

sas<sup>®</sup> forum  
**SWITZERLAND 2012**

make connections • share ideas • be inspired

# BIG Data

*Trend oder Hype?*



THE  
POWER  
TO KNOW.

Guido Oswald  
Solution Architect @SAS Switzerland

# BIG Data..... Wer?



# BIG Data..... Wer?



# Wikipedia sagt:

- Als **Big Data** werden besonders *große Datenmengen* bezeichnet, die mit Hilfe von Standard-Datenbanken und Daten-Management-Tools nicht oder nur unzureichend verarbeitet werden können.
- Problematisch sind hierbei vor allem die *Erfassung*, die *Speicherung*, die *Suche*, *Verteilung*, *Analyse* und *Visualisierung* von großen Datenmengen.
- Das Volumen dieser Datenmengen geht in die **Terabytes**, **Petabytes**, **Exabytes** und **Zettabytes**.

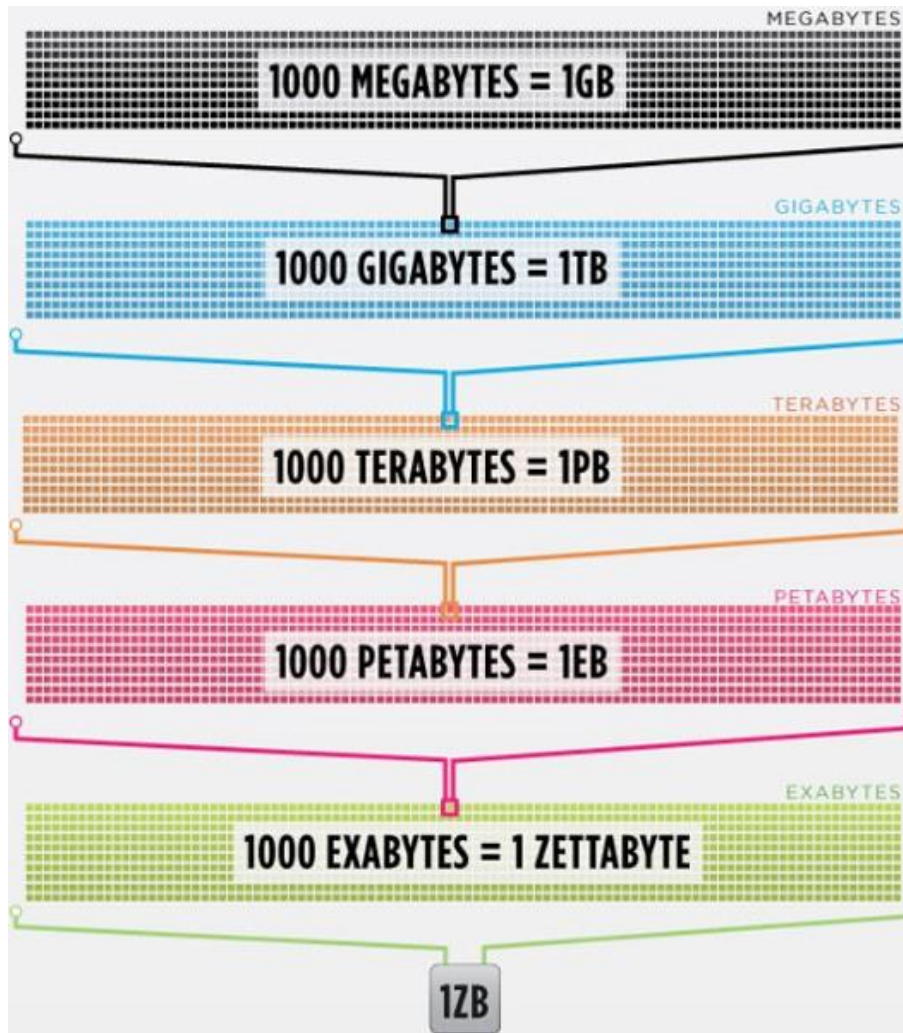


# Ist BIG Data relevant für mich?

- Zettabytes ?!
