
Die Bedeutung von Big Data für mein Unternehmen

Was geht mich das an?

Erwartungen

- Sie möchten die Bedeutung von Big Data für Ihr Unternehmen wissen.
 - Die 129.000 Ergebnisse von Google genügen Ihnen offenbar nicht.
 - Die unzähligen Artikel in Zeitungen und Zeitschriften befriedigen Sie auch nicht.
 - Sie erwarten, dass ich die Antwort weiss, obwohl ich weder Sie noch Ihr Unternehmen kenne.
-

Wir stellen uns vor

... wie immer, aber "anderscht"

Mein Unternehmen & Big Data

- Zu sich: Wer bin ich (beruflich)?
 - Was habe ich studiert, gelernt, usw.?
 - Was mache ich beruflich?
 - Zu seinem Unternehmen
 - Was macht das Unternehmen (raison d'être)?
 - Mein Beitrag zum Erfolg des Unternehmens?
 - Zu (Big) Daten
 - Was interessiert, bzw. fasziniert mich an Daten?
 - Meine Erfahrung mit Daten, bzw. mit Big Data?
-

Richard & seine Unternehmen

- (Aus)bildung
 - MS Mathematik (USA, Austauschjahr ETHZ)
 - Germanistik (USA, Wien; abgebrochen vor MA)
 - Berufe: Informatiker (“ungelernt”)
 - Ciba-Geigy AG: Analytiker für OR (Unternehmensforschung)
 - Friedrich Miescher-Institut: Instruktor für Bio-Informatik
 - UBS AG: Programmierer/Analytiker, Gruppe für zukunftsorientierte Technologien, Solution Engineer Document Management, Content-Architekt Intranet/Internet
 - Roche: Machine Learning, Knowledge Engineering
-

Richard & (Big) Data

- Erfahrung
 - OR: Bedarfsprognosen → Produktionsplanung, Simulationen
 - Bio-Informatik: Der Mensch als Summe seiner Genen, Gendatenbanken
 - (Content-Architekt: Aufbereitung von Wissen für die maschinelle Verarbeitung)
 - Maschinelles Lernen, Semantik
- Persönliche Bedeutung
 - Entscheidungen Anderer betreffen mich. Ihre Daten?
 - Ich als Datensatz (Surf-, Kaufverhalten, Gesundheitsdaten, Meinungen)
- Berufliche Interessen
 - Text in Wissen umwandeln (Text Mining, Machine Learning, NLP)
 - Hochdimensionale Daten “schonend” analysieren (topologische Datenanalyse)

Der Chef empfiehlt

- Ist (Big) Data wirklich so neu?
 - **Eine kurze Geschichte von (Big) Data**

- Wieso ist Big Data plötzlich in aller Mund?
Mash-up zweier Präsentationen der Bloor Group
 - **Think Big: How to Design a Big Data Information Architecture**
 - **The Inevitable Shift: How Big Data Impacts Enterprise Architecture**

- Was macht man mit Big Data?
Auszüge zweier Präsentationen des Fraunhofer-IAIS
 - **Big Data – Vorsprung durch Wissen**
Chancen erkennen und nutzen
 - **Living Lab Big Data**
Konzeption einer Experimentierplattform

- Schlussfolgerungen

Eine kurze Geschichte von (Big) Data

Die Motivation

Big Data ist in aller Munde (PwC)

- **Big Data – Ein Weg zur digitalen Wertschöpfungsstrategie?**
 - „Das Thema Big Data gewinnt durch die rasant ansteigenden Datenmengen für Unternehmen zunehmend an Bedeutung.“
 - „In Zukunft werden sich Unternehmen verstärkt mit der Frage befassen, wie sie die ‚verborgenen Schätze‘ verwerten und Daten als strategische Ressource nutzen können.“
 - „Vielzahl von zusätzlichen Datenquellen [z.B.] soziale Netzwerke, internetbasierte Kommunikation, mobile Internetnutzung, Geodaten, etc.“
 - Empfehlung: „sich zunächst sehr genau fragen: Welche strategischen Fragestellungen soll die Datenanalyse ‚beantworten‘?“

Big Data ist in aller Munde (HBR)

□ **Big Data: The Management Revolution**

- “Simply put, because of big data, managers can measure, and hence know, radically more about their businesses, and directly translate that knowledge into improved decision making and performance.”
- “We expect companies that were born digital to accomplish things that business executives could only dream of a generation ago. But in fact the use of big data has the potential to transform traditional businesses as well.
- “As the tools and philosophies of big data spread, they will change long-standing ideas about the value of experience, the nature of expertise, and the practice of management.”
- “Isn't 'big data' just another way of saying 'analytics'?” → No: Volume, Velocity, Variety
- “The more companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results.”

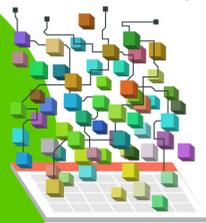
40 ZETTABYTES

[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005



It's estimated that 2.5 QUINTILLION BYTES

[2.3 TRILLION GIGABYTES]
of data are created each day



Volume SCALE OF DATA

6 BILLION
PEOPLE
have cell
phones



WORLD POPULATION: 7 BILLION

Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
[161 BILLION GIGABYTES]



**30 BILLION
PIECES OF CONTENT**
are shared on Facebook
every month



Variety DIFFERENT FORMS OF DATA



By 2014, it's anticipated there will be
**420 MILLION
WEARABLE, WIRELESS
HEALTH MONITORS**

**4 BILLION+
HOURS OF VIDEO**
are watched on
YouTube each month



400 MILLION TWEETS
are sent per day by about 200
million monthly active users



The New York Stock Exchange captures
**1 TB OF TRADE
INFORMATION**
during each trading session



Modern cars have close to
100 SENSORS
that monitor items such as
fuel level and tire pressure

Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected
there will be
**18.9 BILLION
NETWORK
CONNECTIONS**
— almost 2.5 connections
per person on earth



**1 IN 3 BUSINESS
LEADERS**
don't trust the information
they use to make decisions



Poor data quality costs the US
economy around
\$3.1 TRILLION A YEAR



**27% OF
RESPONDENTS**

Veracity UNCERTAINTY OF DATA

in one survey were unsure of
how much of their data was
inaccurate



Eindrücke

- Was man versteht:
 - Big Data sei etwas völlig Neues, etwas nie zuvor da Gewesenes – halt revolutionär.
 - Das Potential von Big Data sei gewaltig, doch man müsse rasch handeln.
 - Die Konkurrenz schlafe nicht.
- Vor dem Revolutionären fühle ich mich orientierungslos. Wo kein Hinten ist, ist auch kein Vorne. Ich komme mir hilflos vor. Ich bin jedem ausgeliefert, der besser zu wissen behauptet.
- Das Evolutionäre hingegen ist wie ein Film. Verstehe ich etwas nicht, kann ich zurückspulen und dessen Entwicklung wiedersehen ... und verstehen.

Die Geschichte

- “If you want to understand today, you have to search yesterday.” – Pearl Buck
 - “In order to predict the future, understand how the past became the present. That is more or less how today will become tomorrow.” – Anonymous
 - Seit 1978 bin ich in der Informatik tätig und habe viele „Next Big Things“ kommen und gehen sehen. Sie waren fast ausnahmslos revolutionär, unternehmenskritisch, usw.
 - Die Informatik ist nicht so jung, dass ihre Geschichte nicht relevant wäre, uns (Informatiker) nichts lehren könnte.
-

Eine kurze Geschichte von (Big) Data

(Big) Data im Laufe der Zeitalter

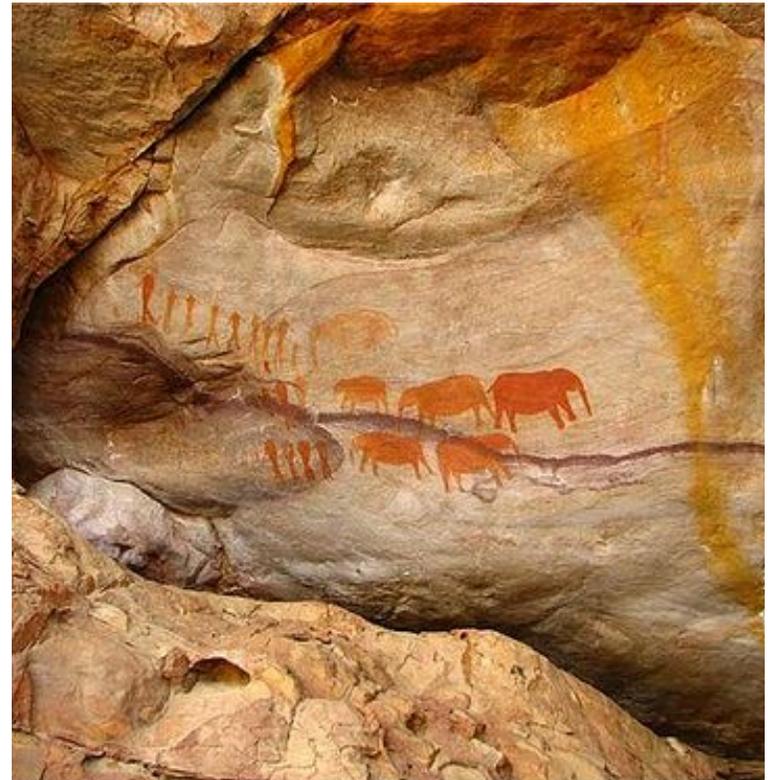
(Big) Data im Laufe der Zeitalter

- Seit wann gibt es den Begriff Data?
 - Seit wann gibt es Daten?
 - Wie hat sich der Umgang mit Daten im Laufe der Zeitalter entwickelt?
-

Der Begriff „Data“

- Herkunft Latein *dare* „geben“ → *datum* „gegeben“ (*data* Mehrzahl)
- Im 17. Jh. kommt „datum“ über die Philosophie ins Englische (und wohl gleichzeitig ins Deutsche) und bezeichnet etwas, was als wahr angenommen wird, d.h. eine Gegebenheit, eine Annahme (→ Mathematik).
- „Data“ beschränken sich nicht auf Numerisches, Messbares, wie oft gemeint wird.
- „Data“ gibt es nicht erst seit der Erfindung des elektronischen Rechners.

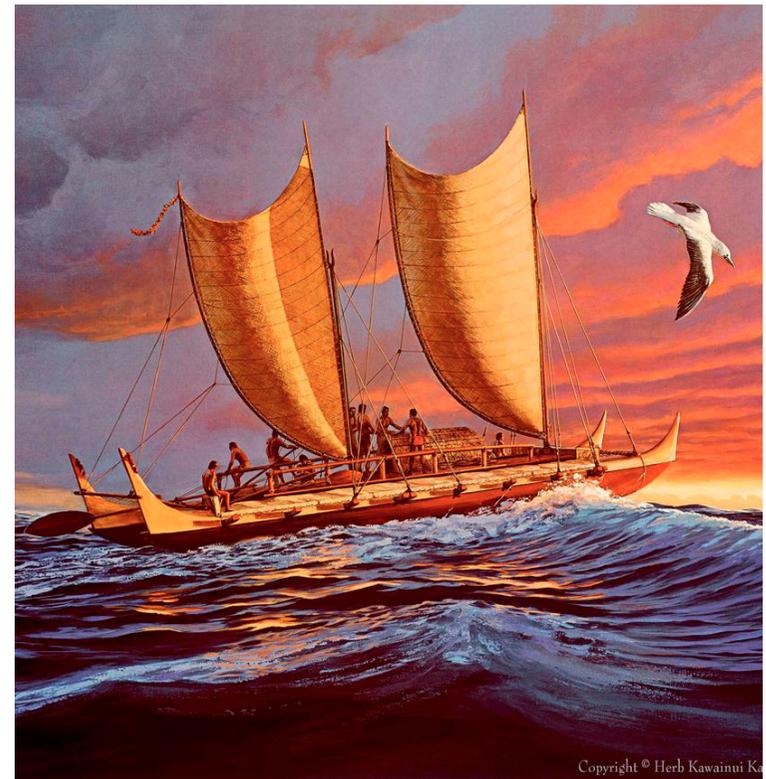
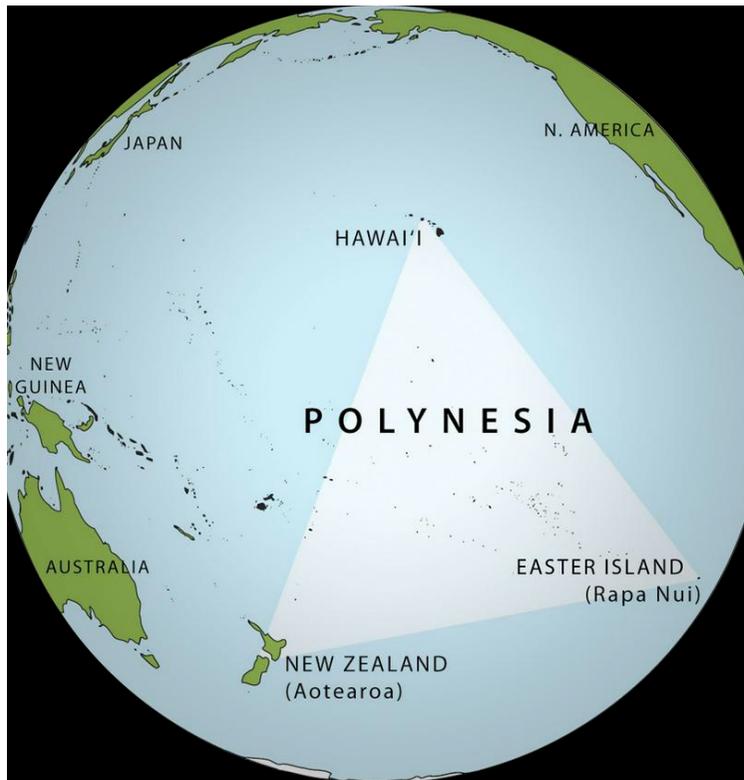
Chauvet-Grotte, Südfrankreich



