



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

MNM  
TEAM  
MÜNICH NETWORK MANAGEMENT TEAM



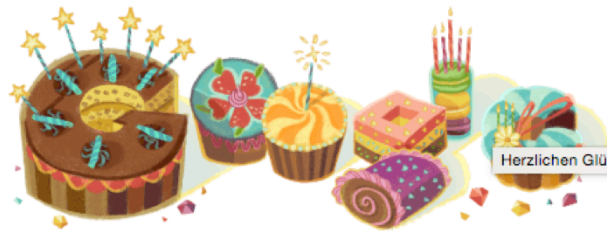
## Gegenwart und Zukunft von Big Data

Dieter Kranzlmüller

Munich Network Management Team  
Ludwig-Maximilians-Universität München (LMU) &  
Leibniz-Rechenzentrum (LRZ)  
der Bayerischen Akademie der Wissenschaften



- \* Ein Hype
- \* Eine spannende (Informatik-)Herausforderung
- \* Eine Chance
- \* Aber auch: Eine permanente Verletzung der Werte unseres Rechtsraumes, hervorgerufen durch eine paranoide Sammelwut



Herzlichen Glückwunsch zum Geburtstag, Dieter!

In Google suchen oder URL eingeben



Ungefähr 820.000.000 Ergebnisse (0,31 Sekunden)

**Big Data Analytics - See How Big Data Works in Real-Time**

**Anzeige** [www.pentaho.com/](http://www.pentaho.com/) ▾

Download the Free eBook.

171 Personen folgen Pentaho auf Google+

Big Data Visualization  
2015 Gartner BI Report

Hadoop Business Analytics  
Predictive Analytics

**BARC-Studie Big Data - SAS.com**

**Anzeige** [www.sas.com/Big-Data](http://www.sas.com/Big-Data) ▾

Big Data Analytics in D-A-CH: Kostenfreie BARC Research Study!

**Big Data – Wikipedia**

[de.wikipedia.org/wiki/Big\\_Data](http://de.wikipedia.org/wiki/Big_Data) ▾

Big Data [ˈbɪɡ ˈdeɪtə] (von englisch big „groß“ und data „Daten“) bezeichnet Datenmengen, die zu groß oder zu komplex sind oder sich zu schnell ändern, ...  
Hintergrund - Beispiele - Verarbeitung von Big Data - Kritik

**Alles über das Thema Big Data - computerwoche.de**

[www.computerwoche.de/k/big-data,3457](http://www.computerwoche.de/k/big-data,3457) ▾

Hintergründe und aktuelle Ratgeber rund um das Thema Big Data – mit welchen Strategien Lösungsansätzen können Unternehmen die wachsende Datenflut ...  
Hintergrund - Ratgeber - News - Video

**News-Themen**



**Big Data: Der Kampf um die Daten im Auto**

PC-Welt - vor 1 Tag

Ein modernes Auto ist auch ein gewaltiger Datenspeicher, denn es ist mit bis zu 70 ...

**Anzeigen**

**Big Data Analytik**

[www.intel.de/Big\\_Data](http://www.intel.de/Big_Data) ▾

Intel® liefert Technologie für jede Big Data-Herausforderung

**Google BigQuery**

[cloud.google.com/BigQuery](http://cloud.google.com/BigQuery) ▾

Analyse Big Data In The Cloud Using SQL. No Servers To Install. Try Now

**IBM Big Data & Analytics**

[www.ibm.com/de/](http://www.ibm.com/de/) ▾

Ohne Analytics ist Big Data nur ein lautes Hintergrundrauschen!

**Predictive Analytics**

[www.open-ls.de/](http://www.open-ls.de/) ▾

1 Million verborgene Schätze.  
1.000 Möglichkeiten. 100% Überblick

**Oracle Big Data**

[www.oracle.com/BigData](http://www.oracle.com/BigData) ▾

Explore An Industry Recognized High Performance Solution For Big Data.

**Talend - Big Data Säulen**

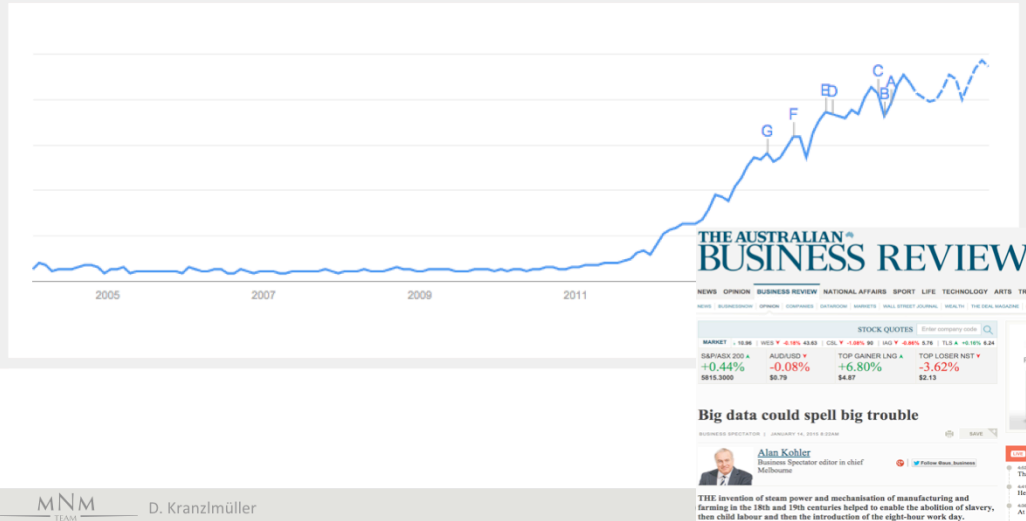
[www.talend.com/4PillarsOfBigData](http://www.talend.com/4PillarsOfBigData) ▾

Ihre Big Data Management Lösung. Whitepaper heute gratis erhalten!

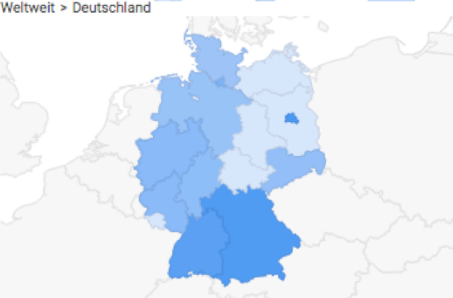
http://www.google.de/trends/explore#q=Big%20Data

Interesse im zeitlichen Verlauf

Nachrichtenschlagzeilen  Prognose



## Regionales Interesse ?



	Region   Stadt	
Indien	100	<div style="width: 100%;"></div>
Singapur	64	<div style="width: 64%;"></div>
Republik Korea	56	<div style="width: 56%;"></div>
Hongkong	51	<div style="width: 51%;"></div>
Taiwan	49	<div style="width: 49%;"></div>
Berlin	100	<div style="width: 100%;"></div>
Bayern	96	<div style="width: 96%;"></div>
Baden-Württemberg	91	<div style="width: 91%;"></div>
Hamburg	69	<div style="width: 69%;"></div>
Rheinland-Pfalz	60	<div style="width: 60%;"></div>
Nordrhein-Westfalen	60	<div style="width: 60%;"></div>
Hessen	54	<div style="width: 54%;"></div>

## Verwandte Suchanfragen ?

Themen	Beliebteste Zunehmende		Suchanfragen	Beliebteste Zunehmende	
	Beliebteste	Zunehmende		Beliebteste	Zunehmende
Data - Website Category	100		big data analytics	100	
Big data - Industry	95		data analytics	100	
Analytics - Software Genre	10		hadoop big data	80	
Apache Hadoop - Software	5		hadoop	80	
Data analysis - Industry	5		google big data	40	
Big Data - Musical Group	0		big data dangerous	30	
			big data pdf	30	

[Google.org - Startseite](#) (auf Englisch)

[Denguefieber-Trends](#)

**Grippe-Trends**

Startseite

Land/Region auswählen

[Wie funktioniert's?](#)

[Häufig gestellte Fragen](#)

**Grippe-Häufigkeit**

sehr hoch

hoch

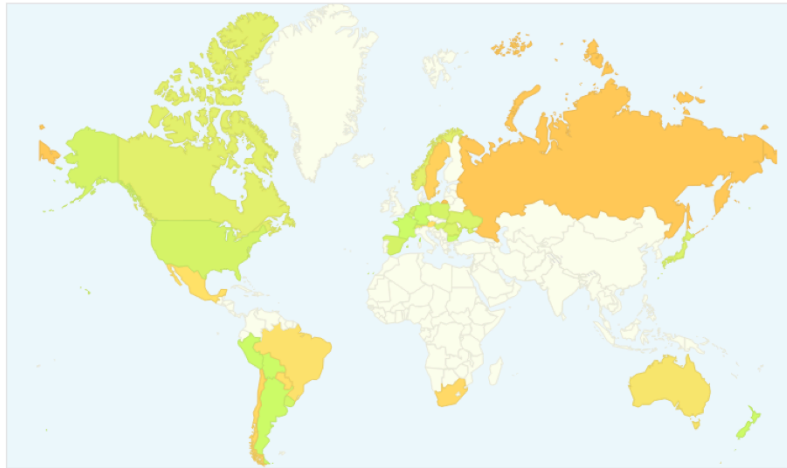
mittel

niedrig

minimal

## Grippe-Trends weltweit verfolgen

Google hat festgestellt, dass die Häufigkeit bestimmter Suchbegriffe Anhaltspunkt für die Häufigkeit von Grippefällen sein kann. Für die Google Grippe-Trends werden Daten der Google-Suche gesammelt und ausgewertet. Auf Grundlage der Ergebnisse wird anschließend die Häufigkeit von Grippefällen geschätzt. [Weitere Informationen »](#)



[Daten zur weltweiten Grippe-Häufigkeit herunterladen](#)



## Einstellungen für Werbung

### Einstellungen für Google Anzeigen

Durch Werbung können Webdienste und -inhalte kostenlos bereitgestellt werden. Mit diesen Einstellungen können Sie festlegen, welche Arten von Google Anzeigen für Sie eingeblendet werden.

