for storing data than paper.

1997
Michael Lesk publishes "**How much information is there in the world?**"
Lesk concludes that "*There may be a few thousand petabytes of information all told; and the production of tape and disk will reach that level by the year 2000.*"

Big Data – A Brief History (2/3)

August 1999
Steve Bryson, David Kenwright, Michael Cox, David Ellsworth, and Robert Haimes publish "**Visually exploring gigabyte data sets in real time**" in the Communications of the ACM**. It is the first CACM article to use the term "Big Data"**. The article opens with the following statement: "*Very powerful computers are a blessing to many fields of inquiry. They are also a curse; fast computations spew out massive amounts of data. Where megabyte data sets were once considered large, we now find data sets from individual simulations in the 300GB range. But understanding the data resulting from high-end computations is a significant endeavor. As more than one scientist has put it, it is just plain difficult to look at all the numbers. And as Richard W. Hamming, mathematician and pioneer computer scientist, pointed out, the purpose of computing is insight, not numbers.*"

November 2000
Francis X. Diebold presents to the Eighth World Congress of the Econometric Society a paper titled **"Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting**" in which he states "*Recently, much good science, whether physical, biological, or social, has been forced to confront—and has often benefited from—the "Big Data" phenomenon. Big Data refers to the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology*."

February 2001
Doug Laney, an analyst with the Meta Group, publishes a research note titled "**3D Data Management: Controlling Data Volume, Velocity, and Variety.**" *A decade later, the "3Vs" have become the generally-accepted three defining dimensions of big data, although the term itself does not appear in Laney's note.*

Big Data – A Brief History (3/3)

September 2008
A **special issue of *Nature* on Big Data** "*examines what big data sets mean for contemporary science*."

December 2008
Randal E. Bryant, Randy H. Katz, and Edward D. Lazowska publish "**Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society**" They write: "*Big-data computing is perhaps the biggest innovation in computing in the last decade. We have only begun to see its potential to collect, organize, and process data in all walks of life*."

February 2010
Kenneth Cukier publishes in The Economist a Special Report titled, "**Data, data everywhere**." "*…the world contains an unimaginably vast amount of digital information which is getting ever vaster more rapidly… The effect is being felt everywhere, from business to science, from governments to the arts. Scientists and computer engineers have coined a new term for the phenomenon: 'big data.'*"

May 2011
James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers of the McKinsey Global Institute publish "Big data: The next frontier for innovation, competition, and productivity." They estimate that "*by 2009, nearly all sectors in the US economy had at least an average of 200 terabytes of stored data (twice the size of US retailer Wal-Mart's data warehouse in 1999) per company with more than 1,000 employees*" In total, the study estimates that 7.4 exabytes of new data were stored by enterprises and 6.8 exabytes by consumers in 2010.