Chauvet-Grotte: Eine Datenbank

- Auswirkung von Speichermedium & Aufzeichnungstechnik
 - Unpraktisch, Kopien zu erstellen
 - Häufige Rückkehr zu den Höhlen
 - Begrenzter Aktionsradius. Existenzbedrohend?
 - Häufige Erfassung dürfte die Qualität der Daten erhöht haben
 - Erinnerungen waren in mehreren Köpfen noch frisch
 - Big Data?
-

Die Polynesier

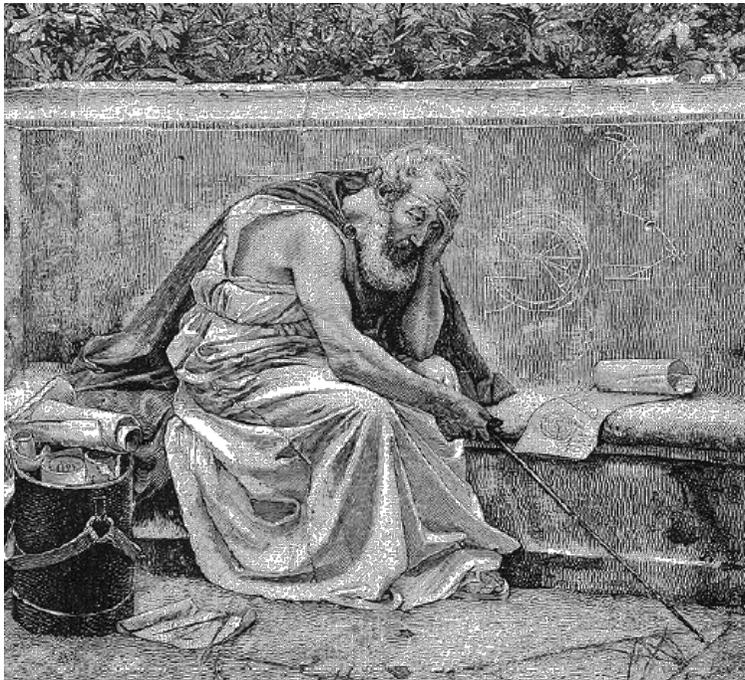


Die Polynesier: Mobile Datenbanken

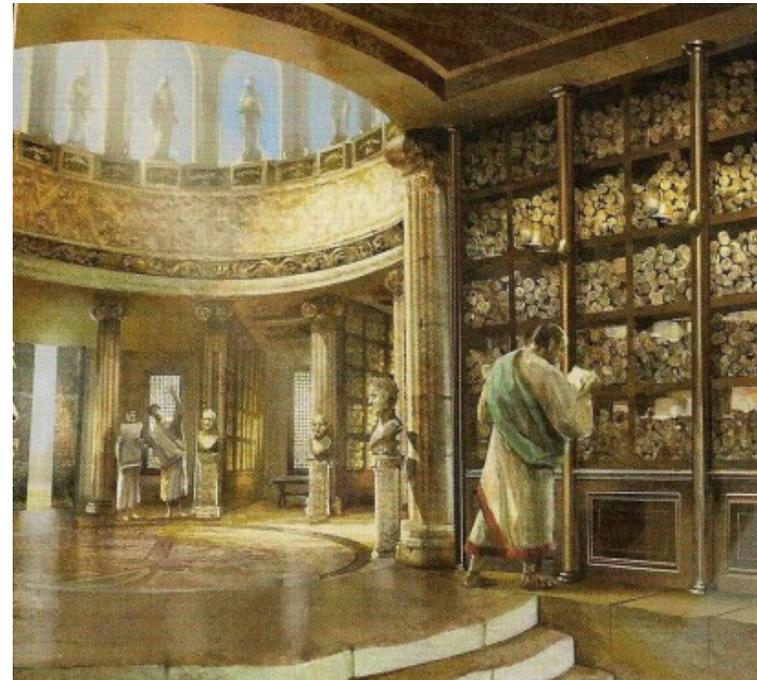
- Sie bewahrten ihr gesamtes Kulturerbe in Erzählungen (*chants*) auf, die skandiert und getanzt wurden.
 - Redundanz: Daten + poetisches „Korsett“ + Tanzsprache
 - Mehrfache Speicherung
 - Navigationskenntnisse wurden *oriori* (Navigationspriestern) anvertraut und von ihnen an Auserwählte mündlich weitergegeben.
 - Erhöht die Zuverlässigkeit der Daten
 - Tragbare Datenbanken, die nach Bedarf aktualisiert und repliziert wurden.
 - Big Data?
-

Alexandria ca. 245 v. Chr.

Eratosthenes v. Kyrene



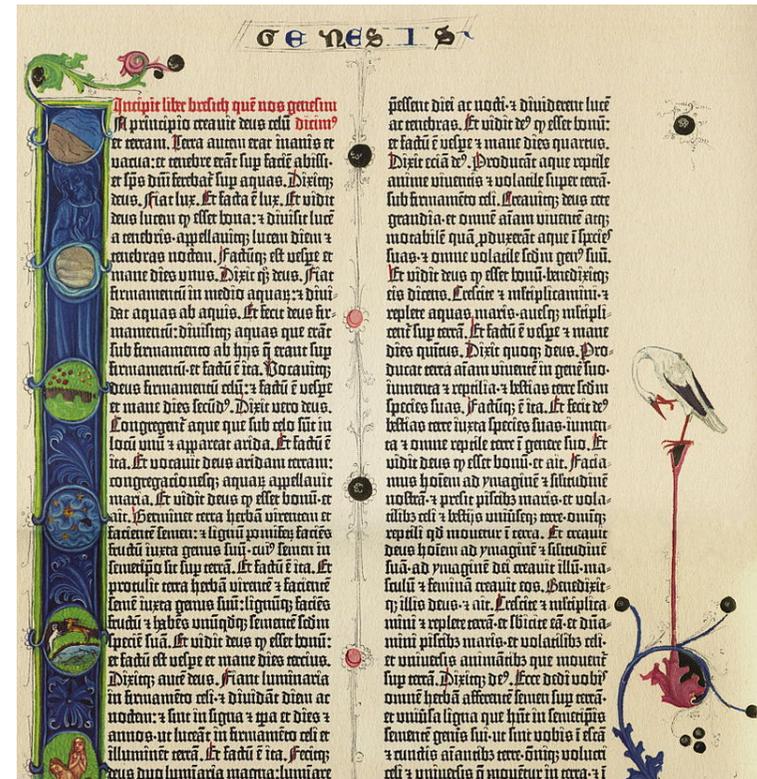
Museum & Bibliothek



Die Bibliothek von Alexandria

- Die Bibliothek war eine Sammlung von Daten aus allen Lebensbereichen.
 - Das Museum wurde errichtet, weil immer mehr Wissen produziert und konsumiert wurde. Forscher konsumieren und produzieren Wissen.
 - Bestehendes Wissen musste übersetzt und transkribiert werden. Oft gab es mehrere Versionen desselben Dokuments.
 - Big Data?
-

Johannes Gutenberg (~1400 – 1468)



Die Erfindung des Buchdrucks

- Demokratisierte die Verbreitung und die Entdeckung von Wissen
 - Die Massenproduktion von gedruckten Büchern
 - Begünstigte die Weitergabe von Informationen und Ideen
 - Förderte die Entstehung von Interessengemeinden
 - Beschleunigte die Produktion von (Fach-)wissen
 - Erst im 20. Jh. entstand die Befürchtung, dass die Bibliotheken die schieren Mengen an Drucksachen bald nicht mehr bewältigen würden.
 - Schon im 17. Jh. erkannten die ersten Bibliothekswissenschaftler, dass der Bestand einer Bibliothek katalogisiert werden muss.
-

Informationsexplosion

- Die Auswertung der US-Volkzählung von 1880 dauerte acht Jahre. Unter gleich bleibenden Umständen würde die Zeit für die Auszählung der nächste Volkzählung nicht reichen.
 - 1932 erlebten die USA einen Bevölkerungsboom.
 - Erforderte gründlichere, besser organisierte Akten.
 - Führte zu mehr Forschung.
 - Bis 1940 konnten die Bibliotheken die Nachfrage nach Publikationen neuerer Forschungsergebnisse kaum decken.
-

Informationsexplosion

- **The Scholar and the Future of the Research Library**, F. Reiter, 1944:
 - Amerikanische Universitätsbibliotheken verdoppelten alle 16 Jahre ihre Bestände
 - Bis 2040 werde Yale 200 Millionen Bände besitzen, die mehr als 10.000 Km Regale und über 6.000 Fachkräfte für das Katalogisieren allein benötigten
 - **Science since Babylon**, D. Price, 1961:
 - Die Anzahl Fachzeitschriften und -zeitungen verdoppele sich alle 15 Jahre.
 - Das wissenschaftliche Wissen wachse exponentiell.
-

Die Geschichte lehrt uns ...

- Daten gibt es schon lange vor der Erfindung des Computers.
 - Für die damaligen Verhältnisse haben unsere Vorfahren genauso mit Volume, Velocity, Variety und Veracity gekämpft wie wir heute. Was gross, hoch oder vielfältig ist, ist relativ. Big Data ist wohl kaum jünger als Data selbst.
 - Wir können uns in Ruhe überlegen, ob uns Big Data etwas bringt.
-

THINK BIG: HOW TO DESIGN A BIG DATA
INFORMATION ARCHITECTURE

THE INEVITABLE SHIFT: HOW BIG DATA
IMPACTS ENTERPRISE ARCHITECTURE

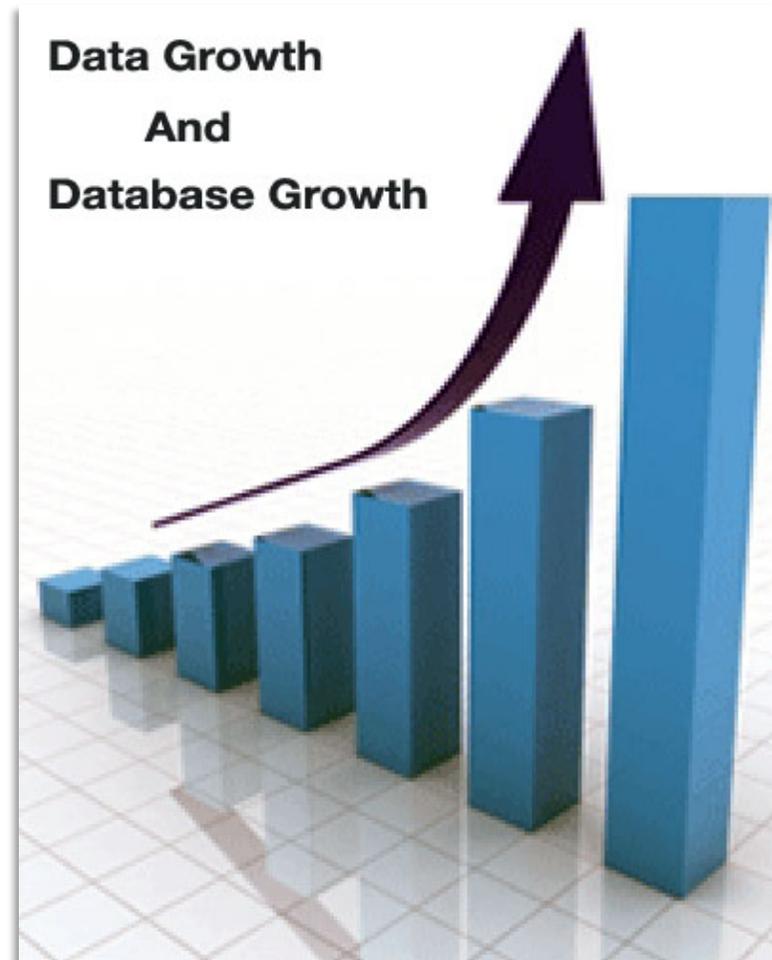
Robin Bloor & Eric Kavanagh, The Bloor Group