#### Werbung bei Google



#### Google Anzeigen im Web

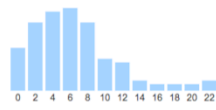


Geschlecht	Männlich <a href="#">Google-Profil aufrufen</a>	Männlich Basierend auf Ihrem Google-Profil
Alter	45-54 <a href="#">Google-Profil aufrufen</a>	45-54 Basierend auf Ihrem Google-Profil
Sprachen	Nicht verfügbar	Deutsch <a href="#">Bearbeiten</a> Basierend auf den von Ihnen gewählten Sprachen
Interessen	Abenteuerspiele und 26 weitere <a href="#">Bearbeiten</a> Basierend auf Ihren bisherigen Aktivitäten auf Google	Action- und Abenteuerfilme <a href="#">Bearbeiten</a> Basierend auf den von Ihnen angesehenen Videos

- #### Interessen
- Abenteuerspiele
  - Autos und Fahrzeuge
  - Bankwesen
  - Computer und Elektronik
  - Computer- und Videospiele
  - Fernsehen und Video
  - Film- und Fernsehproduktion
  - Fußball
  - Gesetz und Regierung
  - Haus und Garten
  - Hunde
  - Hygiene- und Toilettenartikel
  - Kunst und Unterhaltung
  - Lateinamerikanische Musik
  - Musik und Audio
  - Nachrichten



### Stündliche Suchaktivität



### Tägliche Suchaktivität



### Aktivität



Mehr Trends ansehen >>

Einträge entfernen

Weitere Web- & App-Aktivitäten ansehen

### Heute

- Gesucht nach what does google know about me  
Cloud Fender Tips & Tricks • 6 links... - cloudfender.com
- Gesucht nach what does facebook know about me  
What Does Facebook Know About You : The... - daylandoes.com
- Gesucht nach big data
- Gesucht nach Om Shanti  
Startseite bei Restaurant Om Shanti... - restaurantomshanti.de
- Gesucht nach Suchabfragen Google  
http://www.google.de/trends/?hl=de
- Gesucht nach Google Analytics  
https://www.google.com/intl/de/analytics
- Gesucht nach big data
- Gesucht nach Big Data
- Gesucht nach Von Oettingenstraße 67, München nach Wildbad Kreuth, Bildungszentrum Wildbad Kreuth, 83708
- Bildungszentrum Wildbad Kreuth - google.de



Angezeigt werden 18 Nachrichten für Big Data

### Big Data brings big opportunities for UM students

This spring, seven students will be the first from the University of Montana to enter the job market armed with UM's new Big Data analytics certificate.

Montana Kaimin | 21 hours ago

### Big Data Start-Up Rocana Raises \$15M From Google Ventures & ... Catalyst

SAN FRANCISCO-(BUSINESS WIRE)-Big data start-up Rocana lands \$15 million in Series B Funding from Google Ventures and General Catalyst; company aims to leverage Hadoop to keep data centers up and running.

Business Wire | 19 hours ago

### Japan Big Data (By Solution(Hadoop,Big Data Analytics,Big Data Network

LONDON, April 29, 2015 /PRNewswire/ -- Japan big data analytics market is estimated to grow at a CAGR of 34% from 2014-2019. Although China holds the larger shares in the APAC big data analytics market, ...

PR Newswire | 2 days ago

### HDS big-data should help health work

Hitachi Data Systems big-data expertise and enterprise IT specialized products types of information has been expanding traditional storage through develop

PC World | 2

Die Suchmaschine, die Sie nicht verfolgt. Mehr erfahren.

#### WERBUNG CenturyLink Big Data

Learn About CenturyLink Technology Big Data IT Solutions Now.  
CenturyLinkTechnology.com/BigData | Integrated Data Services | ROI Analysis Tool

#### Big data - Wikipedia, the free encyclopedia

**Big data** is a broad term for **data** sets so large or complex that traditional **data** processing applications are inadequate. Challenges include analysis, capture, **data** curation, search, sharing, storage, transfer, visualization, and information privacy.

W en.wikipedia.org/wiki/Big\_data

#### What Is Big Data? - Forbes

**Big data** is new and "ginormous" and scary -very, very scary. No, wait. **Big data** is just another name for the same old **data** marketers have always used, and it's not all that **big**, and it's something we should be embracing, not fearing. No, hold on. That's not it, either.

f forbes.com/sites/isaarthur/2013/08/15/what-is-big-d...

#### Big Data (band) - Wikipedia, the free encyclopedia

**Big Data** is an American electronic music project helmed by producer Alan Wilkis. They are best known for their single "Dangerous", featuring Joywave, which reached number one on the Billboard Alternative Songs chart in August 2014. **Big Data's** first EP, 1.0, was released on October 1 ...

W en.wikipedia.org/wiki/Big\_Data\_(band)

#### IBM Big Data - What is Big Data - United States

What is **Big Data**? **Big data** is being generated by everything around us at all times. Every digital process and social media exchange produces it.

ibm.com/big-data/us/en/



Quelle: [www.go-globe.com](http://www.go-globe.com) (2011)

LMU LUWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

## Awesome World Map of Facebook Connections



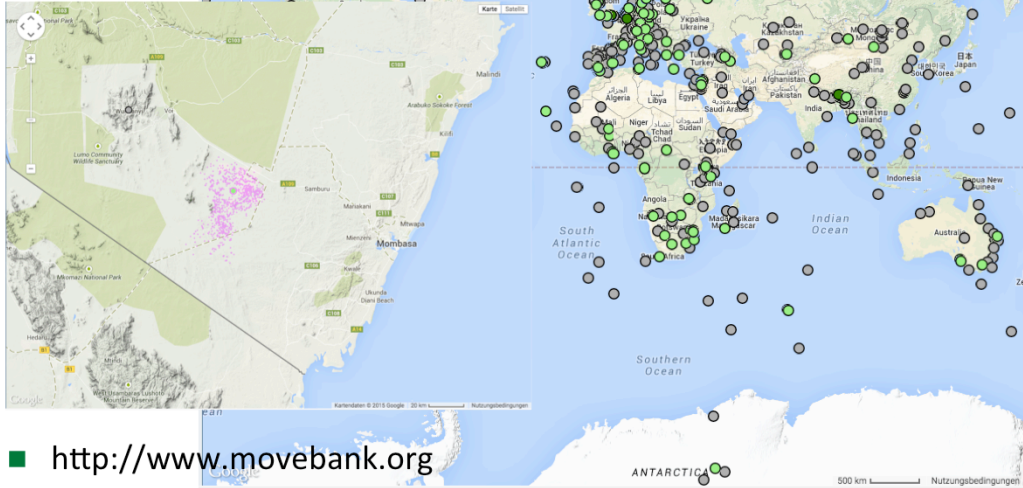
facebook

December 2010

Paul Butler a Facebook Intern, visualizes the relationships between 10 million pairs of Facebook friends  
Quelle: <http://maptd.com/awesome-world-map-of-facebook-connections>

MNM D. Kranzmüller Gegenwart und Zukunft von Big Data 12

Remarkably, the map doesn't include any of the usual features you find in a normal map eg ocean-land borders. Yet it's still possible to see these through the structures built from human relationships.



■ <http://www.movebank.org>

**Autorate überfällig:**

- So war Rita Torres nachts auf der Autobahn unterwegs, als ihr Wagen in voller Fahrt plötzlich stoppte. *"Es war ein Albtraum. Es gab keinerlei Vorwarnung, nichts. Ich hatte eine Riesenangst. Ich hätte einen Unfall verursachen können."* Es gelang ihr, den Wagen von der Straße zu schieben und abschleppen zu lassen.
- Später erfuhr sie, dass ihr Kreditgeber sie ausgebremst hatte, weil sie angeblich mit ihren Rückzahlungen im Verzug war.
- <http://www.heute.de/wegfahrsperrn-in-den-usa-wer-autorate-nicht-zahlt-wird-ausgebremst-37164702.htm>

Autobahn A44 bei Willich \* Quelle: photosforfree.de \*  
Fotograf/Zeichner: Rolf van Melis \* Datum: 2005 \* Sonstiges: gemeinfreie Lizenz

<http://blog.fefe.de/?ts=aa1cfcad>

- Versicherung
- Software-Update

- Süddeutsche Zeitung (SZ), 15. Juli 2013
  - „Zahlen für jede Vollbremsung“
  - Katrin Berkenkopf
  - <http://www.sueddeutsche.de/auto/kfz-versicherung-zahlen-fuer-jede-vollbremsung-1.1721828>
  
- Security Insider, 30. April 2015
  - „Das Auto ist ein Computer und ebensowenig vor Hackern sicher“
  - Malte Pollmann, Thomas Kuther
  - <http://www.security-insider.de/themenbereiche/bedrohungen/sicherheitsluecken/articles/487097/>
  
- Vergleich: Facebook / Auto



<http://de.engadget.com/2015/03/31/fotos-so-sieht-es-in-facebooks-neuem-hauptquartier-aus/>



[www.bmw.de](http://www.bmw.de)