  - $10^{21}$  Byte
  - 1.000.000.000.000.000.000.000 Byte



# Zettabytes



engadget 

2011

1.8  
Zettabytes

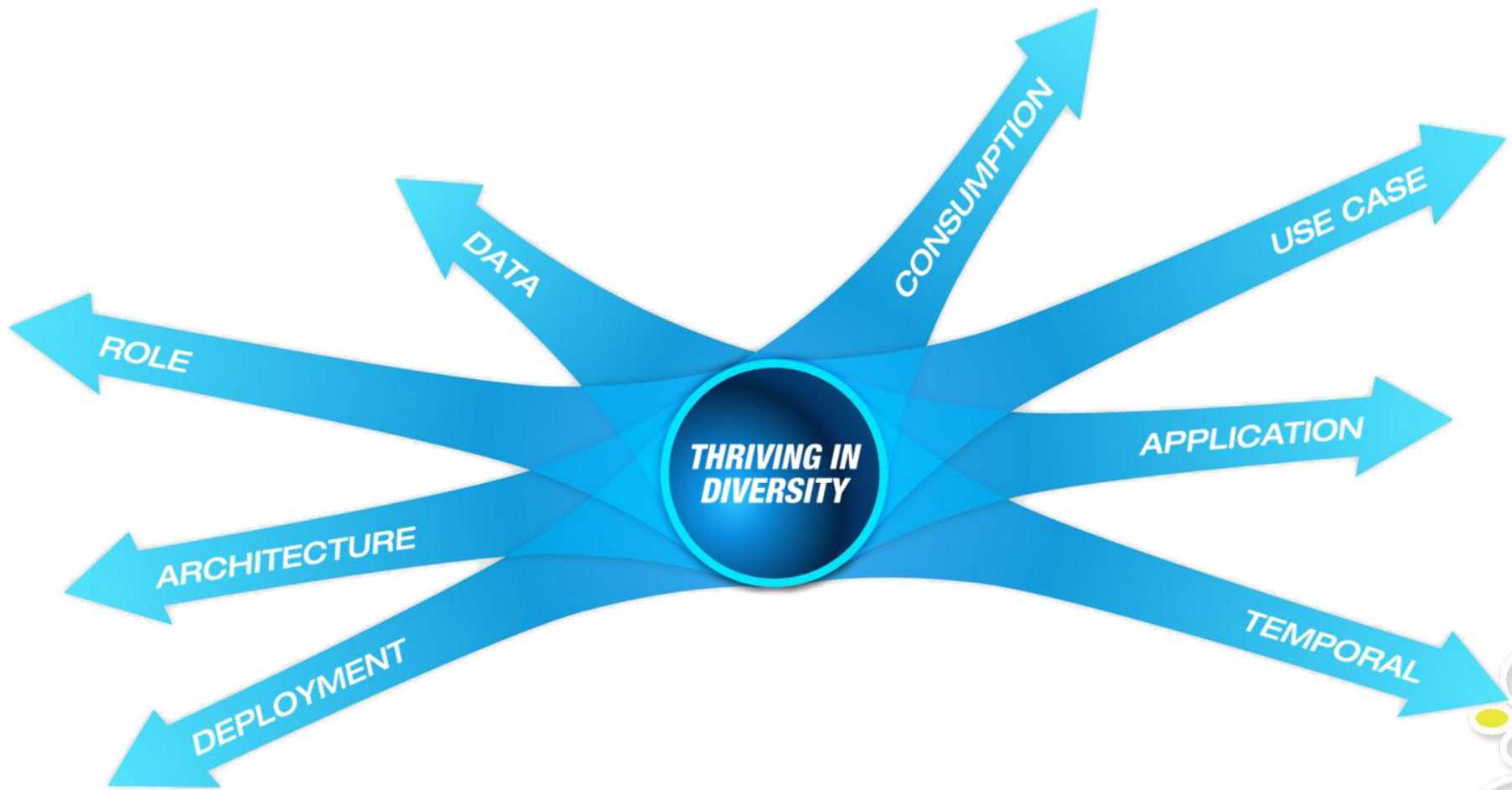
2020

79.2  
Zettabytes


44x

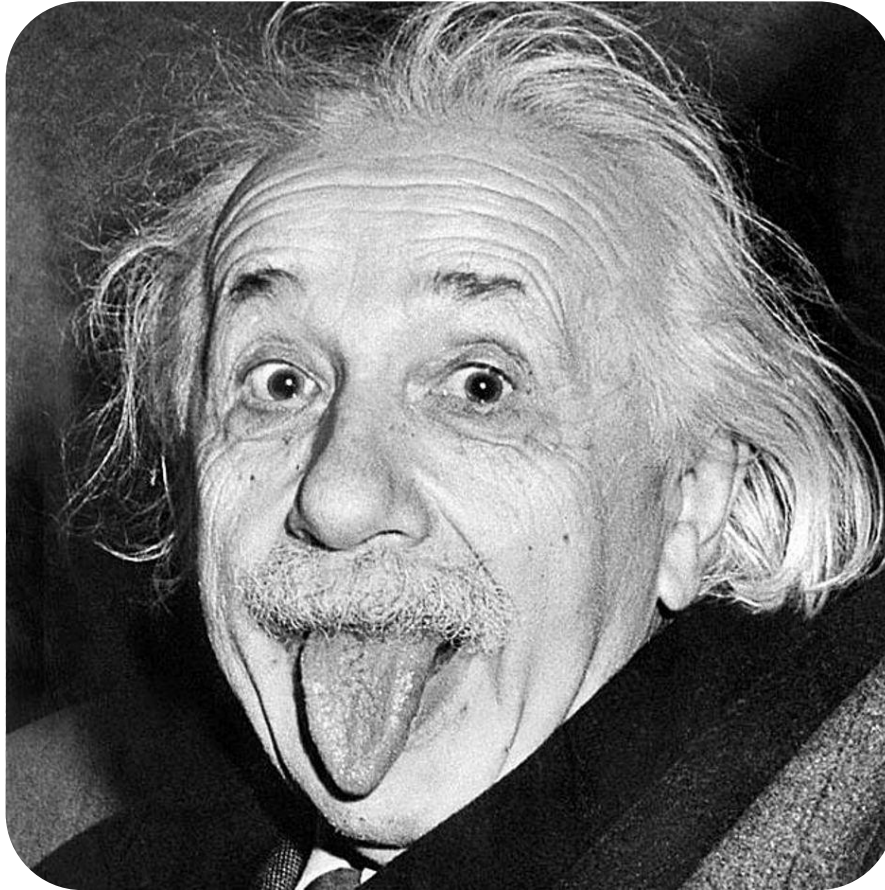


# Diversität der Daten wächst!

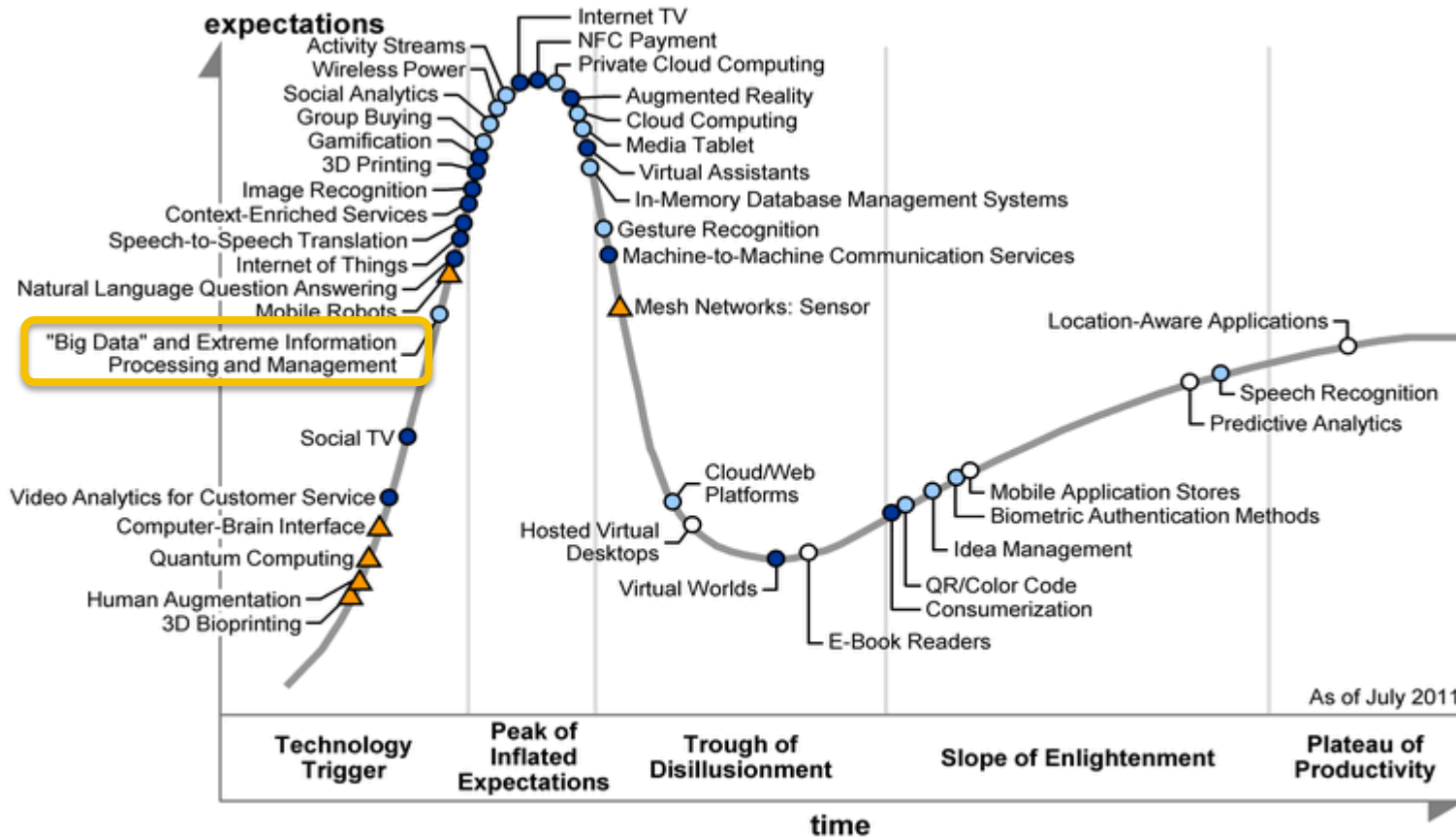





$$E = mc^2$$



# Gartner "Hype Cycle"



Years to mainstream adoption:

○ less than 2 years   ● 2 to 5 years   ● 5 to 10 years   ▲ more than 10 years   ⊗ obsolete before plateau

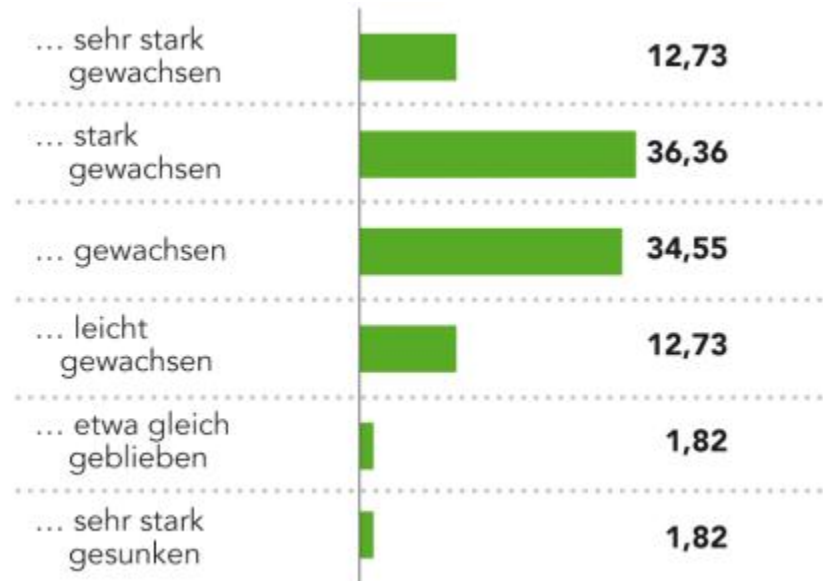
# Gartner IT Trends 2012

1. Tablets
2. Mobile Applikationen und Interfaces
3. Kontextuelles und soziales Benutzererlebnis
4. Das Internet der Dinge
5. App-Stores und Marktplätze
6. Next-Generation-Analytics
7. Big Data
8. In-Memory-Computing
9. Extrem stromsparende Server
10. Cloud Computing