Big Data - A Poorly Defined Term



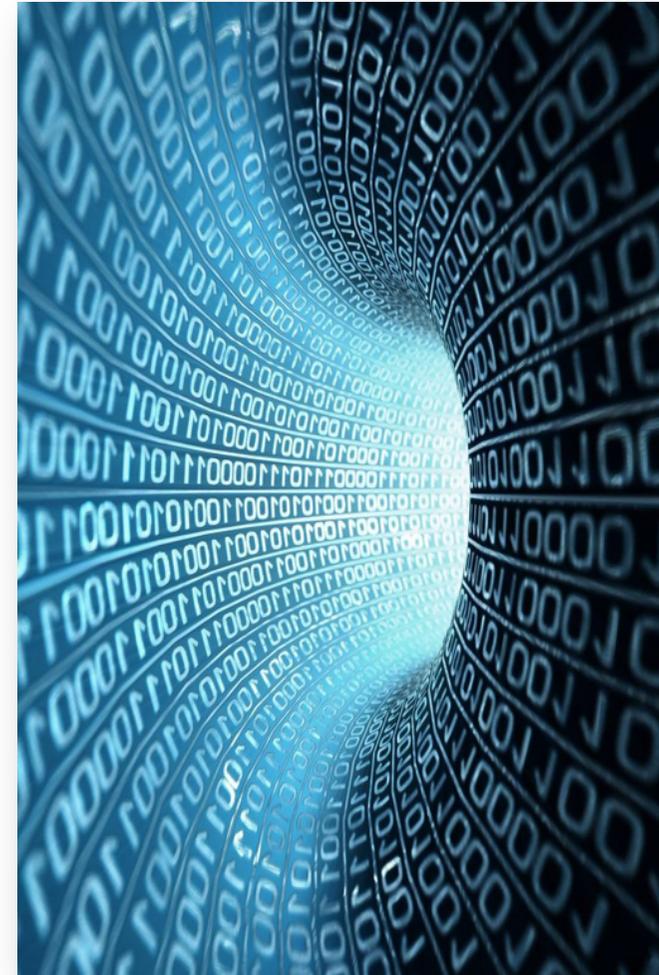
The Visible “Big Data” Trend

- ◆ Corporate data volumes grow at about **55% per annum** - exponentially
- ◆ Data has been growing at this rate for, maybe, 40 years
- ◆ There is nothing new about big data. It clings to an established **exponential trend**

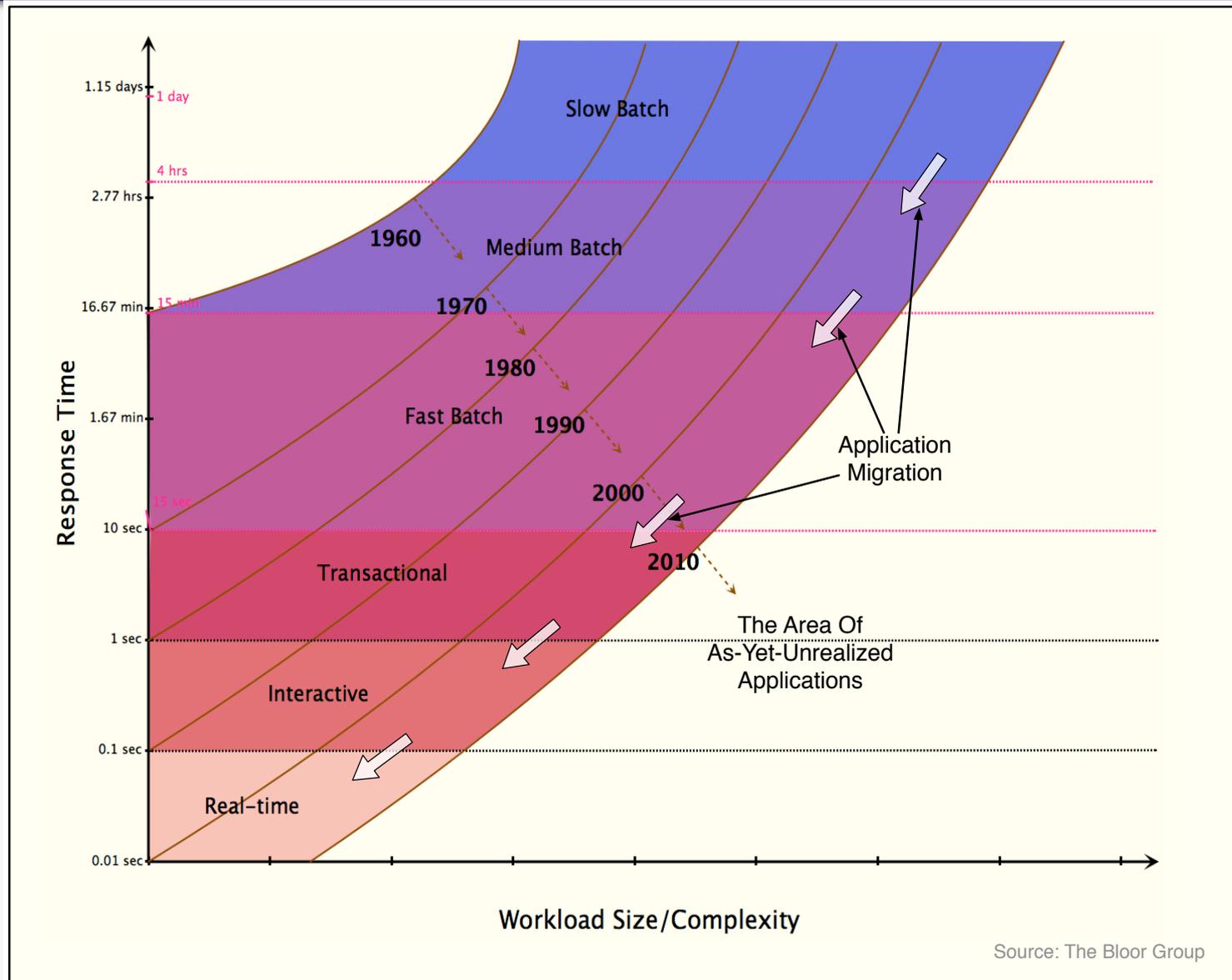


The Invisible Trend: Moore's Law Cubed

- ◆ The biggest databases are **new databases**
- ◆ They grow at the cube of Moore's Law
- ◆ Moore's Law = **10x every 6 years**
- ◆ VLDB: 1000x every 6 years
 - 1991/2 megabytes
 - 1997/8 gigabytes
 - 2003/4 terabytes
 - 2009/10 petabytes
 - **2015/16 exabytes**

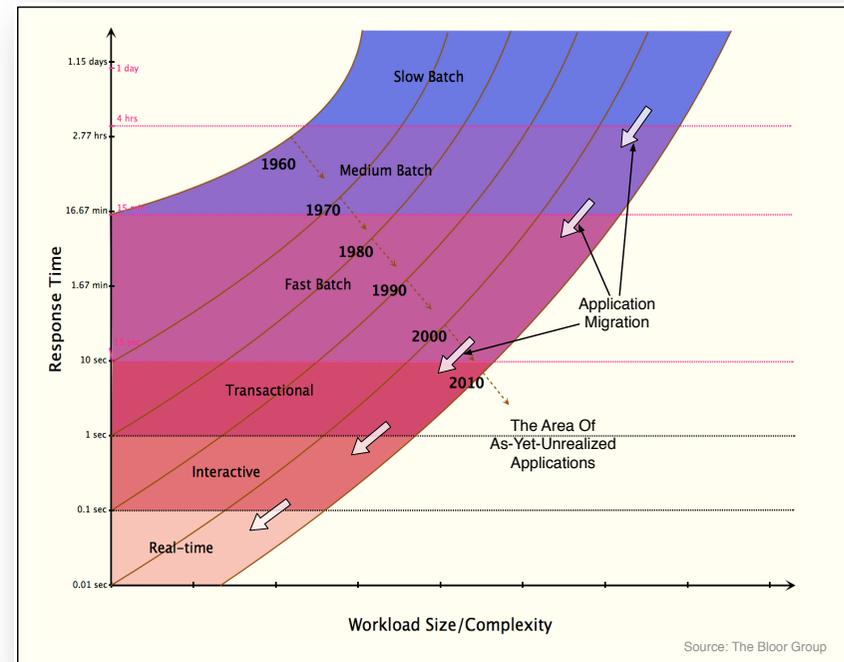


Technology Evolution (Bloor Curve)



The Traditional Force of Disruption

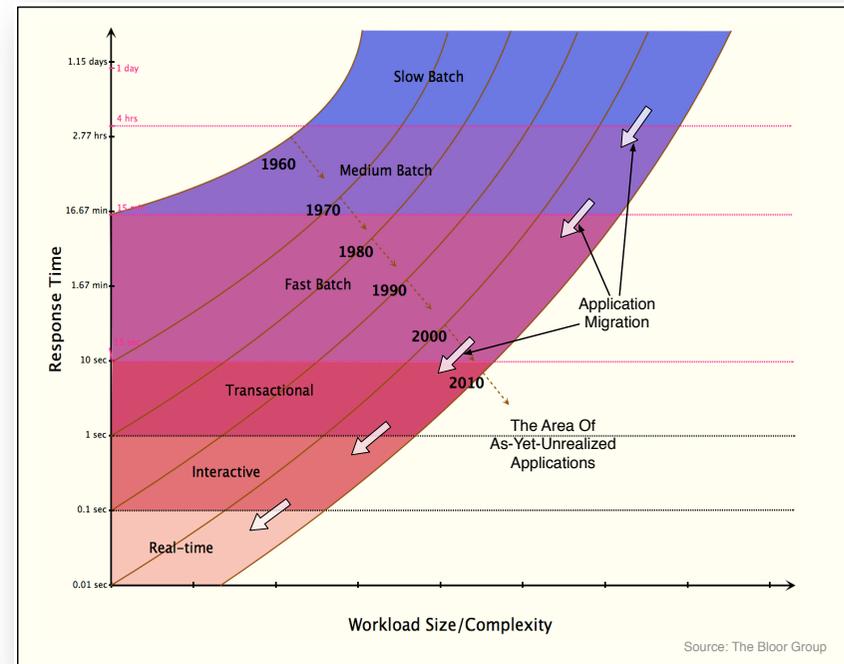
- ◆ **Software architectures change:** centralized, C/S, 3 tier/web, SOA, etc.
- ◆ Applications **migrate** according to latencies
- ◆ Dominant applications and software brands can die via “**The innovator’s dilemma**”
- ◆ Wholly new applications appear **because of lower latencies**, e.g., VMs, CEP



This Curve is Compromised

Two DISRUPTIVE forces have changed the curve:

PARALLELISM
and
The CLOUD



Parallelism: The Imp is Out of the Bottle

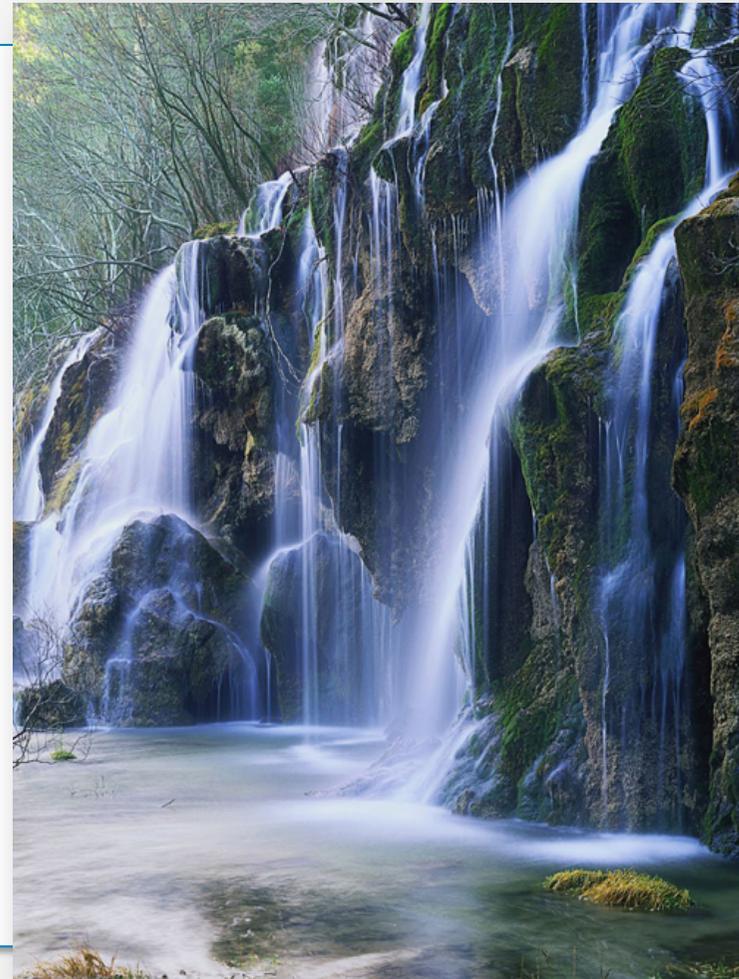
- ◆ Multicore chips enabled **parallelism**
- ◆ It has changed the whole **performance** equation
- ◆ It enabled **Big Data**
- ◆ Big Data is really **Big Processing**