Facebook (incl. Backend) : 60 Mio LOC

Premium Car: 120 Mio LOC

<http://www.informationisbeautiful.net/visualizations/million-lines-of-code/>

- Inside Science News Service, 6. März 2015:
  - „Sewage Bacteria Reveal Cities' Obesity Rates“
  - „Urban 'microbiome' can offer glimpses into disease trends“
  - Originally published: Mar 6 2015 - 11:45am
  - By: Brian Owens, Contributor

- Untersuchung der städtischen Abwasser erlaubt Rückschlüsse auf Bakterien im menschlichen Körper
- Anteil der Fettleibigkeit einer Bevölkerung kann mit 90% Genauigkeit vorhergesagt werden

- <http://www.insidescience.org/content/sewage-bacteria-reveal-cities-obesity-rates/2641>



„Manhole Brighton“ von Les Chatfield aus Brighton, England

„Manhole Brighton“ von Les Chatfield from Brighton, England - Manhole.  
Lizenziert unter CC BY 2.0 über Wikimedia Commons - [http://commons.wikimedia.org/wiki/File:Manhole\\_Brighton.jpg#/media/File:Manhole\\_Brighton.jpg](http://commons.wikimedia.org/wiki/File:Manhole_Brighton.jpg#/media/File:Manhole_Brighton.jpg)



- Zur Verbrechens-aufklärung und –bekämpfung
- Überwacht und vergleicht Daten aus verschiedenen Quellen, u.a.:
  - 3.000 Videokameras
  - Nummernschildscanner
  - Strahlungsdetektoren
  - div. Datenbanken



Quelle: Zeit (2012)  
<http://nymag.com/daily/intelligencer/2012/08/nypd-domain-awareness-system-microsoft-is-watching-you.html>



Für 40 Millionen Dollar wurde in New York das sogenannte Domain Awareness System installiert, eine gigantische Überwachungszentrale. 3.000 Videokameras in Manhattan liefern ihr Bilder. Zusammen mit Strahlungsdetektoren, Nummernschildscannern und Verknüpfungen zu diversen Datenbanken sollen sie helfen, Verbrechen aufzuklären.



aka **Intelligence Community  
Comprehensive National  
Cybersecurity Initiative  
Data Center**

Camp Williams, Bluffdale, Utah

[http://en.wikipedia.org/wiki/Utah\\_Data\\_Center](http://en.wikipedia.org/wiki/Utah_Data_Center)

1,5 million square feet  
2 billion US\$ building  
2 billion US\$ hardware  
65 MW power  
1 Yottabyte Storage



To process “all forms of communication, including the complete contents of private emails, cell phone calls, and Internet searches, as well as all sorts of personal data trails - parking receipts, travel itineraries, bookstore purchases, and other digital 'pocket litter'.”

8

1,5 Msqf ~ 139350 qm = ca. 14 x LRZ Fläche

Stromverbrauch: ca. 9 Mal LRZ

Speicher: ca.  $10^{24}$  Bytes: ca. 5 Millionen Mal LRZ

- “During a single day last year, the NSA’s Special Source Operations branch collected 444,743 e-mail **address books** from Yahoo, 105,068 from Hotmail, 82,857 from Facebook, 33,697 from Gmail and 22,881 from unspecified other providers, according to an internal NSA PowerPoint presentation. Those figures, described as a typical daily intake in the document, correspond to a rate of more than 250 million a year.”
  
- The Washington Post, 15.10.2013
  - [http://www.washingtonpost.com/world/national-security/nsa-collects-millions-of-e-mail-address-books-globally/2013/10/14/8e58b5be-34f9-11e3-80c6-7e6dd8d22d8f\\_story.html](http://www.washingtonpost.com/world/national-security/nsa-collects-millions-of-e-mail-address-books-globally/2013/10/14/8e58b5be-34f9-11e3-80c6-7e6dd8d22d8f_story.html)



„Hubble ultra deep field“ von NASA and the European Space Agency. - <http://hubblesite.org/newscenter/archive/releases/2004/07/image/a/warn/>.  
Lizenziert unter Gemeinfrei über Wikimedia Commons - [http://commons.wikimedia.org/wiki/File:Hubble\\_ultra\\_deep\\_field.jpg#/media/File:Hubble\\_ultra\\_deep\\_field.jpg](http://commons.wikimedia.org/wiki/File:Hubble_ultra_deep_field.jpg#/media/File:Hubble_ultra_deep_field.jpg)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN **Big Data – Definition – Mehr als nur „Big“** LRZ

3 V' s: "Big data is high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."

Beyer, M. & Laney, D. (2012): The Importance of 'Big Data': A Definition, Gartner Research Report

The diagram illustrates the three V's of Big Data:

- Volume:** Represented by a pink box. It lists data sizes: GigaByte, TeraByte, PetaByte, and ExaByte. The description is "Sehr große Mengen von Daten".
- Variety:** Represented by a dark red box. It lists data types: Text, Audio, Sensor Data, Video, and Social Network Data. The description is "Unterschiedlichste Datenquellen und -formate".
- Velocity:** Represented by a teal box. It lists data processing methods: Batch, Periodic, Real Time, and Near Real Time. The description is "Anfallen und Auswertung der Daten in (Near) Real-Time".

Quelle: Arnold Picot, Munich School of Management, LMU

MNM D. Kranzmüller Gegenwart und Zukunft von Big Data 21

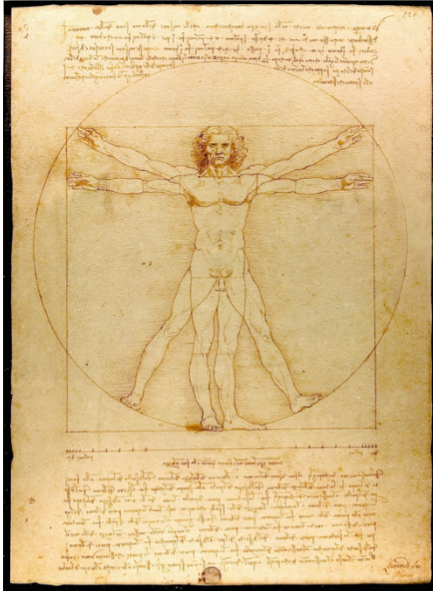
"Big Data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets." (Boyd/Crawford, 2012)

Früher bezog sich der Ausdruck Big Data auf **Datenmengen die nur mit einem Supercomputer analysiert werden konnten**. Heute mit Desktop und Standard-Software möglich.

**Volume:** Laut IDC ist die weltweite Datenmenge zwischen 2005 und 2010 um den Faktor fünf gewachsen. Bereits **2010 durchbrach die weltweit in einem Jahr erzeugte Gesamtmenge an Informationen die Zettabyte Schallmauer** (Zahl mit 21 Nullen). Auf Facebook werden täglich im 20-Minuten-Rhythmus 2,7 Millionen Bilder eingestellt. Im Jahr 2010 wurden insgesamt 700 Milliarden Videos auf YouTube abgespielt.

**Variety:** Als Regel lässt sich sagen, dass heute lediglich **15% der Daten strukturiert sind, ca. 85% der Daten sind unstrukturiert**. Zentrale Herausforderung ist die rasant wachsende Komplexität, die aufgrund der mangelnden Struktur entsteht.

**Velocity:** Die Daten verlieren u.U. schnell ihren Wert



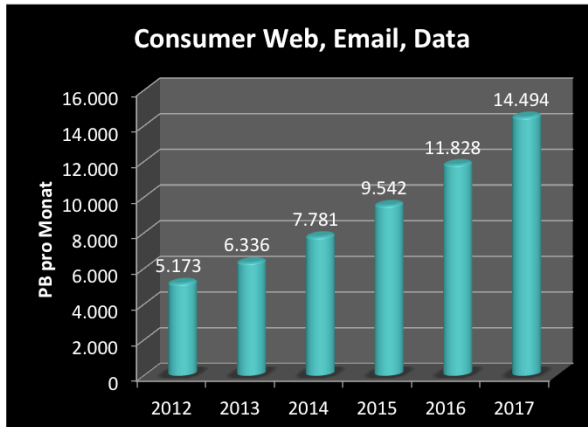
Besteht aus ungefähr

**$3,72 \times 10^{13}$**   
(37,2 Billionen)

Zellen

Quelle: Bianconi, Eva et al: *An estimation of the number of cells in the human body*. In: *Annals of Human Biology* 40 (2013), Nr. 6, S. 463–471

Quelle: [http://upload.wikimedia.org/wikipedia/commons/2/22/Da\\_Vinci\\_Vitruve\\_Luc\\_Viatour.jpg](http://upload.wikimedia.org/wikipedia/commons/2/22/Da_Vinci_Vitruve_Luc_Viatour.jpg)



Besteht 2013 aus  
ungefähr

**633.600 x 10<sup>13</sup>**  
(6,3 Trillionen)

Bytes/Monat(!)

<sup>\*)</sup> Nur Consumer Web, Email, Instant Messaging und anderer Datenverkehr (ohne File Sharing)

Quelle: Cisco Visual Networking Index: Forecast and Methodology, 2012-2017, Cisco 2013

Entspricht der Anzahl aller Körperzellen aller Einwohner von Saarbrücken (Einwohner 2012 gemäß Statistischem Bundesamt, 2013)



Quelle: <http://www.google.com/about/datacenters/>

Entspricht in etwa der Anzahl aller Körperzellen aller Einwohner Deutschlands (Einwohner 2012 gemäß Angaben der Weltbank von 2013)

ungefähr  
**280.000.000 x**  
**10<sup>13</sup>**  
(2,8 Zettabyte oder 2,8 Trilliarden)  
**Bytes**

Quelle: Spiegel Online vom 11.12.2012 (<http://www.spiegel.de/netzwelt/web/das-internet-der-dinge-erzeugt-2-8-zettabyte-daten-a-872280.html>)





SuperMUC – 155.000 Cores, 3,2 Petaflop Peak Performance  
Ca. 3.200.000.000.000.000 Gleitkommaoperationen / Sekunde

iPhone 4s: 2 cores, ca. 120 Mflop/s = 120.