# BIG Data kommt

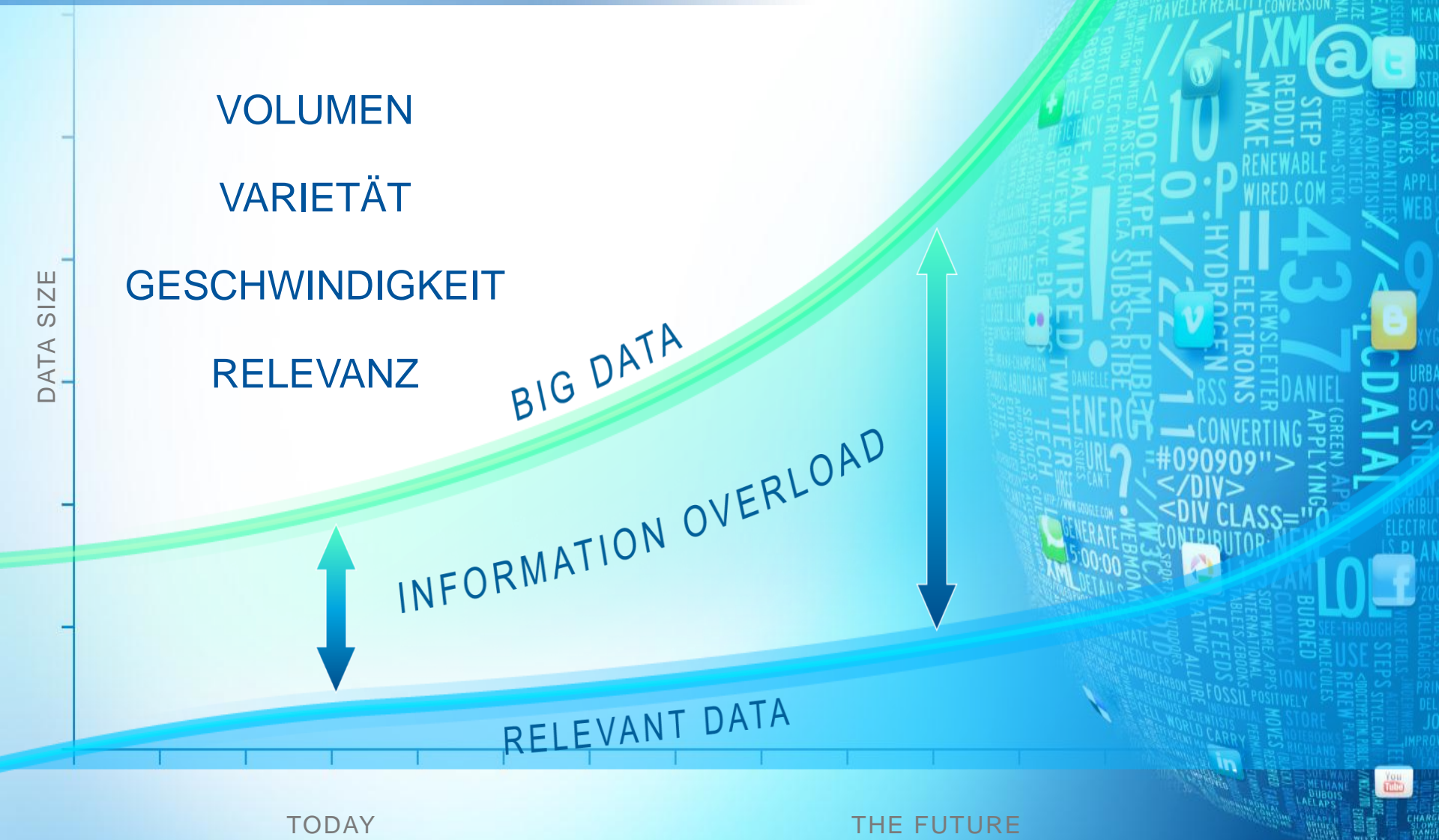
COMPUTERWOCHE



Wie hat sich nach Ihrer Einschätzung die Datenmenge, die in Ihrem Unternehmen verarbeitet wird, in den letzten zwölf Monaten verändert? Die Datenmenge ist ... (n=116 , Angaben in Prozent)

Quelle: CW-Marktstudie 2011: Big Data.