Memory is Becoming Hierarchical Store

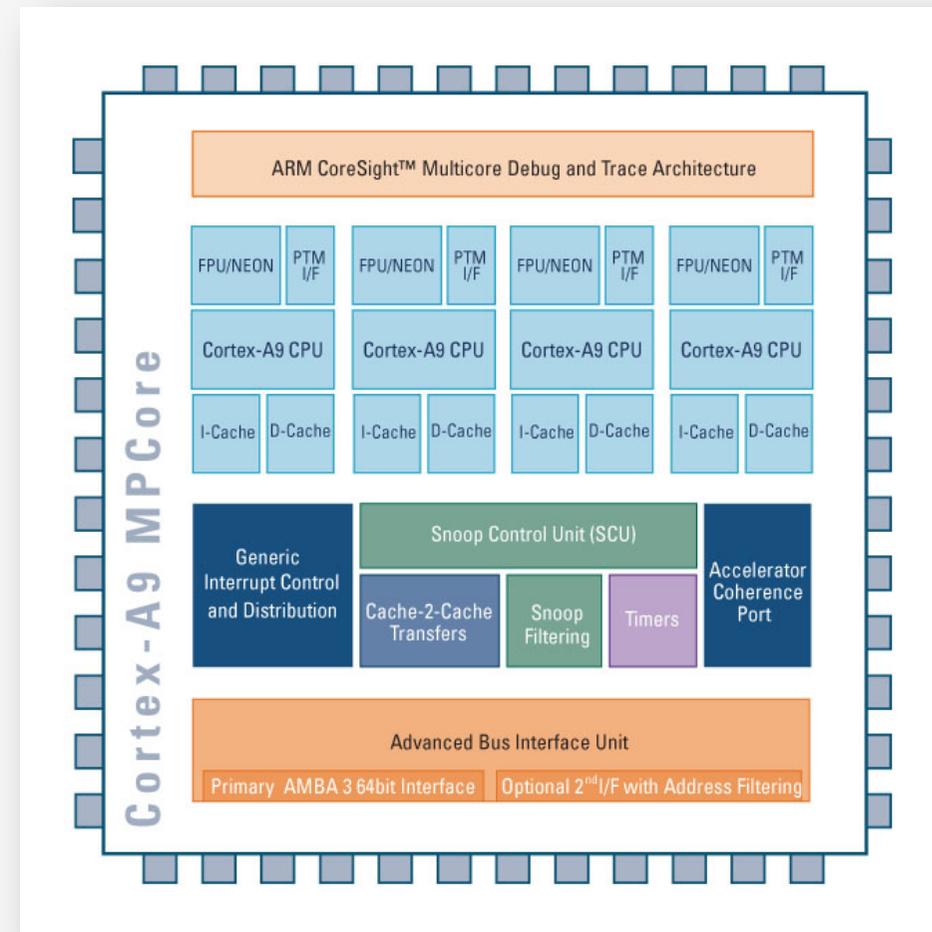
- ◆ On chip speed v RAM
 - L1(32K) = 100x
 - L2(246K) = 30x
 - L3(8-20Mb) = 8.6x
- ◆ RAM v SSD
 - RAM = 300x
- ◆ SSD v Disk
 - SSD = 10x
- ◆ Disk will soon **turn up its toes**

Note: Vector instructions and data compression



Putting a SoC in IT

- ◆ It's possible that the **CPU/Memory** split will vanish, possibly soon
- ◆ This requires the emergence of the **commodity SoC**
- ◆ There are already **ARM SoCs** that run Linux
- ◆ Grids of SoCs would **replace** grids of servers



Big Data???

It's not really about

BIG DATA

It's about

BIG PROCESSING

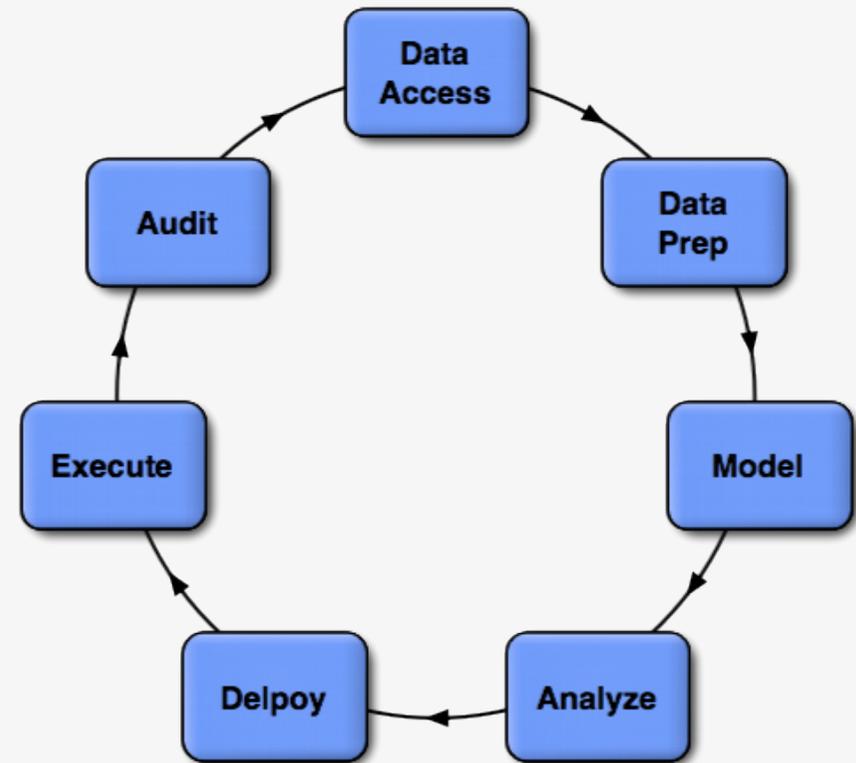
The Big Data Applications

It's pretty much
all about

BI & ANALYTICS

A Process, Not an Activity

- ◆ Data Analytics is a multi-disciplinary end-to-end **process**
- ◆ Until recently it was a walled-garden. But recently the **walls were torn down** by...
 - Data availability
 - Scalable technology
 - Open source tools

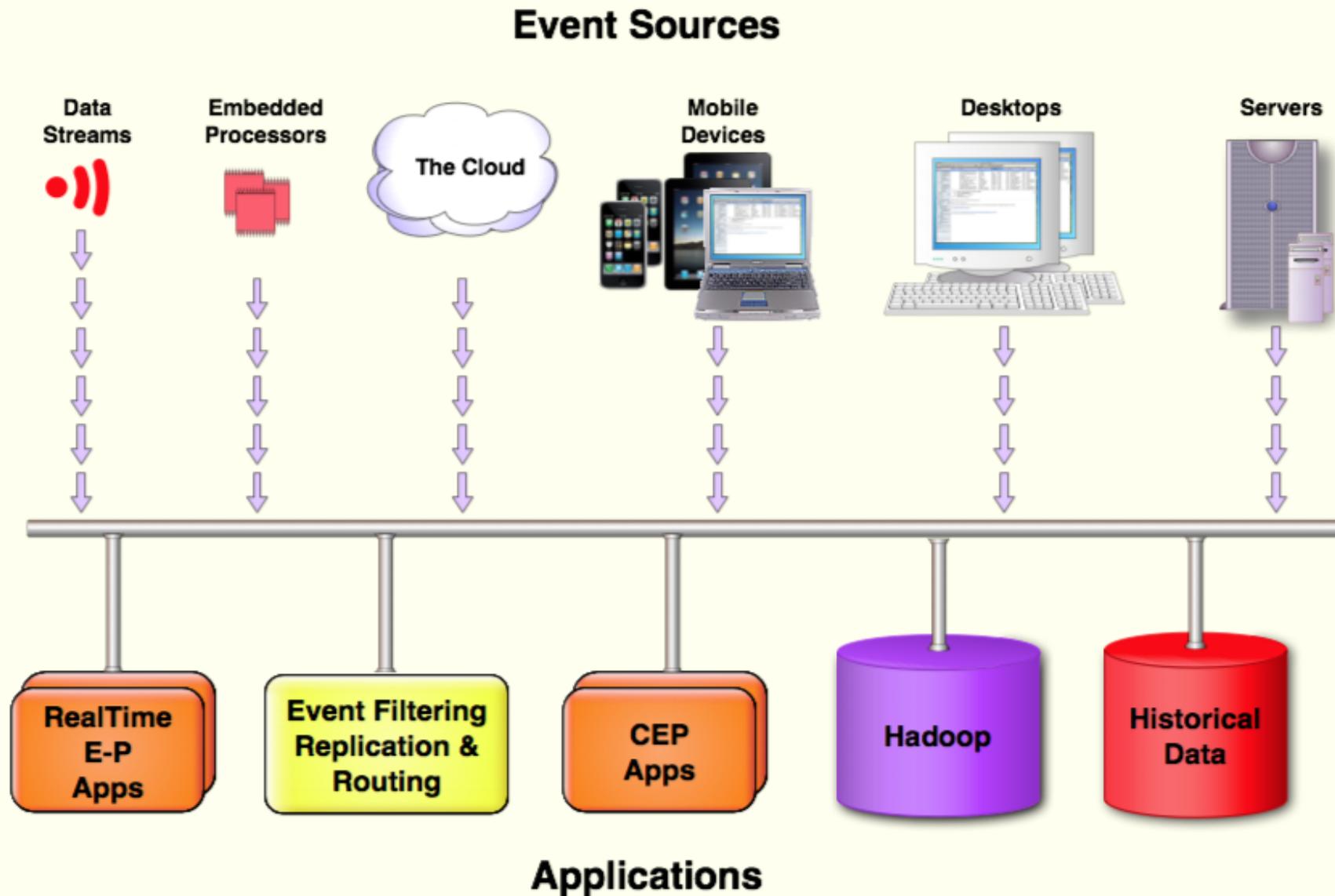


Atoms and Molecules

The **ATOM** of data has
become the **EVENT**

A **TRANSACTION** is a
MOLECULE of **ATOMIC**
EVENTS

Event Driven/Big Data Architecture?



Atoms and Molecules

DATA FLOW

is becoming a driving factor



This suggests the need
for a

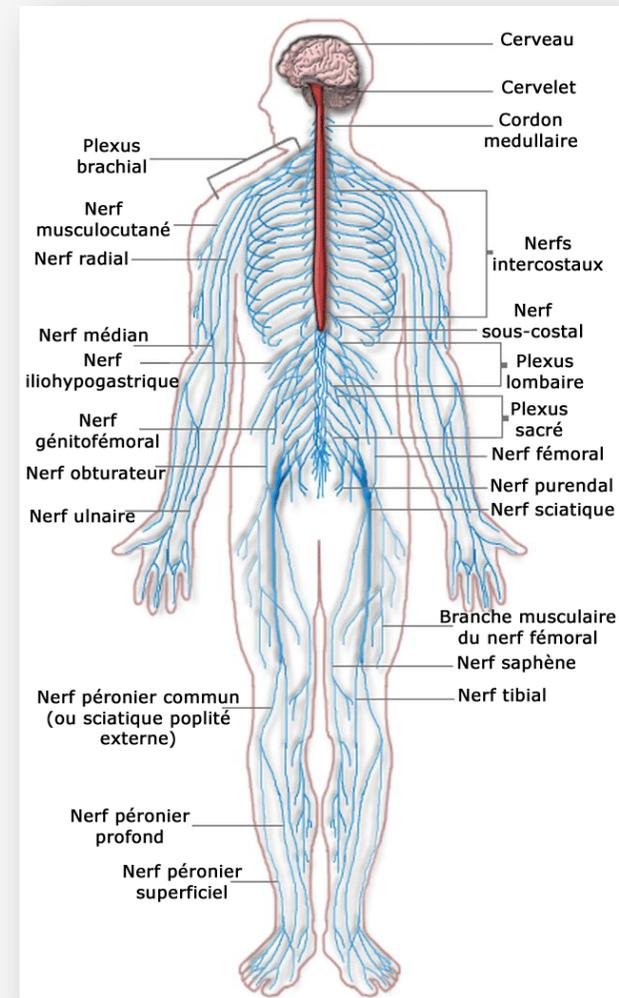
DATA RESERVOIR



The Bloor Group

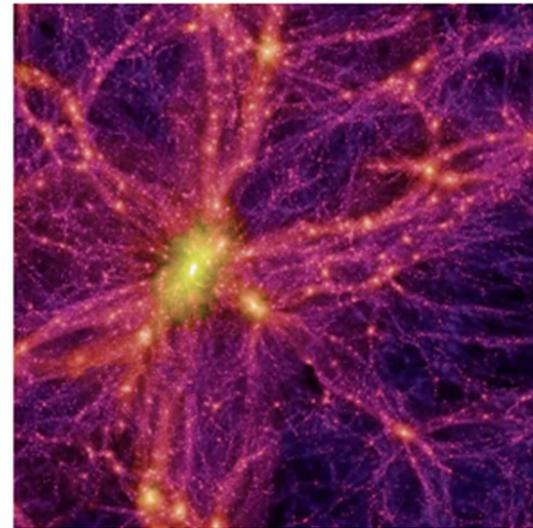
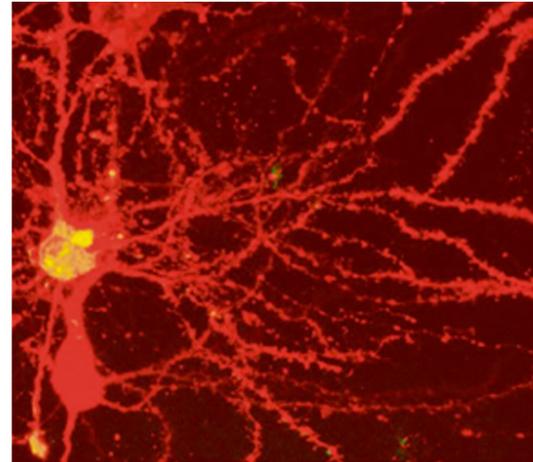
The Biological System

- ◆ Our human control system works at different speeds:
 - Almost instant reflex
 - Swift response
 - Considered response
- ◆ Organizations will gradually implement similar control systems
- ◆ This suggests a data-flow-based architecture
- ◆ The EDW is memory



The Corporate Biological System

- ◆ Right now this division into two different data flows is already occurring
- ◆ Currently we can distinguish between:
 - Real-time/Business-time applications
 - Analytical applications
- ◆ We should build specific architectures for this



The CRITICAL Workload Issue

- ◆ Previously, we viewed database workloads as an **i/o optimization problem**
- ◆ With analytics the workload is a very **variable mix** of i/o and calculation
- ◆ No databases were built precisely for this - **not even Big Data databases**



Some Architectural Principles

- ◆ The new atom of data is **the event**
- ◆ SUSO, scale up before scale out
- ◆ **Take the processing to the data**, if you can
- ◆ Hadoop is a component **not a solution**



Big Data – Vorsprung durch Wissen

Chancen erkennen und nutzen

Prof. Dr. Stefan Wrobel, Fraunhofer IAIS

Big Idea

Store forever...