SuperMUC => 26 Millionen Mal schneller als iPhone 4s

Hauptspeicher: 340 Tbyte

Stromverbrauch: bis zu 3 Mwatt

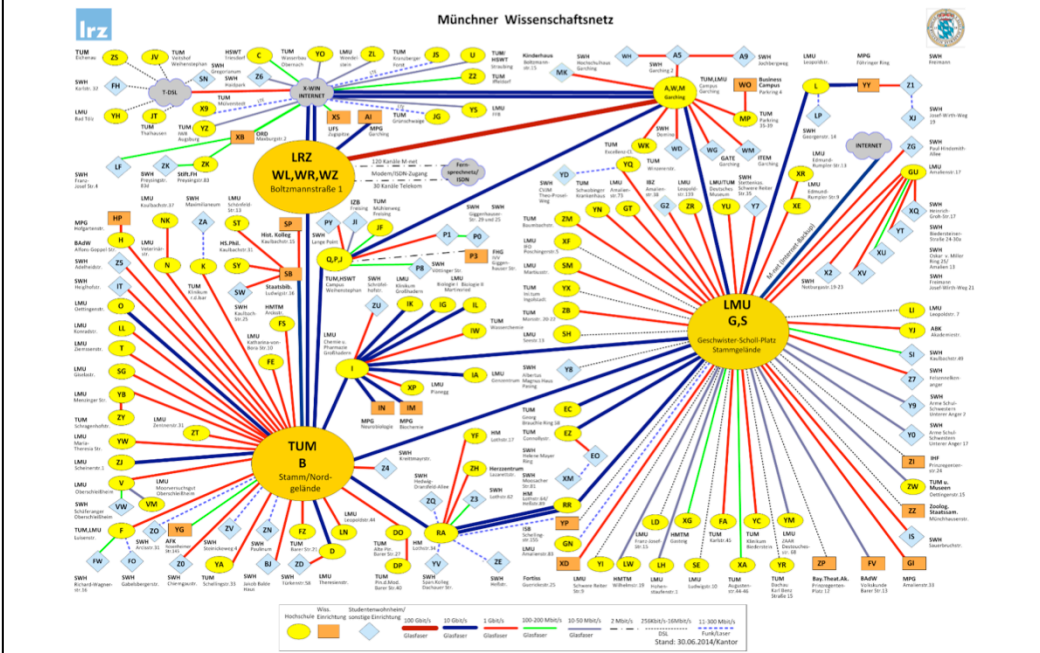
Stromverbrauch SuperMUC 2012 Durchschnitt: ~2 MW/h: 17.520.000 kWh/a  
~ 2825 Einwohner

Stromverbrauch pro Einwohner 2011 im EU-Durchschnitt: 6.201 kWh/a



- **Gemeinsames Rechenzentrum für alle Münchner Hochschulen**
  - 160 Mitarbeiter, 43 wiss. und stud. Hilfskräfte
  - Über 90.000 Studierende,
  - 30.000 Professoren, wiss. und nichtwissenschaftliche Angestellte
  
- **IT-Services für Universitäten, Kunst- und Fachhochschulen in München und andere Großkunden wie BSB, BVB, Hochschulstart**
  - Münchner Wissenschaftsnetz (MWN)
  - Mail-, Web-, Hosting-, Cloud- und Storage Cloud- Services, Landeslizenzen, Spezialdienste

# Das Münchner Wissenschaftsnetz (MWN)



### ■ Kennzahlen

- 16 Router
- 1.460 Switches
- 2.500 Accesspoints
- 72 gemietete dark fibre Leitungen
- 40+ private dark fibre Leitungen
- >100.000 Endgeräte
- 50 Lokationen mit 540 Gebäuden

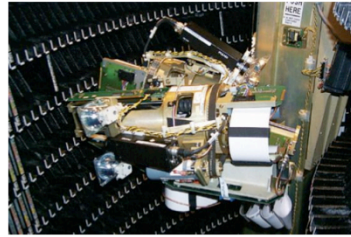
### ■ Verbindet Münchens Universitäten mit 10 Gbit/s mit dem Internet

- 1.100/480 TByte/Monat ein-/ausgehend
- 20 PByte/Monat über Backbone

- **Gemeinsames Rechenzentrum für alle Münchner Hochschulen**
  - 160 Mitarbeiter, 43 wiss. und stud. Hilfskräfte
  - Über 90.000 Studierende,
  - 30.000 Professoren, wiss. und nichtwissenschaftliche Angestellte
  
- **IT-Services für Universitäten, Kunst- und Fachhochschulen in München und andere Großkunden wie BSB, BVB, Hochschulstart**
  - Münchner Wissenschaftsnetz (MWN)
  - Mail-, Web-, Hosting-, Cloud- und Storage Cloud- Services, Landeslizenzen, Spezialdienste
  - Backup- und Archivzentrum (33 Petabyte, über 18 Milliarden Dateien)

### ■ Belegung

- 33 Petabyte Daten
- 18 Milliarden Dateien
- 82 TB eingehende Daten pro Tag
- 9.363 registrierte Systeme



### ■ Ausstattung

- 5 Bandbibliotheken
- 126 Bandlaufwerke
- 46.000 Bandkassetten
- Ca. 3 PB Platten Cache



Dezimalpräfixe	
Name (Symbol)	Bedeutung <sup>[G 1]</sup>
Kilobyte (kB) <sup>[G 2]</sup>	$10^3$ Byte = 1000 Byte
Megabyte (MB)	$10^6$ Byte = 1 000 000 Byte
Gigabyte (GB)	$10^9$ Byte = 1 000 000 000 Byte
Terabyte (TB)	$10^{12}$ Byte = 1 000 000 000 000 Byte
Petabyte (PB)	$10^{15}$ Byte = 1 000 000 000 000 000 Byte
Exabyte (EB)	$10^{18}$ Byte = 1 000 000 000 000 000 000 Byte
Zettabyte (ZB)	$10^{21}$ Byte = 1 000 000 000 000 000 000 000 Byte
Yottabyte (YB)	$10^{24}$ Byte = 1 000 000 000 000 000 000 000 000 Byte

Quelle: <http://de.wikipedia.org/wiki/Byte>



Bilder: [https://www.lrz.de/services/compute/linux-cluster/lx\\_timeline/](https://www.lrz.de/services/compute/linux-cluster/lx_timeline/)



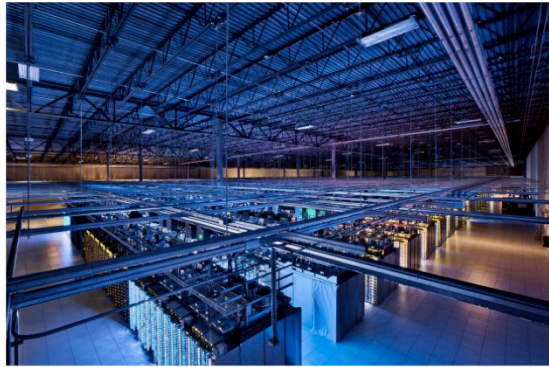
Erster Linux-Cluster (1999-2002)

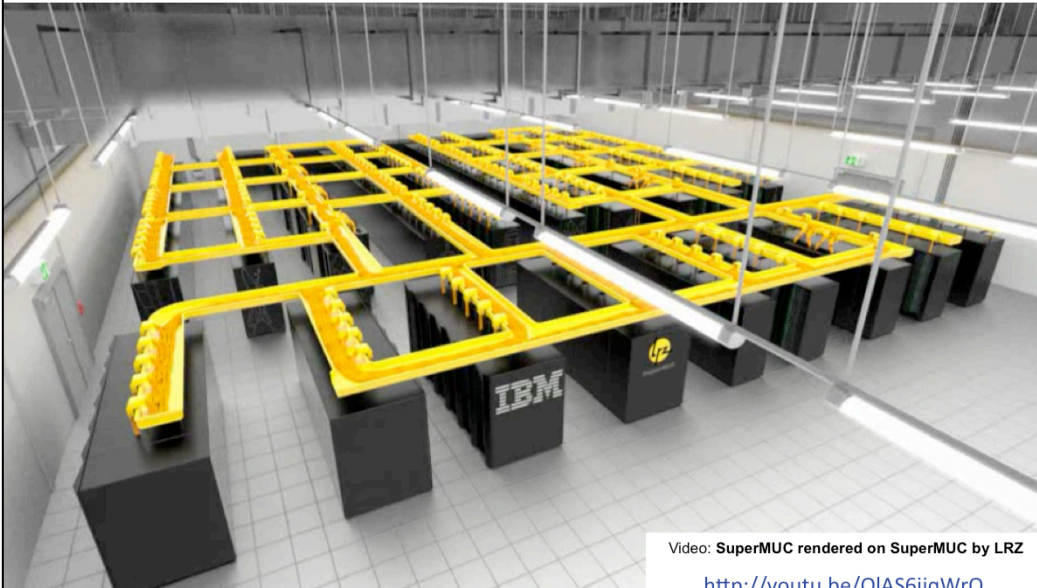
Linux-Cluster (2012)





<http://www.golem.de/news/datacenter-google-zeigt-seine-rechenzentren-1210-95143.html>  
<http://www.google.com/about/datacenters/gallery/#/>