# Wettbewerbsvorteil aus BIG Data



# Analytics Lifecycle und BIG Data

## BUSINESS MANAGER



Domain Expert  
Makes Decisions  
Evaluates Processes and ROI

## BUSINESS ANALYST



Data Exploration  
Data Visualization  
Report Creation

## IT SYSTEMS / MANAGEMENT

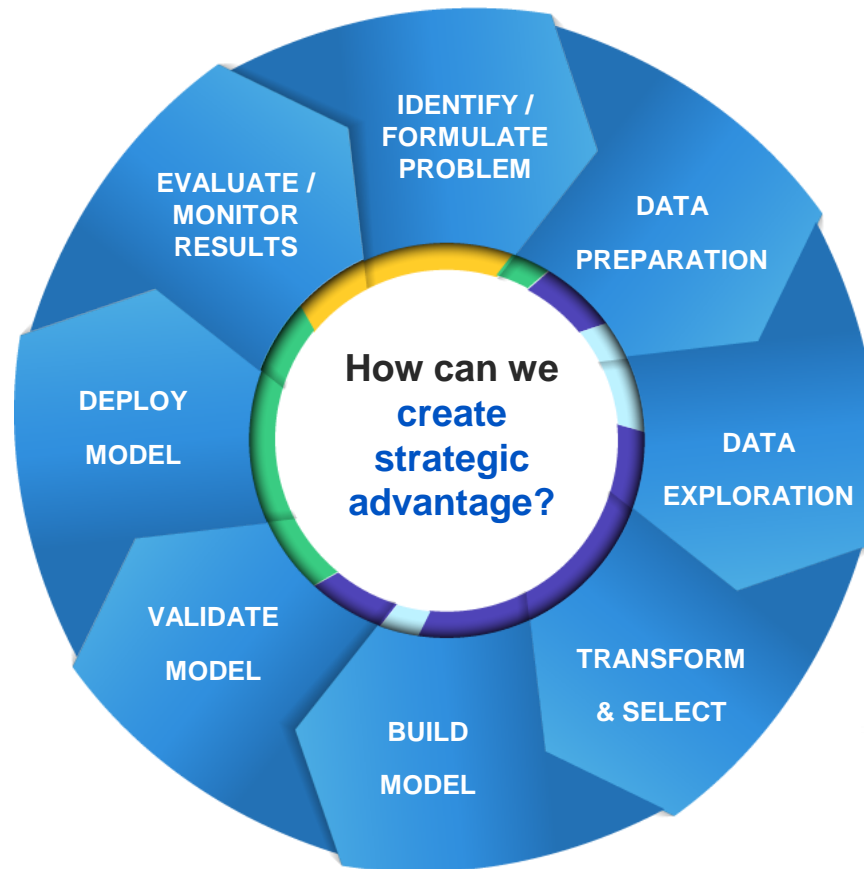


Model Validation  
Model Deployment  
Model Monitoring  
Data Preparation

## DATA MINER / STATISTICIAN



Exploratory Analysis  
Descriptive Segmentation  
Predictive Modeling



# Die Lösung: Supercomputer?



# Die Lösung: Supercomputer?

## “Monte Rosa” im CSCS (Lugano)

- AMD quad core Opteron 2,4 GHz Prozessoren
- 47.872 Kerne
- 297,0 TFLOPS

## ABER:

- Kosten?
- Verteilen der Daten / Verarbeitung?





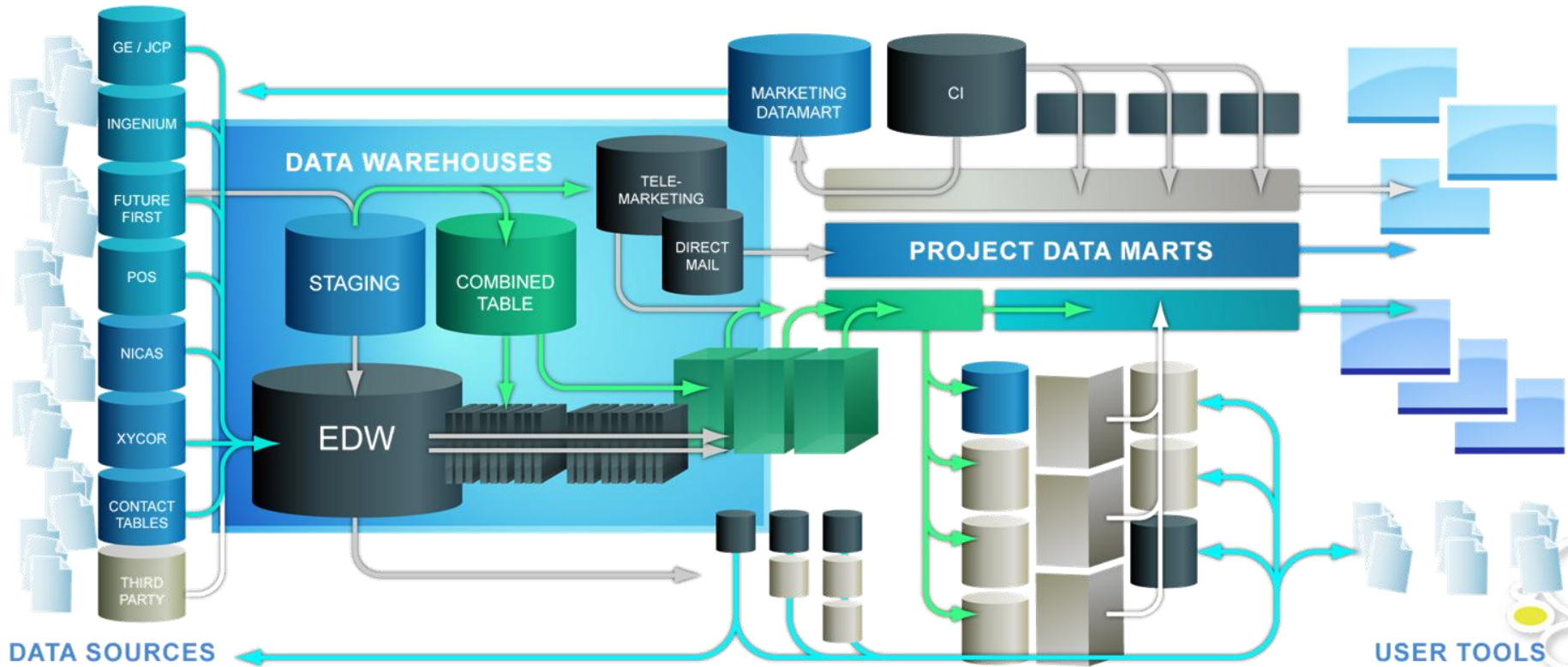
# Lösungsansatz für BIG Analytics

Statt die Daten zur «Arbeit»