- Daten werden heute für zukünftige Auswertungen gespeichert
- Daten werden als Produktionsfaktor genutzt

and use better...

- Datenanalyseverfahren werden immer schneller und effizienter
- Es gibt leistungsstarke Werkzeuge zur Speicherung und Auswertung

➔ Big Data is here to stay!

- Big Data schafft neue Geschäftsmodelle, Produkte und Dienstleistungen – selbst für kleinste Unternehmen
- Big Data ist kein rein technisches Konzept, sondern wird zu einem Schlüsselfaktor in der weltweiten digitalen Revolution in der Wirtschaft

Über das Projekt



Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

Fördergeber

Bundesministerium für Wirtschaft und Technologie,
Laufzeit Juli 2012 bis Dezember 2012

Projekthalte

- Erstellung einer Markt- und Potentialanalyse für Big Data
 - Aufzeigen konkreter Handlungsoptionen für Wirtschaft, Politik und Forschung
 - Zielgruppe: Führungskräfte und Entscheider aus Wirtschaft und Politik
- Konzeption und Aufbau einer Big Data Experimentierplattform
 - Aufzeigen der technische Machbarkeit und des potentiellen Nutzens von Big Data für den Wirtschafts- und Forschungsstandort Deutschland
 - Zielgruppe: Entscheidungsträger und Fachanwender

Unser Vorgehen in der Innovationspotenzialanalyse

Recherche (Bestandsaufnahme)



```
graph TD; A[Recherche (Bestandsaufnahme)] --> B[Branchenspezifische Zukunftsworkshops (Qualitative Untersuchung)]; B --> C[Online-Umfrage (Quantitative Untersuchung)];
```

Branchenspezifische
Zukunftsworkshops
(Qualitative Untersuchung)

Online-Umfrage
(Quantitative Untersuchung)

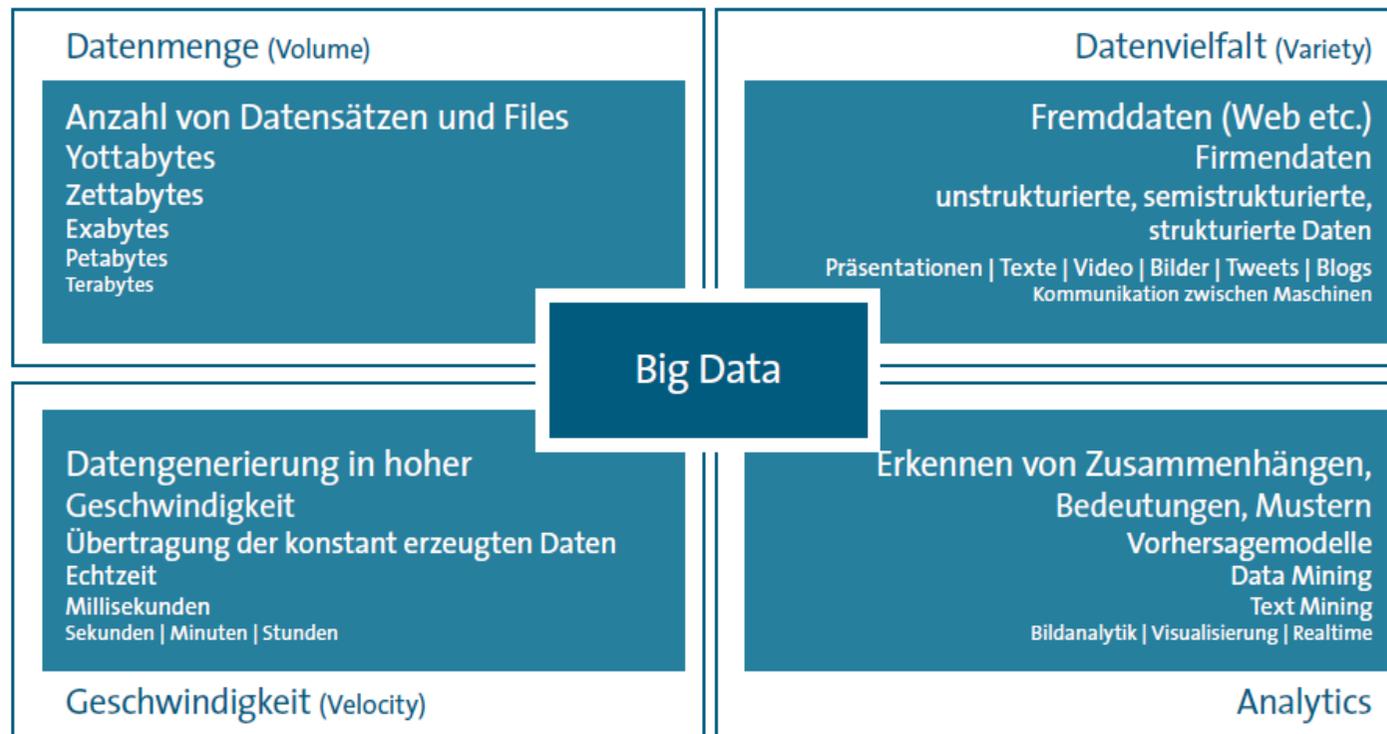
Branchenspezifische Zukunftsworkshops (Qualitative Untersuchung)

- 12.11.2012: Finanz (11 Experten)
 - 15.11.2012: Telekommunikation (12 Experten)
 - 19.11.2012: Marktforschung (4 Experten)
 - 26.11.2012: E-Commerce (7 Experten)
 - 30.11.2012: Versicherung (7 Experten)
- Die Einladung der Experten wurde unterstützt von



Big Data

Das Themenfeld laut BITKOM



Quelle: BITKOM Big Data Leitfaden, 2012. BITKOM AK Big Data.

**Perspektivisch haben wir mit Experten aus Branchenworkshops
mehr als 100 relevante Datenquellen identifizieren können**



Übersicht öffentlicher Datenquellen (Open Data, Auswahl)

**Common
Crawl**

6 Milliarden
Webseiten

1000 Genomes
A Deep Catalog of Human Genetic Variation

200 TB
Genomdaten

Google books

86 Milliarden
ngrams

DBpedia

400 Millionen
Fakten



WIKIPEDIA
Die freie Enzyklopädie

Über 4,1 Millionen
Artikel (in Englisch)

datacatalogs.org

270 Datenkataloge

Open-Source Technologien im Kontext von Big Data (Auswahl)



Kommerzielle Anbieter im Kontext von Big Data (Auswahl)



International wird in Big Data-Forschung investiert

Big Data is here to stay

- Big Data Research Development Initiative (USA, 200 Mio \$)
- NSF Big Data Förderprogramm (USA, 25 Mio \$)
- EU aktuelle und geplante Programme (EU, 51 Mio €)
- Roadmap Horizon 2020 des Big Data Public Private Forum (EU)
- Forschungssponsoring der Wirtschaft (MIT, Intel, Massachusetts)

Fachliches Innovationspotential von Big Data

Mehr speichern:

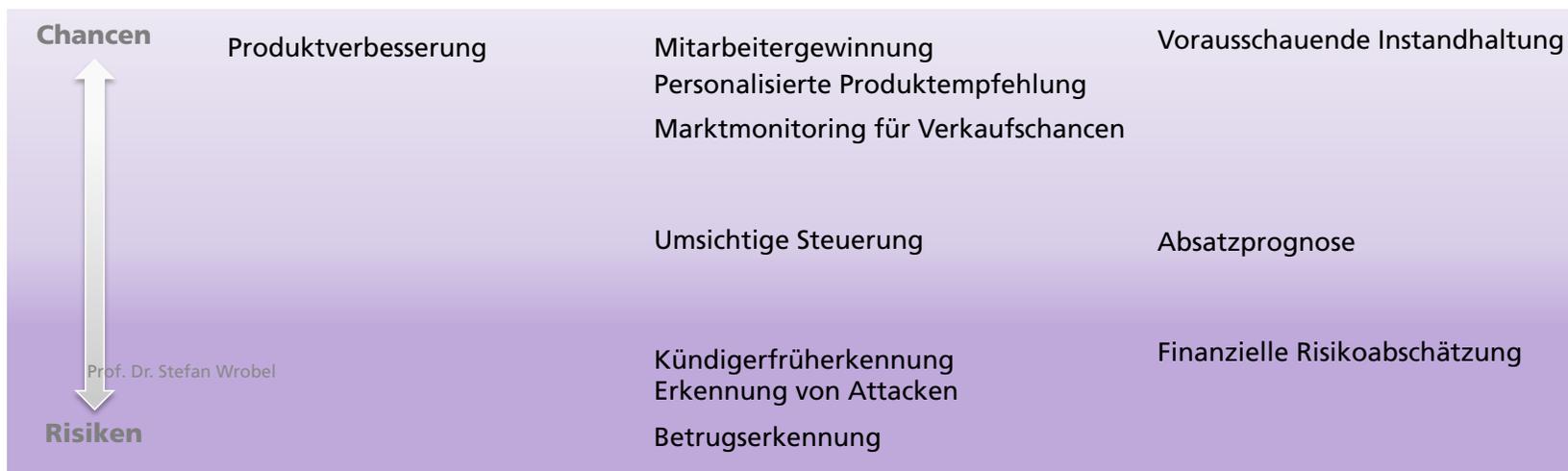
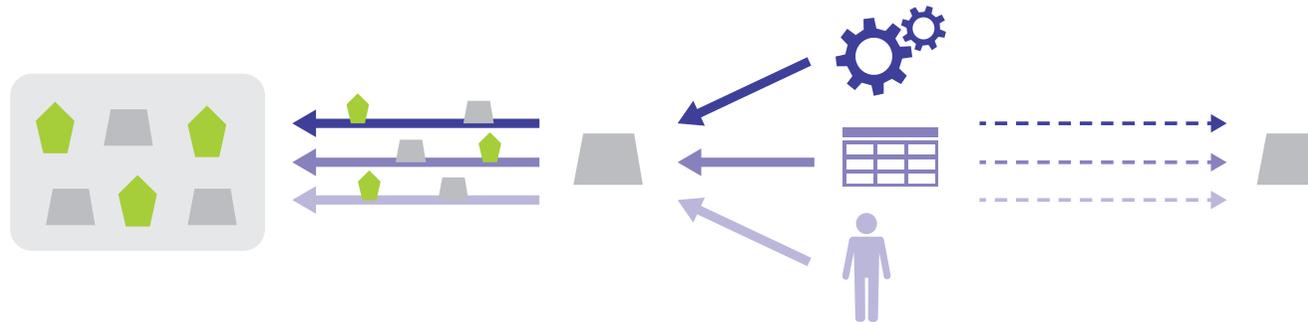
- Rohdaten müssen nicht mehr gelöscht werden
- Längere Historien

Complex Event Recognition

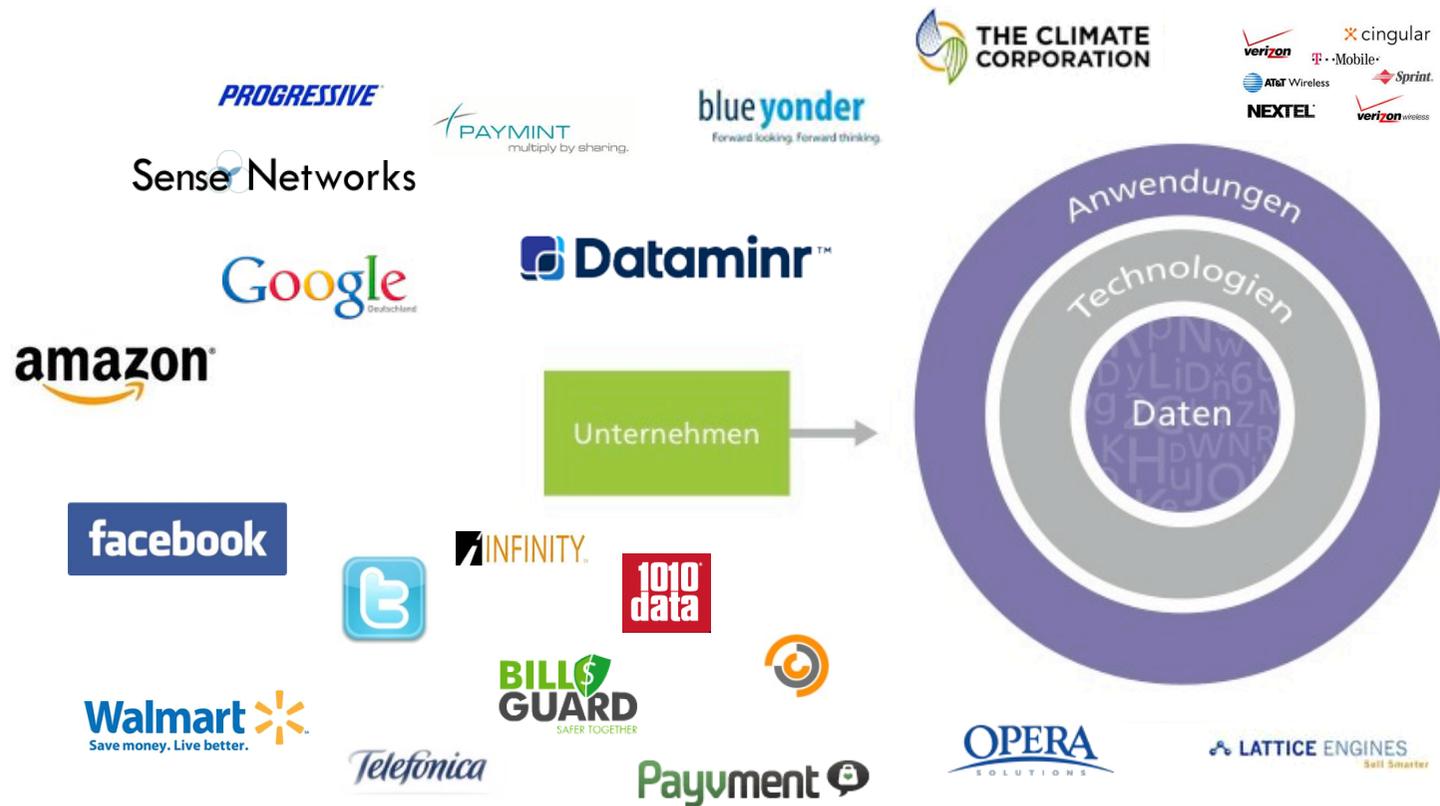
- Mehr beobachten
- Schneller erkennen
- Informationsvorsprung nutzen

Prädiktive Analytik

- Prognosen
- Simulationen



Beispiele für (Big)-Data-driven Enterprises



Einst Bücherversender, heute e-Retailer, morgen Big Data Infrastruktur-Anbieter? Amazon Inc.