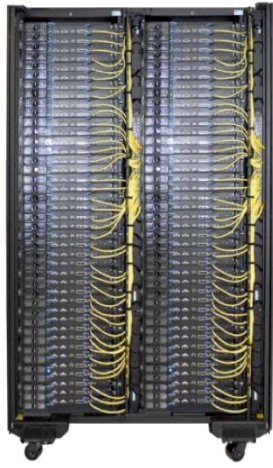




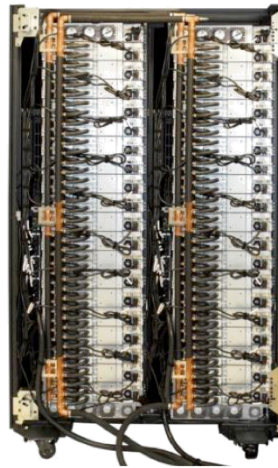
Video: SuperMUC rendered on SuperMUC by LRZ

<http://youtu.be/OIAS6iiqWrQ>



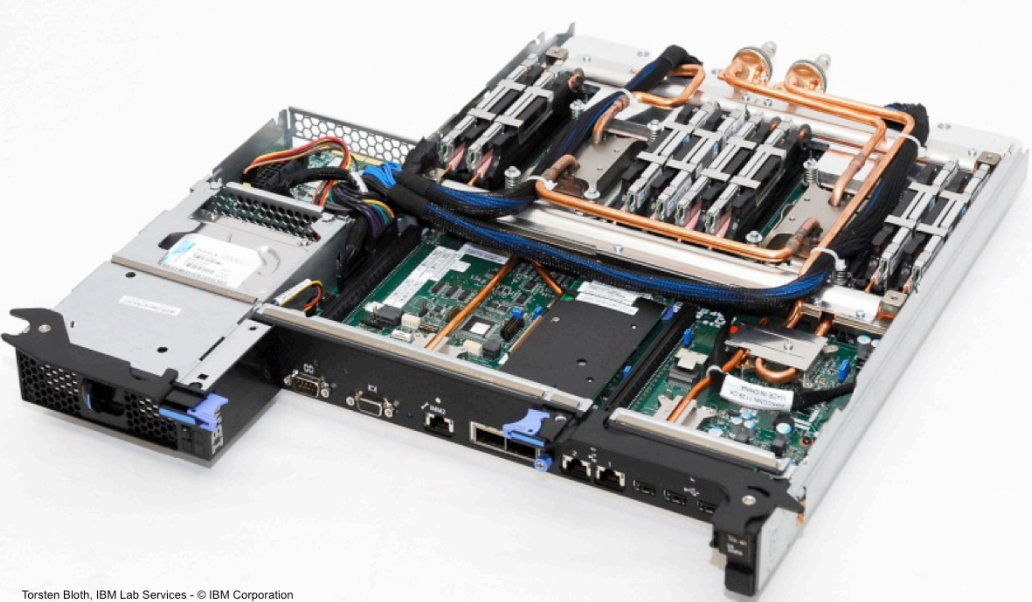


iDataPlex DWC Rack  
w/ water cooled nodes  
(front view)



iDataPlex DWC Rack  
w/ water cooled nodes  
(rear view of water manifolds)

Torsten Bloth, IBM Lab Services - © IBM Corporation





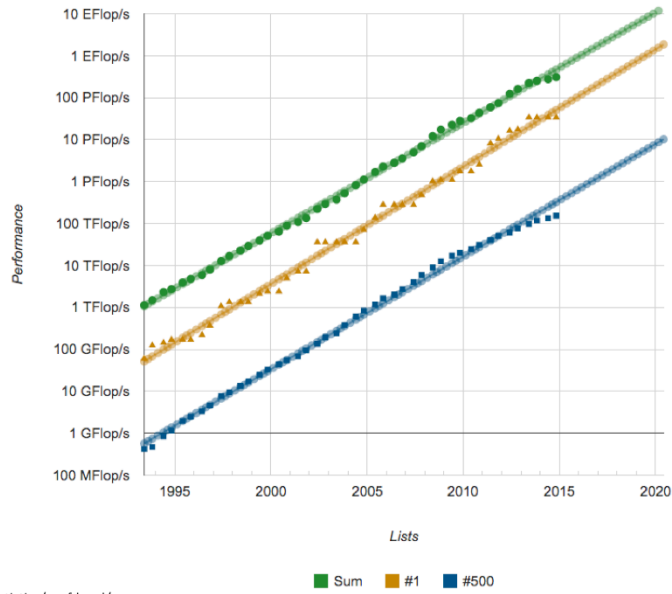
Photos: SIBAM2 (staatl. Hochbauamt München 2)

Rank	Site	Computer/Year Vendor	Cores	R <sub>max</sub>	R <sub>peak</sub>	Power
1	DOE/NNSA/LLNL United States	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom / 2011 IBM	1572864	16324.75	20132.66	7890.0
2	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIII/x 2.0GHz, Tofu interconnect / 2011 Fujitsu	705024	10510.00	11280.38	12659.9
3	DOE/SC/Argonne National Laboratory United States	Mira - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	786432	8162.38	10066.33	3945.0
4	Leibniz Rechenzentrum Germany	SuperMUC - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR / 2012 IBM	147456	2897.00	3185.05	3422.7
5	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050 / 2010 NUDT	186368	2566.00	4701.00	4040.0
6	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XK6, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA 2090 / 2009 Cray Inc.	298592	1941.00	2627.61	5142.0
7	CINECA Italy	Fermi - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	163840	1725.49	2097.15	821.9
8	Forschungszentrum Juelich (FZJ) Germany	JuQUEEN - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	131072	1380.39	1677.72	657.5
9	CEA/TGCC-GENCI France	Curie thin nodes - Bullx B510, Xeon E5- 2680 8C 2.700GHz, Infiniband QDR / 2012 Bull	77184	1359.00	1667.17	2251.0
10	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade System, Xeon X5650 6C 2.66GHz, Infiniband QDR, NVIDIA 2050 / 2010 Dawning	120640	1271.00	2984.30	2580.0

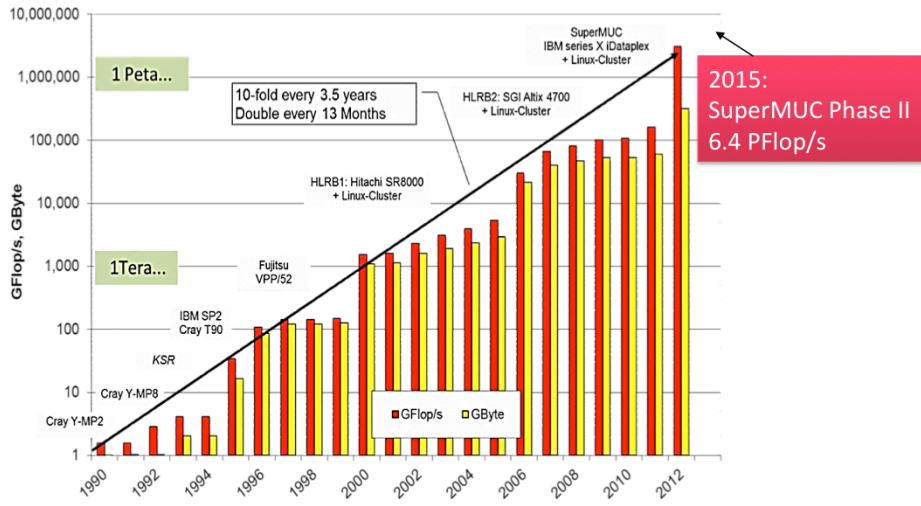
www.top500.org



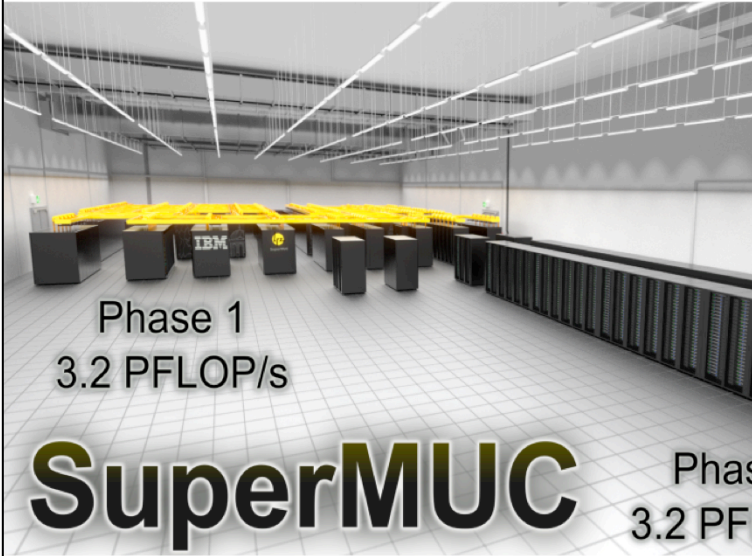
Projected Performance Development



<http://www.top500.org/statistics/perfdevel/>







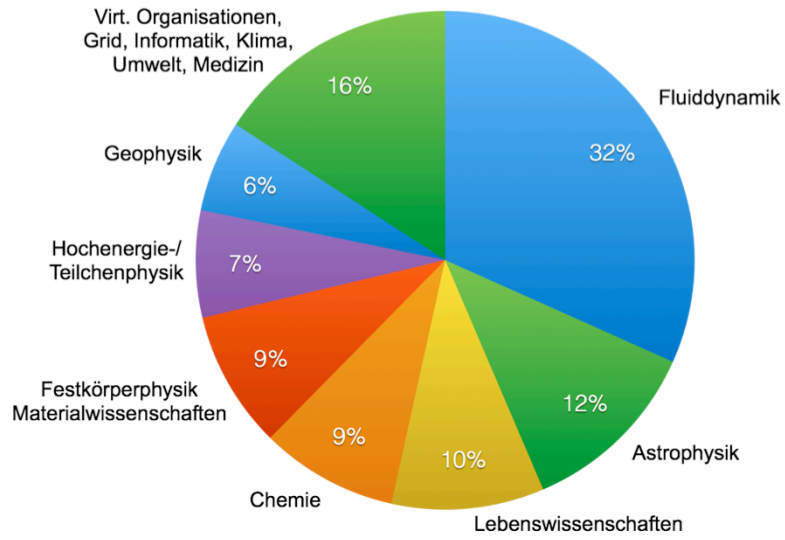
[www.apple.com](http://www.apple.com)