# Lösungsansatz für BIG Analytics

## Traditionelle Architektur

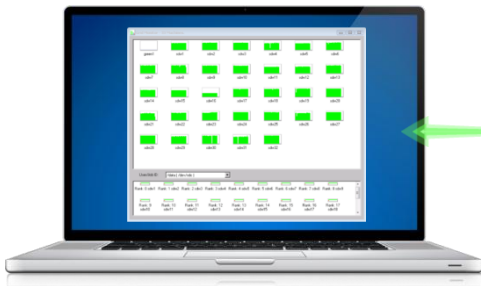


# Lösungsansatz für BIG Analytics

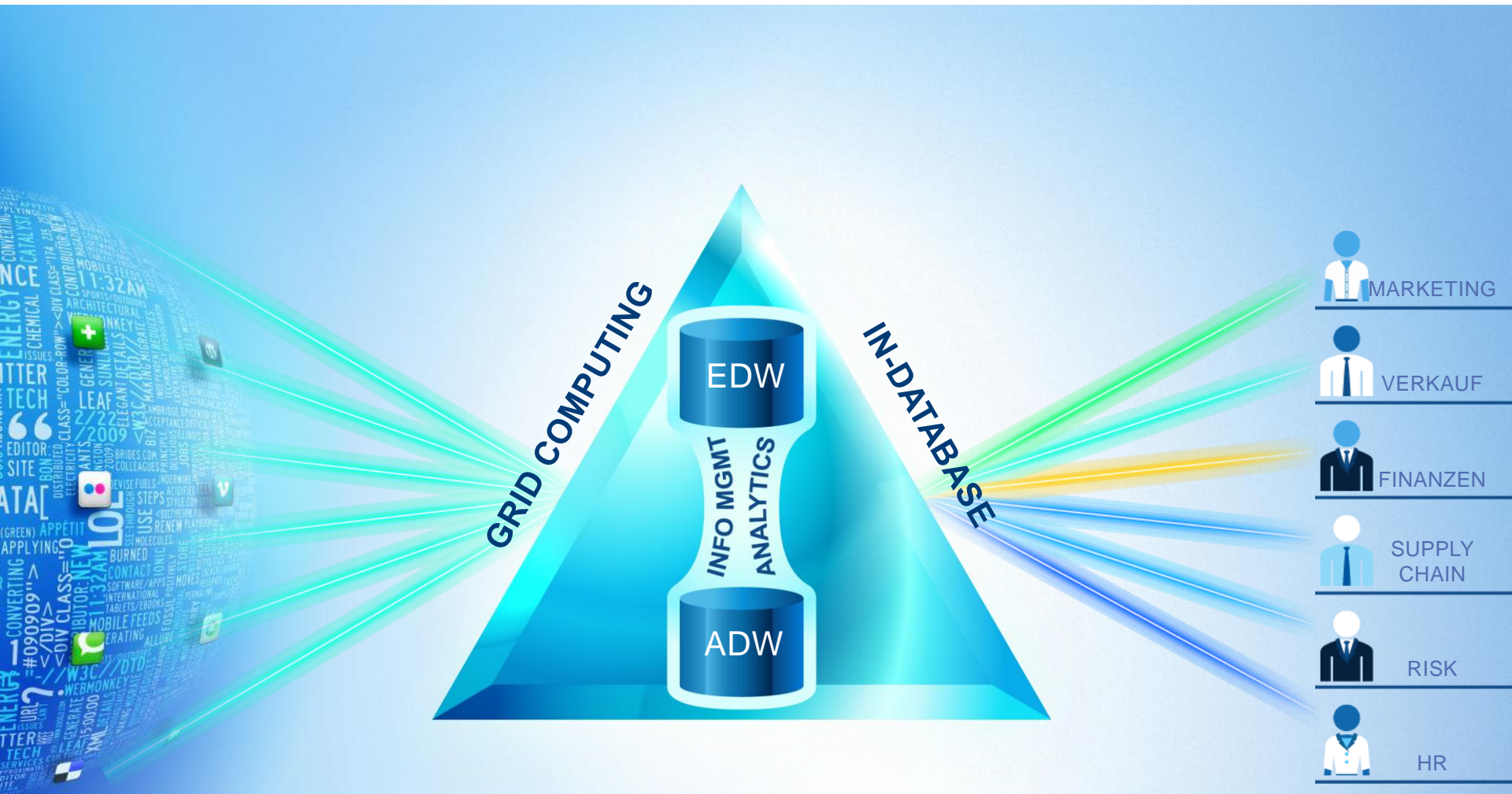
Statt die Daten zur «Arbeit»



Die «Arbeit» zu den Daten

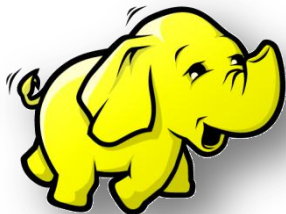


# Lösungsansatz für BIG Analytics



# Die Lösung für BIG Analytics

- Paralleles Datenbanksystem
  - EMC Greenplum
  - Teradata
- Apache Hadoop
  - Erprobt durch Google, Facebook, Yahoo
  - Skaliert praktisch beliebig!
  - Commodity Hardware (Blade Server)