Big Idea

- *“To continue to offer quality products and services using the best technology available and at a reasonable price.”*

Big Data

- 445,000 verfügbare Server über AWS (Schätzung von Accenture)
- Transaktionen mit über 68 Mio. Kunden
- Redshift Cluster mit bis zu 1 Petabyte mietbarem Speicher
- InMemory-Analytics HANA One SAP für \$0.99/h

Big Business

- Umsatz Amazon 48 Mrd. USD (2011)
- Umsatz bei AWS 1 Mrd. USD (2011), 2 Mrd. USD für 2012 (erwartet)
- Nettoerlöse 631 Mio USD (2011)
- Marktkapitalisierung 113 Mrd. USD



© James Duncan Davidson / Quelle: Flickr

Jeff Bezos: “Wir wollen Ungewöhnliches erfinden”

Facebook: Von einem Harvard-Kennenlern-Dienst zu über 1 Mrd. aktiven Nutzern weltweit

Big Idea

- *"We hope to rewire the way people spread and consume information ... We think a more open and connected world will help create a stronger economy with more authentic businesses that build better products and services."*

Big Data

- Über 1 Mrd. aktive Nutzer
- 2,5 Milliarden Inhalte, 2,7 Milliarden Likes und 300 Millionen Fotos pro Tag
- Datenvolumen von mehr als 500 Terabyte täglich

Big Business

- 3,71 Mrd. USD Umsatz (2011)
- 1 Mrd. USD Nettoerlöse (2011)
- Marktkapitalisierung von 68 Mrd. USD



© Guillaume Paumier / Wikimedia Commons

Mark Zuckerberg: "We don't build services to make money; we make money to build better services."

Als Big Data Start-up gestartet, heute ohnegleichen: Google Inc.

Big Idea

- *“Google’s mission is to organize the world’s information and make it universally accessible and useful.”*

Big Data

- 1 Million Server
- 1 Mrd. Suchanfragen pro Tag
- 20 Petabyte tägliches Datenaufkommen
- Analysiert das Suchinteresse und zeigt entsprechende Werbung an (AdWords)

Big Business

- 99% der Umsätze stammen von AdWords
- 37 Mrd USD Umsatz (2011)
- 9.7 Mrd. USD Nettoerlöse (2011)
- Marktkapitalisierung von 228 Mrd. USD (2012)



© Joi Ito / Quelle: Flickr

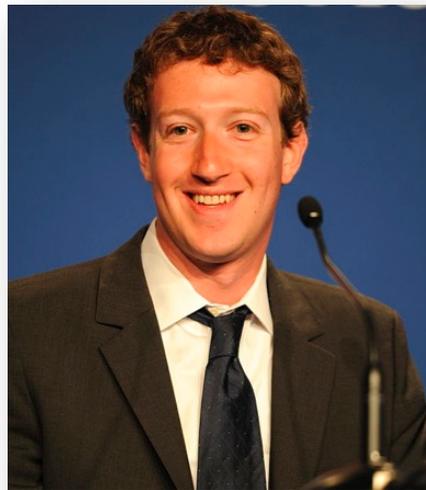
Sergey Brin and Larry Page:
“do no evil”

Was macht diese Unternehmen besonders?

Drei Beobachtungen mit Blick auf Big-Data-Nutzung



Fokussierung auf
Effizienz
durch Big Data



Fokussierung auf
Individualisierung
durch Big Data

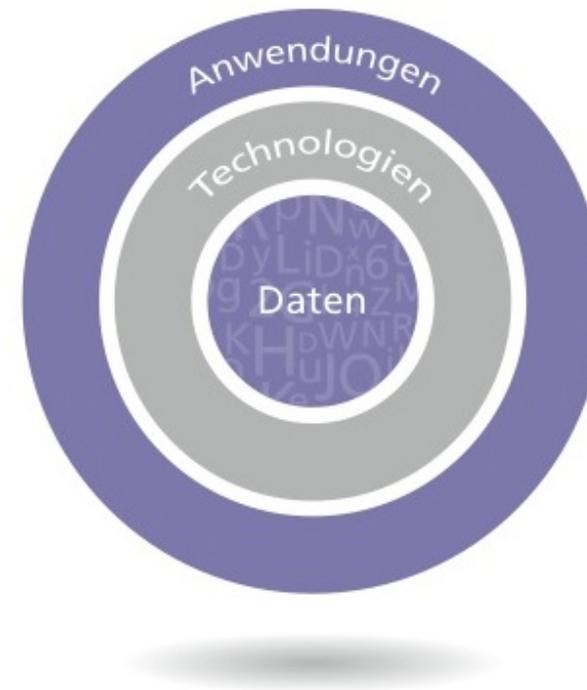


Fokussierung auf
Intelligenz
durch Big Data

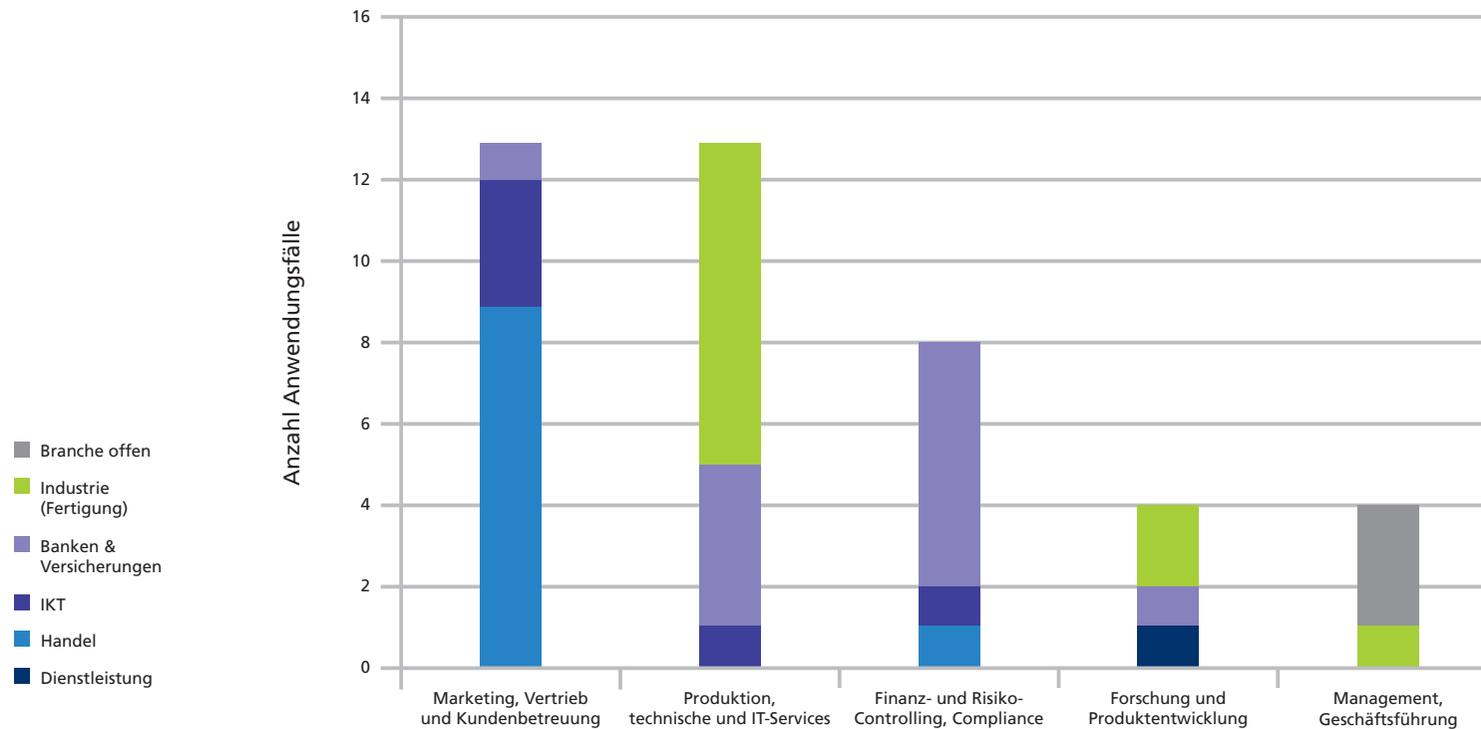
Wir haben international Big Data Use Cases recherchiert und 50 Fälle systematisch aufbereitet

Die wichtigsten Auswahlkriterien waren:

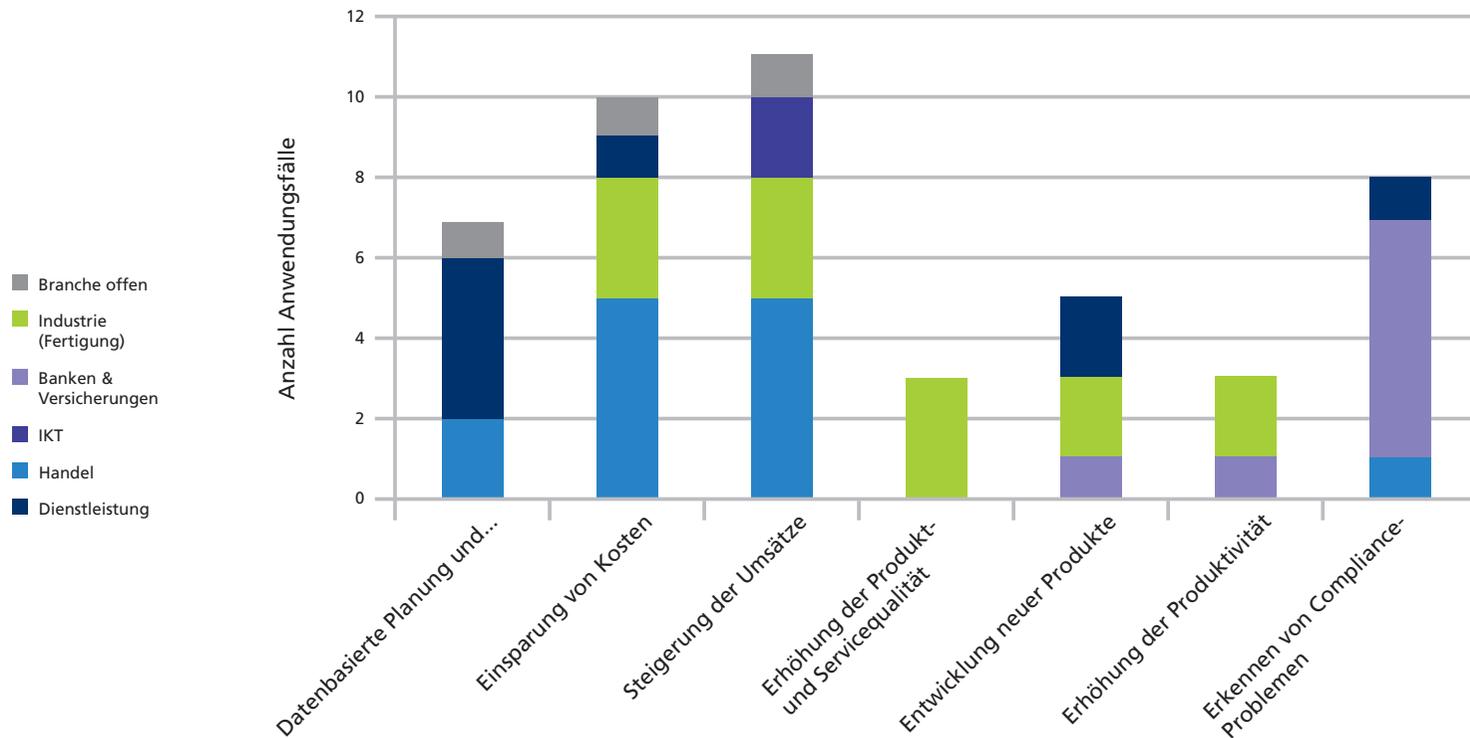
- Verarbeitung von umfangreichen, veränderlichen Datenmengen
- Analytische Fragestellungen
- Abdeckung möglichst verschiedener Branchen und Unternehmensbereiche