Phase 1  
3.2 PFLOP/s

# SuperMUC

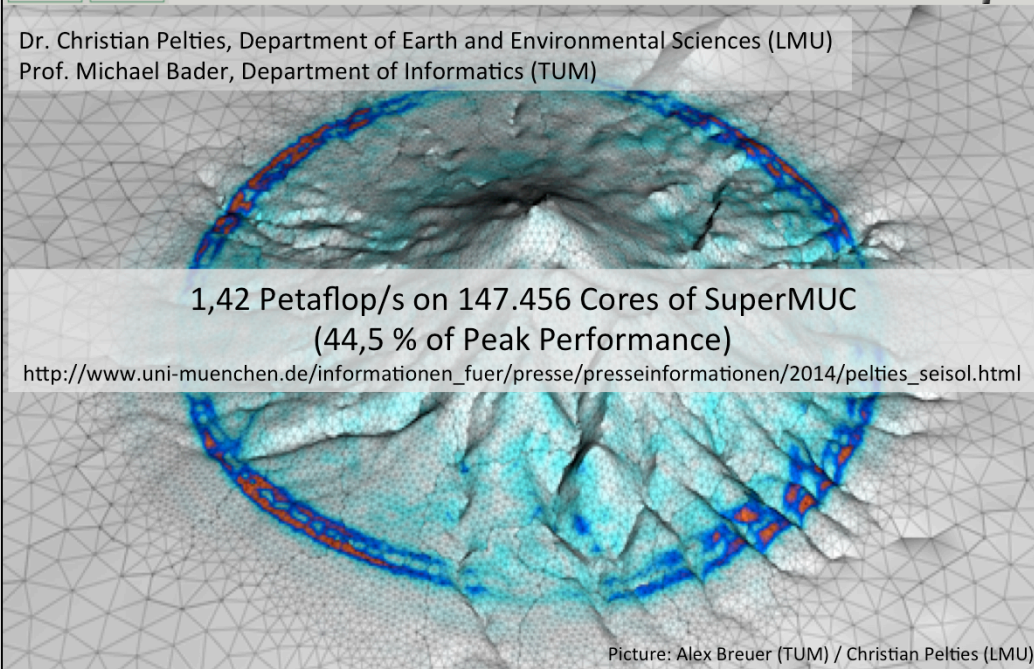
Phase 2  
3.2 PFLOP/s

- Strömungsdynamik: Optimierung von Turbinen, Tragflächenoptimierung, Lärmreduktion, Klimatisierung in Zügen
- Fusionsforschung: Plasma in künftigen Fusionsreaktoren (ITER)
- Astrophysik: Entstehung und Entwicklung von Sternen und Galaxien
- Festkörperphysik: Supraleitung, Oberflächeneigenschaften
- Geophysik: Erdbebenszenarien
- Materialwissenschaften: Halbleiter
- Chemie: Katalysatorreaktionen
- Medizin und Medizintechnik: Blutströmungen, Aneurysmen, Klimatisierung in Operationssälen
- Biowissenschaften: Viruseigenschaften, Genom-Analyse
- Klimatologie: Ozeanströmungen
- ...



**LMU** LUWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN **SeisSol - Numerical Simulation of Seismic Wave Phenomena** **lrz**

Dr. Christian Pelties, Department of Earth and Environmental Sciences (LMU)  
Prof. Michael Bader, Department of Informatics (TUM)



1,42 Petaflop/s on 147.456 Cores of SuperMUC  
(44,5 % of Peak Performance)  
[http://www.uni-muenchen.de/informationen\\_fuer/presse/presseinformationen/2014/pelties\\_seisol.html](http://www.uni-muenchen.de/informationen_fuer/presse/presseinformationen/2014/pelties_seisol.html)

Picture: Alex Breuer (TUM) / Christian Pelties (LMU)

MNM D. Kranzlmüller Energy Efficiency and Extreme Scaling 47

Ludwig-Maximilians-Universität München

Suche

Links und Funktionen

[www.lmu.de](http://www.lmu.de)

[LMU-Portal](#)

[Sitemap](#)

[RSS-Feeds](#)

[Sprachumschaltung](#)

[International Visitors](#)

[Navigationspfad](#)

[Startseite Informationen für: Presse Presseinformationen 2014 Rechenrekord auf dem SuperMUC](#)

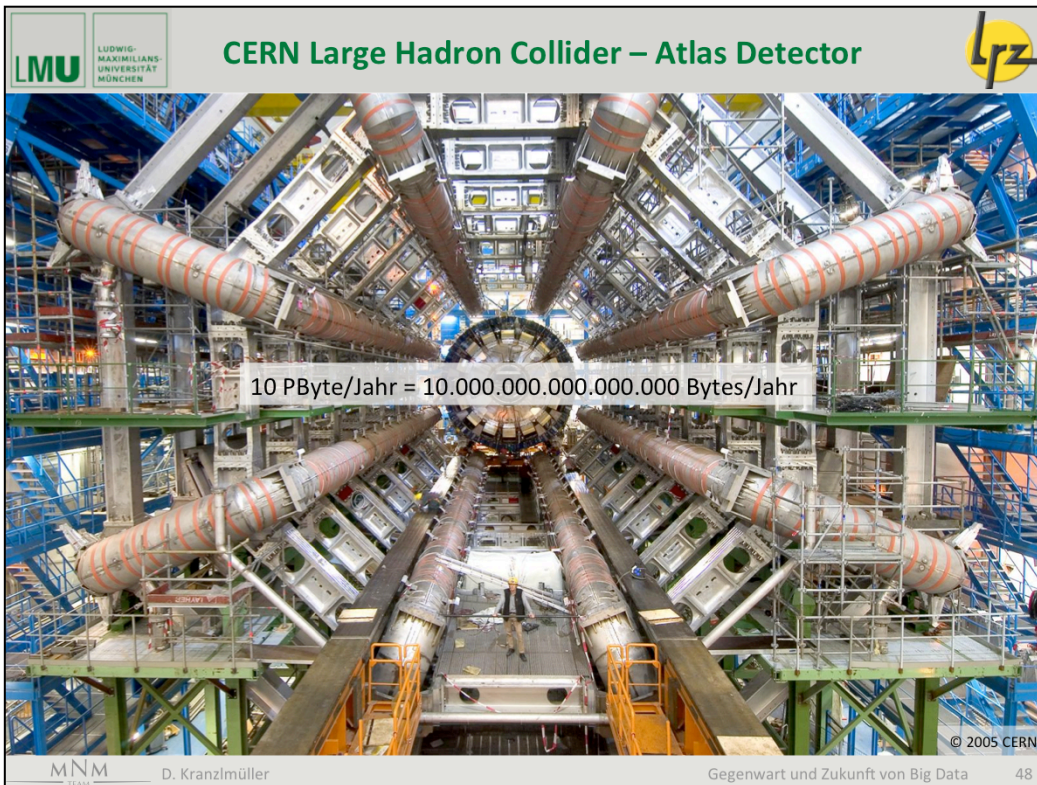
[Hauptnavigation](#)

[AKTUELLES](#)

[ÜBER DIE LMU](#)

[EINRICHTUNGEN](#)

[STUDIUM](#)



Datenstrom ca. 300 GB/s – Filterung auf interessante Ereignisse – Roh Daten: 300 MB/s => 10 PB/Jahr

$10.000.000.000.000.000 = 10^{16}$  Bytes

iPhone 4s: 16 GB =  $16.000.000.000 = 16 \cdot 10^9$

9,3 mm Dicke

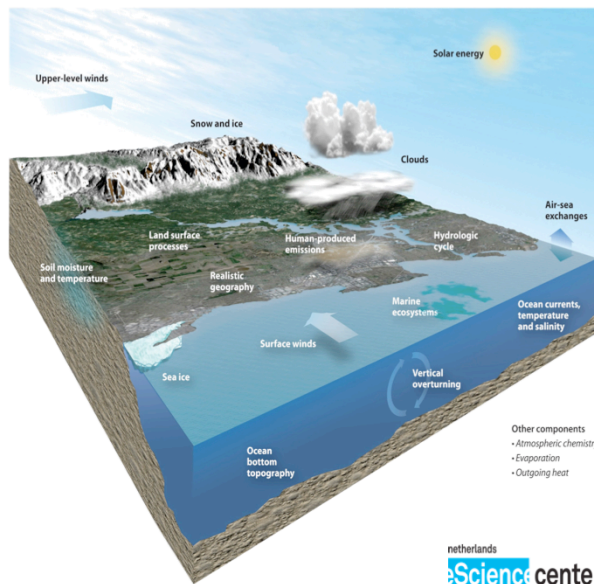
=> 591.300 iPhones = 5499.090 m = 5,5 km Stapel




- Modellierung des Systems "Erde" erfordert eine Kombination vieler Modelle (Atmosphäre, Ozean, Land und Eis)

**Wissenschaftliche Herausforderungen:**


- Komplexität
- Skalierbarkeit
- Datenmenge
- Zuverlässigkeit
- Heterogenität
- ...