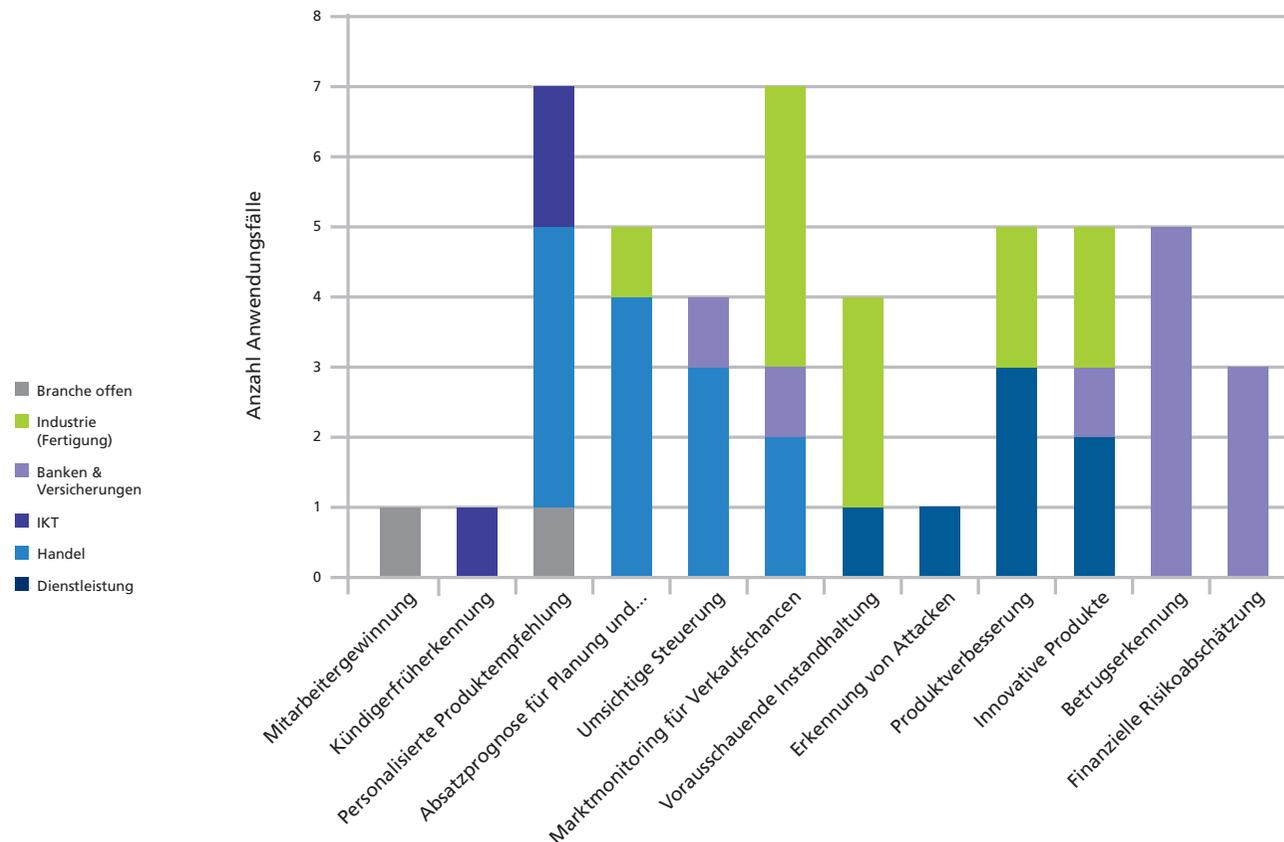
Jede Branche hat demnach charakteristische Unternehmensbereiche für Big Data Anwendungen



Steigerung der Umsätze und Einsparung von Kosten sind die häufigsten Ziele



Für jede Branche lassen sich Schwerpunkte bei den Aufgaben der Big-Data-Anwendungen identifizieren



Ausgewählte Ergebnisse der Online-Umfrage

- 69% aller Befragten in unserer Online-Umfrage wollen strategische Wettbewerbsvorteile durch Big Data erzielen.
- 78% teilen mit, dass sie Ihre personellen Ressourcen im Bereich Big Data verbessern müssen.
- 67% der Befragten teilen mit, dass das Budget für Big Data Themen (Technologien, Analysen, Datenquellen exkl. Personal) steigen sollte.
- Nur 8% der Befragten geben an, dass keine Umsetzungsbarrieren existieren.

Ausgewählte Ergebnisse der Online-Umfrage

- Die Hauptprobleme in der Umsetzung liegen in den Bereichen
 - Datenschutz und -sicherheit (49%)
 - Budget/Prioritätensetzung (45%)
 - technische Herausforderungen des Datenmanagements (38%)
 - Expertise (36%)
 - mangelnde Bekanntheit von Big Data Anwendungen und Technologien (35%).
- Um die bestehenden Defizite zu ändern, wünschen sich 95% der Befragten Förderung in Form von
 - Best Practices, Trainings, gefolgt von Anbieter- und Lösungsübersichten sowie überarbeiteten Datenschutzvorgaben

Stärken, Schwächen, Chancen und Risiken

Stärken

- Starker Mittelstand
- Hochgradige Spezialisierung
- hoher Exportanteil
- werthaltige Produkte (Premium)
- etablierte FuE-Netzwerke (PPP)
- „Deutsches Jobwunder in 2009“

Schwächen

- schwach ausgeprägtes Verständnis der Potentiale
- Geringe Datenkompetenz
- Unsicherheiten in Bezug auf Datenschutz und Sicherheit
- Technologieangebote noch sub-optimal (Komplexität, Sicherheit, Datenschutz)

Chancen

- Viele unrealisierte Effizienzpotenziale
- Individualisierung von Diensten
- Produktion von intelligenten Produkten für den Premium-Sektor
- „Neues deutsches Jobwunder?“

Risiken

- de facto Setzung von Privacy und Security Standards durch globale Anbieter
- datengetriebene Wettbewerber treten in bisherige Märkte ein
- zögerliche Adaption vergrößert den Lernabstand zu den globalen Wettbewerbern

Drei zentrale Chancen für Unternehmen in Deutschland

Big Ideas + Big Data = Big Business

- Chance 1: Effizienteres Unternehmensmanagement
- Chance 2: Massenindividualisierung von Diensten
- Chance 3: Intelligente Produkte



Big Ideas + Big Data = Big Business

– made in Germany

Chancen

- Chance 1: Effizienteres Unternehmensmanagement
- Chance 2: Massenindividualisierung von Diensten
- Chance 3: Intelligente Produkte

Herausforderungen

- Forschungsdynamik schaffen
- Schnelligkeit im Wissenstransfer
- Privacy & Security als Wettbewerbsvorteil

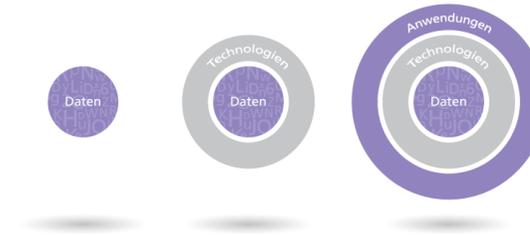


Living Lab Big Data

Konzeption einer Experimentierplattform

Dr. Michael May, Fraunhofer IAIS

Ausgangslage



- Viele Anwenderunternehmen sind in der Orientierungsphase
- Vorstellungen von Big Data in Unternehmen oft diffus, sowohl zu
 - Einsatzmöglichkeiten als auch zu
 - vorhandenen Techniken
- Ergebnisse Befragung und Zukunftsworkshops:
 - Hoher Bedarf an Best Practices, Training, Schulung (95% der Befragten)
- Infrastruktur und Datenmanagement sind für sich genommen zu abstrakt, Lösungspotentiale zu bewerten

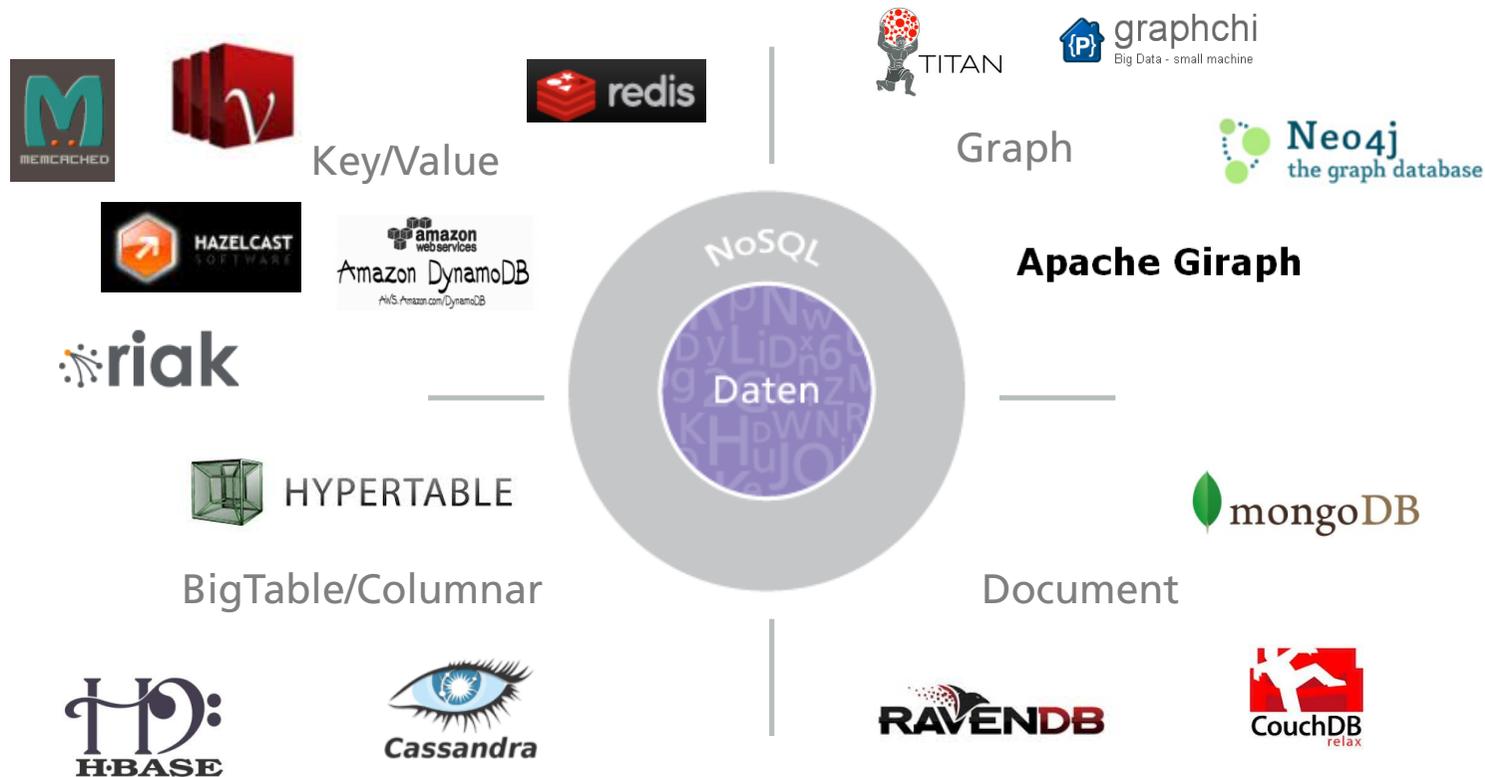
Im Kontext von Big Data sind bereits viele Open-Source Projekte verfügbar (eine Auswahl)



Big Data Anwendungen setzen auf eine Kombination von Tools aus verschiedenen Technologiekategorien



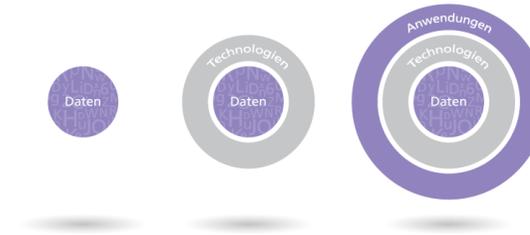
NoSQL Datenbanken – Verteilte nichtrelationale und schemafreie Datenbanken



Im Kontext von Big Data sind bereits viele Anbieter am Markt (eine Auswahl)

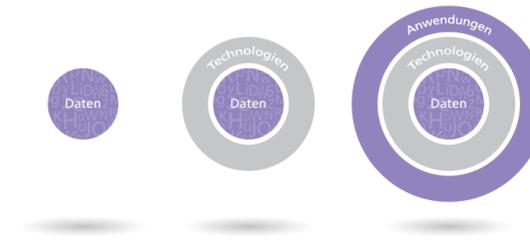


Barrieren



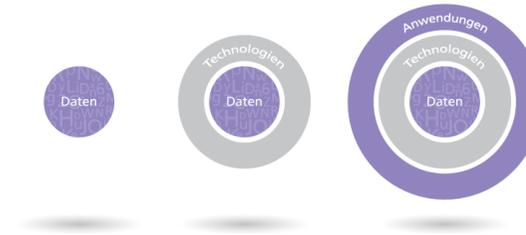
- Die Auswahl der richtigen Technologie ist eine Herausforderung
 - keine Standardisierung
 - unterschiedlicher Entwicklungsstand
 - unzureichende Dokumentation
 - uneinheitliche Terminologie, ...
 - Kaum Vergleichsmöglichkeiten, Benchmarks
- Kaum Personal mit Erfahrung vorhanden
- Anbieter reagieren mit „Out of the Box“-Lösungen (z.B. Appliance)
 - Hohe Kosten als Einstiegshürde