**LMU**  
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

## Enlighten Your Research Global (EYRg)




  

**3.2 PFlop/s**



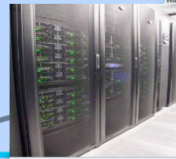
**SUPERMUC(GER)**

**1.0 PFlop/s**



**CARTESIUS(NLD)**


**114 TFlop/s**



**EMERALD(UK)**

**Stampede**  
9050 km



**STAMPEDE(USA)**


**5.1 PFlop/s**

**merald**  
40 km


**Cartesius**

**SuperMUC**  
960 km


  




**IMAU**




**LMU**




**OXFORD e-Research CENTRE**




**UNIVERSITY OF OXFORD**




**NCAR**




**RUTGERS**




**RSMAS**




**VU UNIVERSITY AMSTERDAM**




**lrz**



**UNIVERSITEIT UTRECHT**




**netherlands Science center**



**TACC SURF SARA**

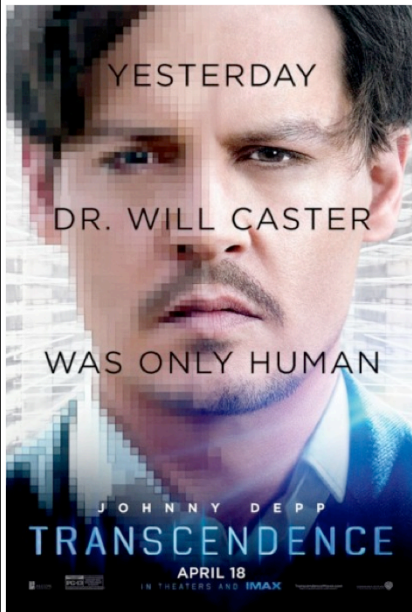


**MNM**

- 1997 schlägt ein Großrechner erstmals den Schachmeister Garry Kasparov (IBM's Deep Blue, der in der Sekunde über 200 Millionen mögliche Spielzüge berechnen konnte)
- Watson beruht darauf natürliche menschliche Sprache zu verstehen, deren Wörter und Kontext zu analysieren, diese Informationen schnell zu verarbeiten und so Antworten auf Fragen in natürlicher Sprache auszugeben
- 2011 schlägt Watson zwei Jeopardy!-Champions



Quelle: [www.ibm.com/watson](http://www.ibm.com/watson)



<http://www.transcendencemovie.com/>

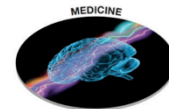
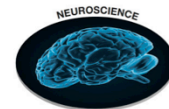
*“Understanding the human brain is one of the greatest challenges facing 21st century science. If we can rise to the challenge, we can gain profound insights into what makes us human, develop new treatments for brain disease and build revolutionary new computing technologies. Today, for the first time, modern ICT has brought these goals within sight.”*




Human Brain Project

*“Molecular-level models require hundreds of exabytes of memory to simulate”*


<https://www.humanbrainproject.eu/>



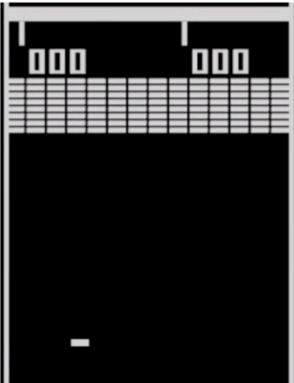


LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN


## Neuronale Netze lernen spielen



- Computerspiel: Breakout
- Atari, April 1976



- *„Human-level control through deep reinforcement learning“*  
 Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dhharshan Kumaran, Daan Wierstra, Shane Legg & Demis Hassabis
- Nature 518, 529–533 (26 February 2015)
- doi:10.1038/nature14236



D. Kranzlmüller

Gegenwart und Zukunft von Big Data

53

Oder: <http://godwincaruana.me/artificial-intelligence-bests-humans-at-classic-arcade-games-science-magazine/>

[http://de.wikipedia.org/wiki/Breakout\\_%28Computerspiel%29](http://de.wikipedia.org/wiki/Breakout_%28Computerspiel%29)

Leitender Entwickler: Steve Wozniak

Das erste Breakout-Spiel war, wie viele frühe Videospiele, noch kein echtes Computerspiel, da es sich nicht um einen Computer mit einem darauf ablaufenden Programm handelte, sondern um einen das Spielprinzip direkt umsetzenden elektronischen Schaltkreis. Diese Spielelektronik wurde von Steve Wozniak entworfen, während das Game-Design von Nolan Bushnell stammte. Nolan Bushnell zeichnete das Spielprinzip auf eine Tafel auf. Steve Jobs, der damals bei Atari arbeitete, überredete seinen Freund Steve Wozniak (damals bei HP), dieses Spiel zu konstruieren. Steve Jobs bekam für das Spiel Breakout von Nolan Bushnell 5.000 Dollar bezahlt. Er gab seinem Freund Steve Wozniak, der das eigentliche Spiel konstruiert hat, nur 350 Dollar. Steve Wozniak benötigte nur 45 TTL-Bauteile, um das Spiel Breakout zu konstruieren. Fast alle späteren Versionen waren dann normale

LMU



rz

MNM  
D. Kranzmüller

Gegenwart und Zukunft von Big Data

54

- Ziel:  
*„To build a never-ending machine learning system that acquires the ability to extract structured information from unstructured web pages. If successful, this will result in a knowledge base (i.e., a relational database) of structured information that mirrors the content of the Web.“*

- NELL (Never-Ending Language Learner)

- Seit Jänner 2010

- Carnegie Mellon University

NELL Knowledge Base Browser

Category	Item	Date	Score
cheese	brin_4_amour_cheese	09-jul-2011	100.0
meat	brillat_sasavin_cheese	09-jul-2011	100.0
candy	brin_cheese	24-jun-2011	100.0
consumerelectronicitem	brin_4_amour_cheese	09-jul-2011	100.0
videogamesystem	brin_4_amour_cheese	24-nov-2011	100.0
product	brin_4_amour_cheese	05-sep-2014	100.0
vehicle	brin_4_amour_cheese	05-sep-2014	100.0
automobilemodel	brin_4_amour_cheese	05-sep-2014	100.0
software	broccio_demi_affine_cheese	19-jun-2011	100.0
musicsoftware	broccio_demi_affine_cheese	19-jun-2011	100.0
videogame	bruder_basil_cheese	11-sep-2011	100.0
personalcareitem	bruder_basil_cheese	11-sep-2011	100.0
drug	bruder_basil_cheese	11-sep-2011	100.0
automobileengine	bruder_basil_cheese	11-sep-2011	100.0
tool	bruder_basil_cheese	11-sep-2011	100.0
mediatype	bruder_basil_cheese	11-sep-2011	100.0
clothing	bruder_basil_cheese	11-sep-2011	100.0
buildingfeature	bruder_basil_cheese	11-sep-2011	100.0
fungus	bruder_basil_cheese	11-sep-2011	100.0
plant	bruder_basil_cheese	11-sep-2011	100.0
sportsequipment	bruder_basil_cheese	11-sep-2011	100.0
officeitem	bruder_basil_cheese	11-sep-2011	100.0
visualizablething	bruder_basil_cheese	11-sep-2011	100.0
visualizableattribute	bruder_basil_cheese	11-sep-2011	100.0
color	bruder_basil_cheese	11-sep-2011	100.0
geometricshape	bruder_basil_cheese	11-sep-2011	100.0
visualizablescene	bruder_basil_cheese	11-sep-2011	100.0
landscapfeatures	bruder_basil_cheese	11-sep-2011	100.0
city	bruder_basil_cheese	11-sep-2011	100.0
visualizableobject	bruder_basil_cheese	11-sep-2011	100.0
weapon	bruder_basil_cheese	11-sep-2011	100.0
musicinstrument	bruder_basil_cheese	11-sep-2011	100.0
food (*)	bruder_basil_cheese	11-sep-2011	100.0
vehicle (*)	bruder_basil_cheese	11-sep-2011	100.0
personalcareitem	bruder_basil_cheese	11-sep-2011	100.0
sportsequipment	bruder_basil_cheese	11-sep-2011	100.0

http://rtw.ml.cmu.edu/rtw/

- Von Computern betriebener Handel mit Wertpapieren
- Hochleistungsrechner handeln selbständig nach Algorithmen
- Marktveränderungen führen zu Handelsentscheidungen innerhalb weniger Microsekunden
  
- Tyler Durden: „Goldman’s \$4 Billion High Frequency Trading Wildcard“. In: Zero Hedge. 17. Juli 2009  
<http://zerohedge.blogspot.de/2009/07/goldmans-4-billion-high-frequency.html>
  
- Deutschland: Bundesrat, 28. Februar 2013  
*„Gesetz zur Vermeidung von Gefahren und Missbräuchen im Hochfrequenzhandel“* (Hochfrequenzhandelsgesetz)



TESLA MODEL S MODEL X MODEL W

## Blog

### Announcing the Tesla Model W

April 1, 2015

PALO ALTO, Calif., April 1, 2015 – Tesla today announced the Tesla Model W. As many in the media predicted, it's a watch that tells time in a way that is infinitely adjustable, able to tell the time in Timbuktu, California, anywhere! This will change your life the same.

Musk at The Summit 2013 in Dublin  
[http://en.wikipedia.org/wiki/Elon\\_Musk](http://en.wikipedia.org/wiki/Elon_Musk)

Tesla Motors © 2015 | Legal | Contact | Careers | United States <http://www.teslamotors.com/>

MNM D. Kranzlmüller Gegenwart und Zukunft von Big Data 57

FEHLER VON NACHRICHTENAGENTUREN

# Teslas Aprilscherz rüttelt die Aktie durch

Autor: Martin Dowideit  
Datum: 02.04.2015 13:20 Uhr

Der Aprilscherz von Tesla war ein plumper Seitenhieb auf Apple. Doch mehrere Nachrichtenagenturen hielten die Schlagzeile für wahr und meldeten ein neues Modell des Elektroauto-Herstellers. Die Börse spielte verrückt.