Ziele des Living Labs



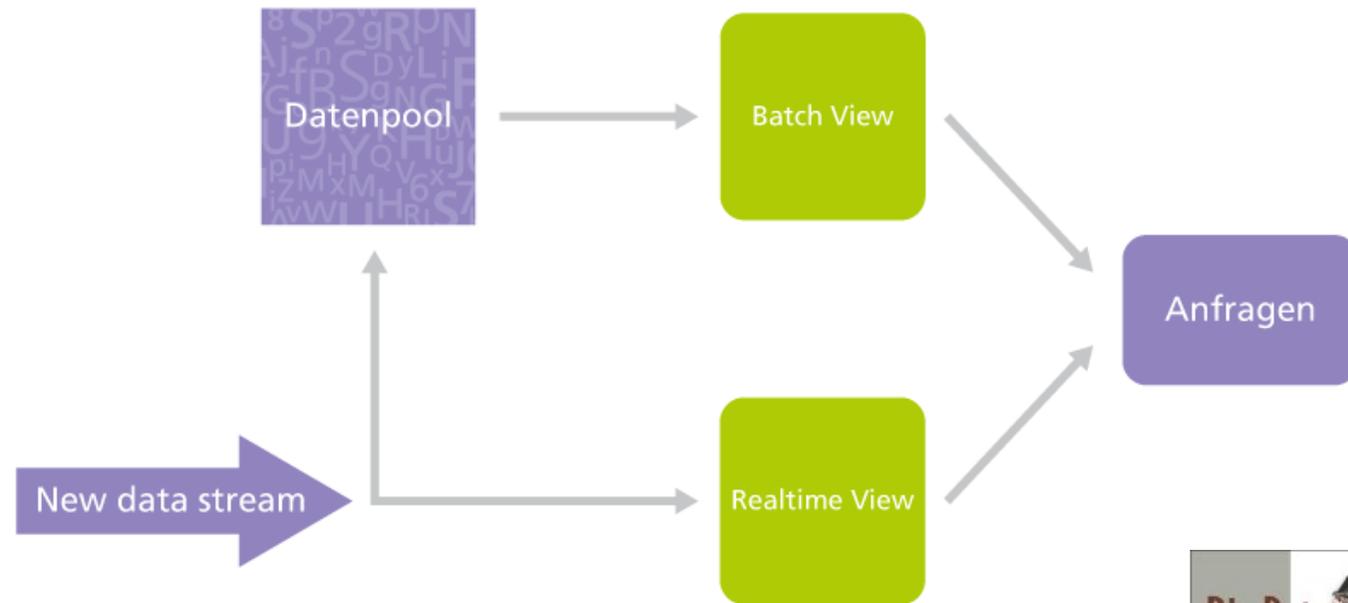
- Lern- und Experimentierumgebung für Unternehmen
- Technik „anfassbar“ machen, Einstiegshürden herabsetzen
- Teil eines Schulungskonzeptes „Data Science“
- Präsentation von Big-Data-Problemlösungen am Beispiel einer durchgängigen Anwendung des gesamten Technologie-Stacks
- Verschiedene öffentlich verfügbare Big-Data-Datenbestände werden eingespielt und stehen zum Testen zur Verfügung.
- Analyseverfahren für Big Data, die im Rahmen des THESEUS-Programms zum Thema Smart Semantics (Quote und Eat&Drink) entwickelt wurden
- Hardwareinfrastruktur steht nicht im Fokus
- Vorwettbewerblich, herstellerneutral, erweiterbar

Design Living Lab

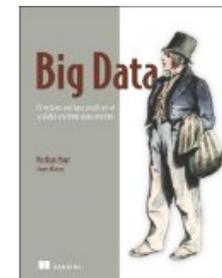


- Nicht eine einzelne Technologie (z.B. Hadoop, NoSQL), sondern eine **Big Data Architektur**, die flexibel genug ist,
 - verschiedenste Einsatzzwecke in verschiedensten Branchen abzudecken
 - es in den einzelnen Komponenten erlaubt, Technologien gegeneinander auszutauschen und z.B. auch eine individuelle Entscheidung zwischen OpenSource und kommerziellen Angeboten zu treffen
- Erst wenn man in Architekturen denkt, erschließen sich Stärken und Schwächen existierender Technologien

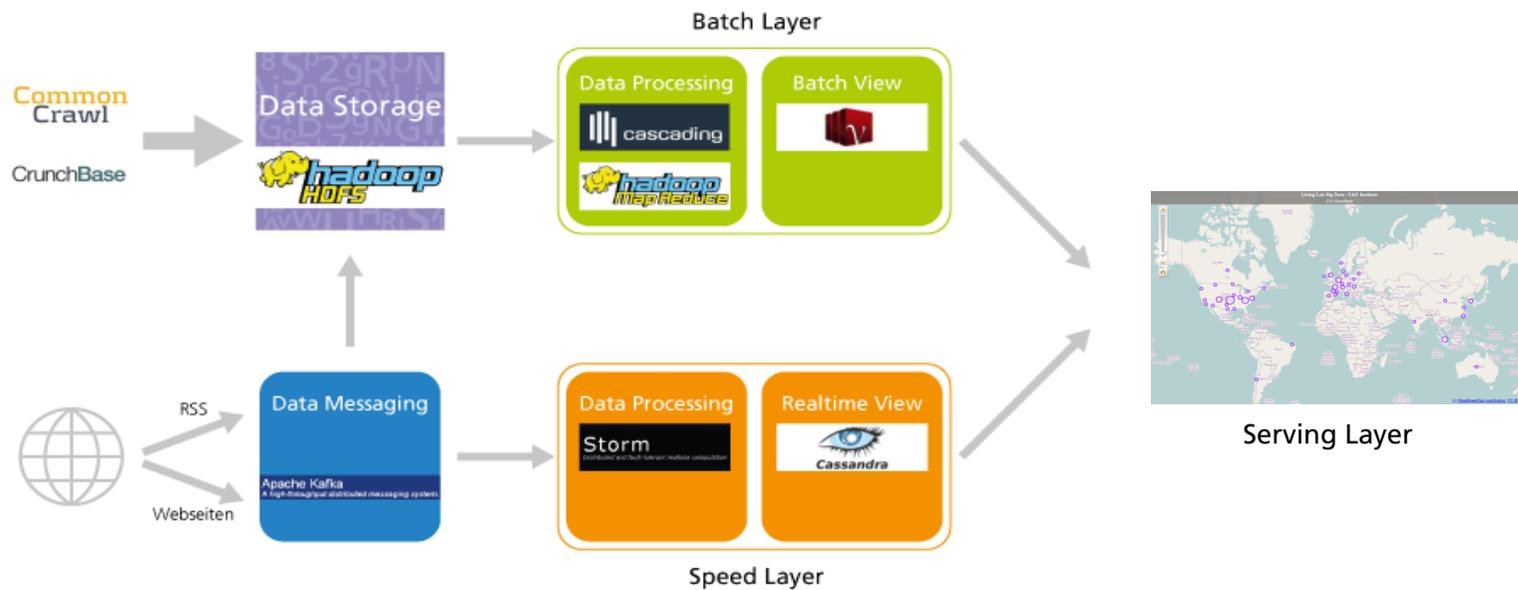
«Lambda Architecture»



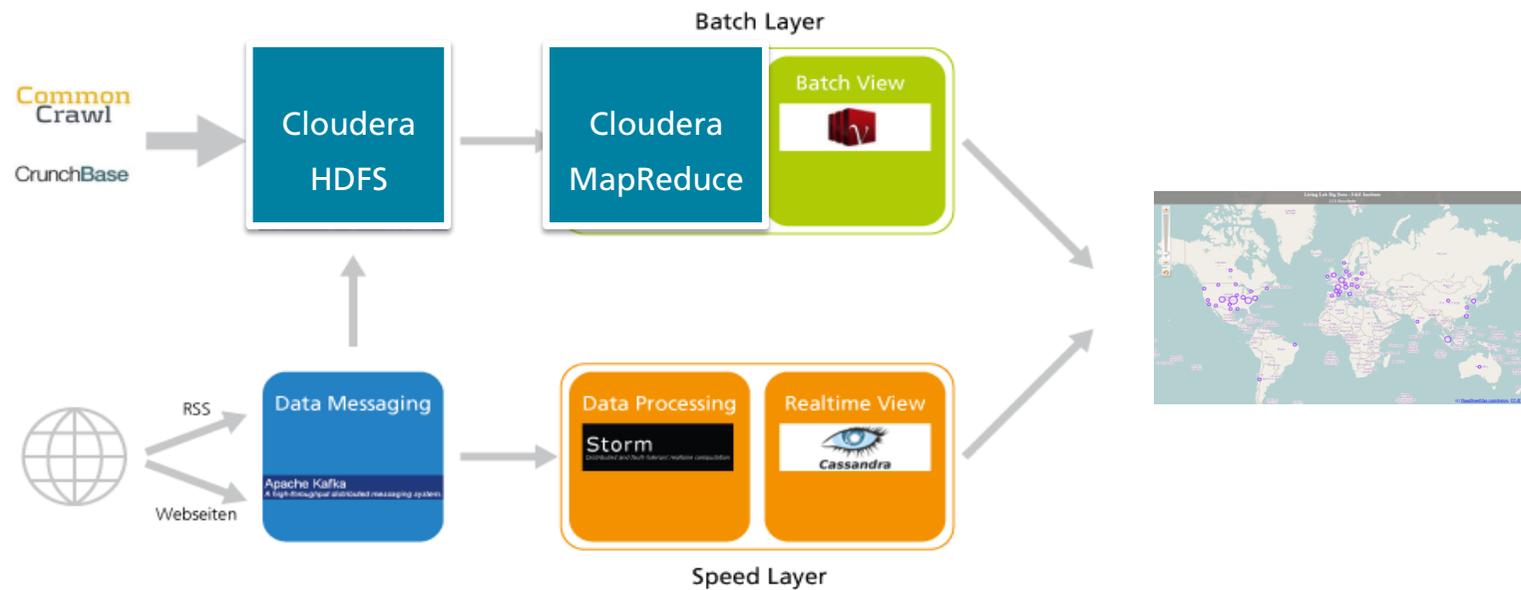
Quelle: Nathan Marz. Big Data: Principles and Best Practices of Scalable Realtime Data Systems. 2013



Living Lab - Architektur



Living Lab Technologiealternativen

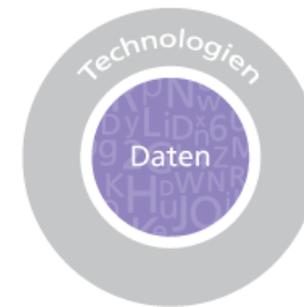


Wie kann das Living Lab genutzt werden?

- Zielgruppe: Data Scientists, Analysten, Entwickler
- Teil des Schulungskonzeptes Data Scientist – „Big Data Management“
- Schulungsmodul
- Individuelle Workshops
- „Bring your own Data“
- In-House-Installation möglich

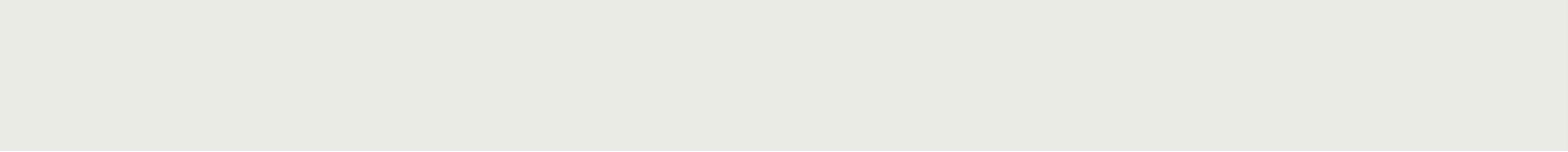
Zusammenfassung

- Das Living Lab macht Big Data „anfassbar“
- Ziel: Orientierung, Best Practices, Schulung
- Zentrales Architekturkonzept: Lambda Architektur
 - Basiert auf Open-Source-Komponenten
 - Durch kommerzielle Komponenten teilweise substituierbar
 - Erlaubt Exploration von Alternativen
- Herstellerunabhängig, vorwettbewerblich





Schlussfolgerungen



Big Data ist nicht neu

- Was „big“ (wohl im Sinner der Vier V's) ist, ist relativ. Es geht um das, was „bigger than usual“ ist.
 - Big Data ist gerade und erst jetzt in aller Mund, weil uns die technischen und analytischen Werkzeuge zur Verfügung stehen, damit etwas Vernünftiges zu machen.
 - Big Data ist umso interessanter, je „billiger“ sie ist. Einfach so „herumliegende“ Big Data sind besonders reizvoll.
 - „Rohdaten“ im eigenen Unternehmen
 - Anonymisierte Mobilfunkdaten
 - Social Media
-

Die Datenanalyse: Die Herausforderung von Big Data

- Ein reizvolles Zusammenspiel von Mathematik und Informatik
 - Straight forward-Algorithmen basieren auf linearer Algebra
 - Die Parallelisierung solcher Algorithmen ist weniger straight forward.
 - Vermehrt kommt maschinelles Lernen zum Zug
 - Unüberwacht: Cluster-Analyse, Datenexploration
 - Überwacht: z.B. in Betrugserkennungsapplikationen
-

Big Data relativiert Erfahrung und Expertise

- Die leichte Zugang zu grossen Mengen billiger Daten verändert die Art, wie wir uns Problemen nähern.
 - Haben „früher“ Experten
 - das Problem studiert,
 - eine Theorie entwickelt,
 - ein mathematisches Modell gebaut,
 - Daten gesammelt, um das Modell zu validieren;
 - Suchen Mathematiker „Auffälligkeiten“ in den Daten, die Experten erklären sollen.
-