Ein neues Elektroauto aus Kalifornien? Automatisierte Handelsalgorithmen setzen solche Schlagzeilen automatisch in Kaufbefehle um - und in Sekundenbruchteilen sprang die Tesla-Aktie um 0,75 Prozent nach oben. Die von Elon Musk gegründete Firma war auf einen Schlag 125 Millionen Dollar mehr wert.

<http://www.handelsblatt.com/finanzen/maerkte/aktien/fehler-von-nachrichtenagenturen-teslas-aprilscherz-ruettelt-die-aktie-durch/11591032.html>

- Alan Turing, 23 Juni 1912 – 7 Juni 1954
- [http://en.wikipedia.org/wiki/Alan\\_Turing](http://en.wikipedia.org/wiki/Alan_Turing)
  - Theoretische Grundlagen für die moderne Informations- und Computertechnologie
  - Entzifferung der Enigma-Verschlüsselung (Bletchley Park)
  - Entwicklung eines der ersten Schachprogramme



<http://www.turingarchive.org/viewer/?id=521&title=4>

- 1950: Turing-Test zum Nachweis künstlicher Intelligenz
- <http://de.wikipedia.org/wiki/Turing-Test>

■ Testablauf:

- Menschlicher Fragesteller führt über eine Tastatur und einen Bildschirm ohne Sicht- und Hörkontakt mit zwei ihm unbekanntem Gesprächspartnern eine Unterhaltung.
- Ein Gesprächspartner ist ein Mensch, der andere Maschine.
- Beide versuchen, den Fragesteller zu überzeugen, dass sie denkende Menschen sind.
- Wenn der Fragesteller nach der intensiven Befragung nicht klar sagen kann, welcher von beiden die Maschine ist, hat die Maschine den Turing-Test bestanden, und es wird der Maschine ein dem Menschen ebenbürtiges Denkvermögen unterstellt.

- 1988: KI-Webapplikation, die den Inhalt der Gespräche durch Kommunikation mit Menschen lernt
- Anders als Smartphone Assistenten: Siri, Cortana
- 2013: 20 Millionen Menschen als Gesprächspartner
- Datenschatz: 180 Millionen Gesprächszeilen, + 1 Million/Woche



48216 people talking

Hallo Cleverbot!

Hallo.

Wie geht's Dir heute?

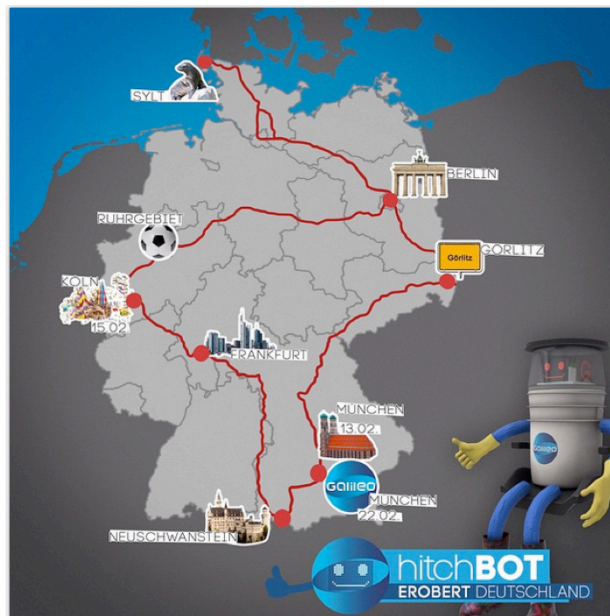
Gut und dir?

Danke. Auch gut.

Bitte. ●●●●● 

- 3. Sept. 2011: Technische Festival
- Indisches Institut IIT Guwahati
- 1334 Stimmen
- Cleverbot wurde **zu 59,3 % als menschlich erkannt**
- Menschliche Konkurrenten erzielten 63,3 %
- NewScientist, September 2011:
- <http://www.newscientist.com/article/dn20865-software-tricks-people-into-thinking-it-is-human.html>

- <http://www.hitchbot.me/>
- „I am hitchBOT — a robot from Port Credit, Ontario.“
- 27. Juli – 21. Aug. 2014
  - 10.000 km durch Kanada
  - 19 Mitfahrgelegenheiten
- 13-22. Februar 2015
  - Deutschland-Tour
  - Galileo/Pro Sieben



<https://instagram.com/p/yrxl2qPiqo/>

- Futurologie (Zukunftsforschung)
- **Singularität** = Zeitpunkt, ab dem sich Maschinen mittels **künstlicher Intelligenz (KI)** selbst verbessern können
- Beschleunigung des technischen Fortschritts
- Zukunft der Menschheit nicht mehr vorhersehbar
  
- 1958: Stanislaw Ulam im Gespräch mit John von Neumann:  
*„Ein Gespräch drehte sich um die stete Beschleunigung des technischen Fortschritts und der Veränderungen im Lebenswandel, die den Anschein macht, auf eine entscheidende Singularität in der Geschichte der Menschheit hinauszulaufen, nach der die Lebensverhältnisse, so wie wir sie kennen, sich nicht fortsetzen könnten.“*
  
- 1993: Vernor Vinge, „*Technological Singularity*“
- 2001: Ray Kurzweil, „*The Law of Accelerating Returns*“

[http://de.wikipedia.org/wiki/Technologische\\_Singularit%C3%A4t](http://de.wikipedia.org/wiki/Technologische_Singularit%C3%A4t)

- Isaac Asimov
- 2. Jänner 1920 – 6. April 1992
- 1942: „Runaround“ (Erzählung)
- 1950: „I, Robot“
  
- Drei Gesetze der Robotik:
  1. Ein Roboter darf keinen Menschen verletzen oder durch Untätigkeit zu Schaden kommen lassen.
  2. Ein Roboter muss den Befehlen eines Menschen gehorchen, es sei denn, solche Befehle stehen im Widerspruch zum ersten Gesetz.
  3. Ein Roboter muss seine eigene Existenz schützen, solange dieser Schutz nicht dem Ersten oder Zweiten Gesetz widerspricht.



Bild: Phillip Leonian from New York World-Telegram & Sun.

„Isaac.Asimov01“ von Phillip Leonian [1] from New York World-Telegram & Sun.[2] - United States Library of Congress. New York World-Telegram and the Sun Newspaper Photograph Collection. Call number: NYWTS - BIOG--Asimov, Isaac, Dr. <item> [P&P]. Reproduction number: LC-USZ62-115121. Lizenziert unter Gemeinfrei über Wikimedia Commons - <http://commons.wikimedia.org/wiki/File:Isaac.Asimov01.jpg#/media/File:Isaac.Asimov01.jpg>



„Hubble ultra deep field“ von NASA and the European Space Agency. - <http://hubblesite.org/newscenter/archive/releases/2004/07/image/a/warn/>.  
Lizenziert unter Gemeinfrei über Wikimedia Commons - [http://commons.wikimedia.org/wiki/File:Hubble\\_ultra\\_deep\\_field.jpg#/media/File:Hubble\\_ultra\\_deep\\_field.jpg](http://commons.wikimedia.org/wiki/File:Hubble_ultra_deep_field.jpg#/media/File:Hubble_ultra_deep_field.jpg)



### Datenebene

- Welche Daten sollten gespeichert werden?
- Wer hat Zugang zu den Datensätzen?
- Wer kann die Daten nutzen?
- ...

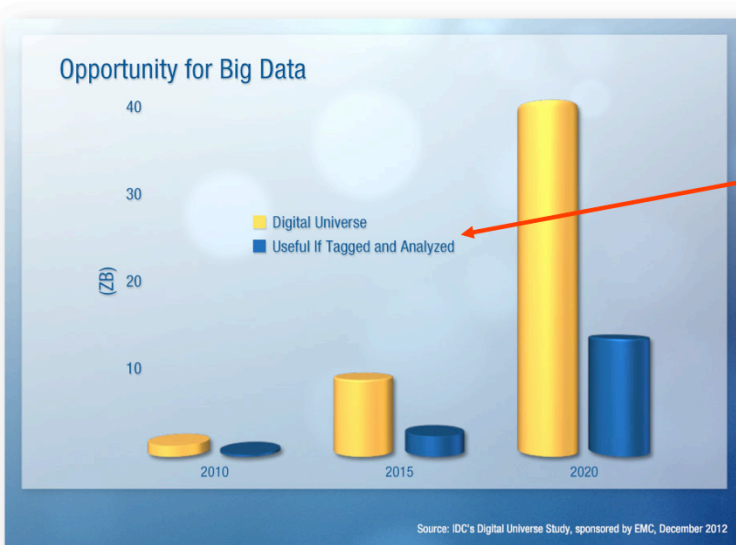
### Technologieebene

- Können die gewünschten Auswertungen kosteneffizient erstellt werden?
- Sind die zukünftig benötigten IT-Kompetenzen bei Mitarbeitern und Unternehmen ausreichend vorhanden?
- ...

### Nutzungsebene

- Welche Daten sind oder werden u.U. personenbezogen?
- Wie kann die gesellschaftliche Akzeptanz erreicht werden?
- ...

Folie: Arnold Picot, LMU



Wer definiert „Nutzen“?

Quelle: The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East; IDC December 2012  
<http://germany.emc.com/leadership/digital-universe/iview/big-data-2020.htm>