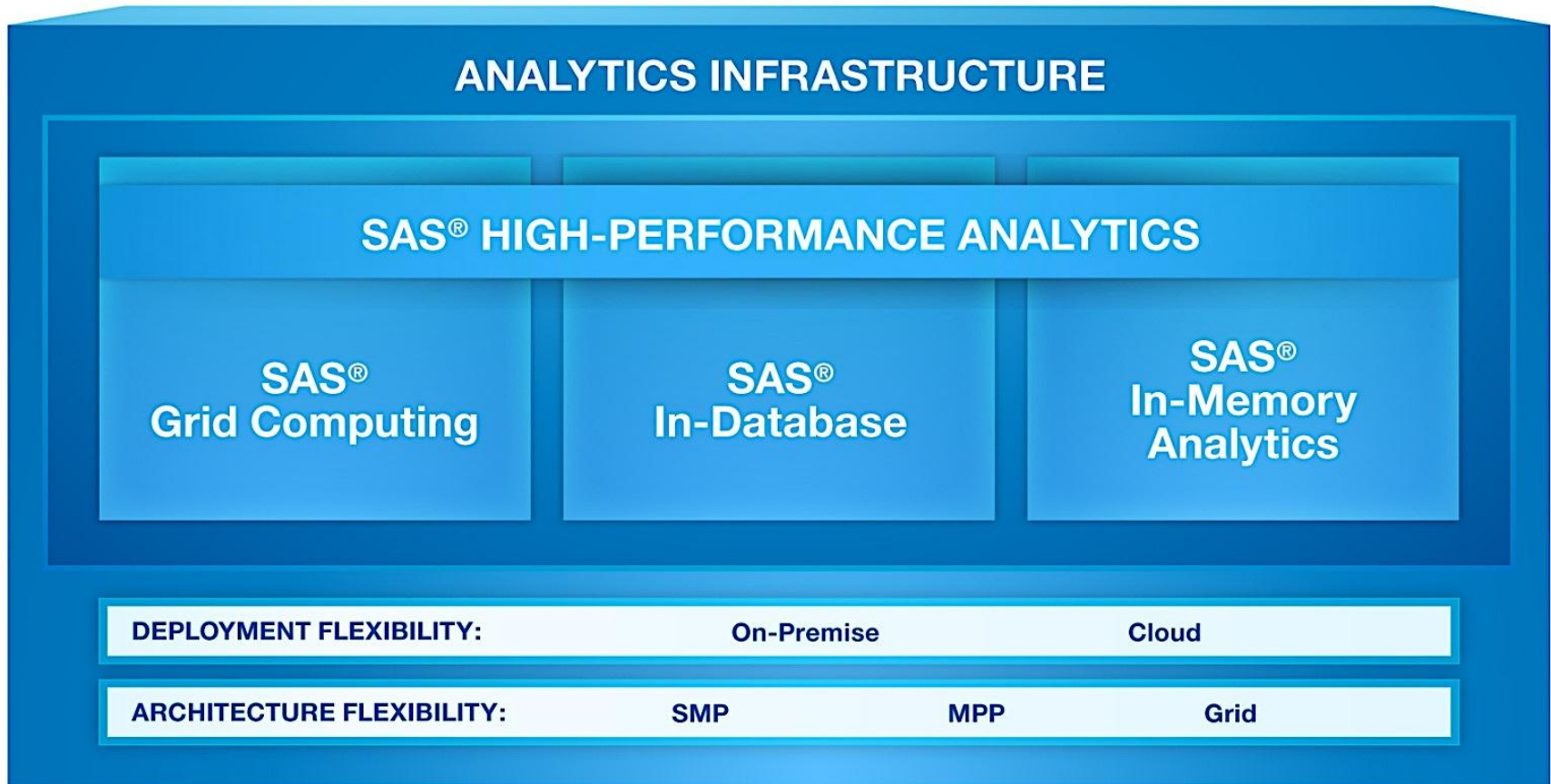


**hadoop**

# Commodity Hardware: Blade Server



# SAS Lösung für BIG Data



# Fazit

- BIG Data ist ein Hype!
- Das Buzzword wird verschwinden, aber die grossen Datenmengen werden bleiben
- Die technischen Voraussetzungen für BIG Analytics sind vorhanden
- BIG Analytics verschafft den entscheidenden Wettbewerbsvorteil ...schon heute!

[WWW.SAS.COM/BIGDATA](http://WWW.SAS.COM/BIGDATA)





**sas**.forum  
**SWITZERLAND 2012**

make connections • share ideas • be inspired

Vielen Dank für Ihre  
Aufmerksamkeit!



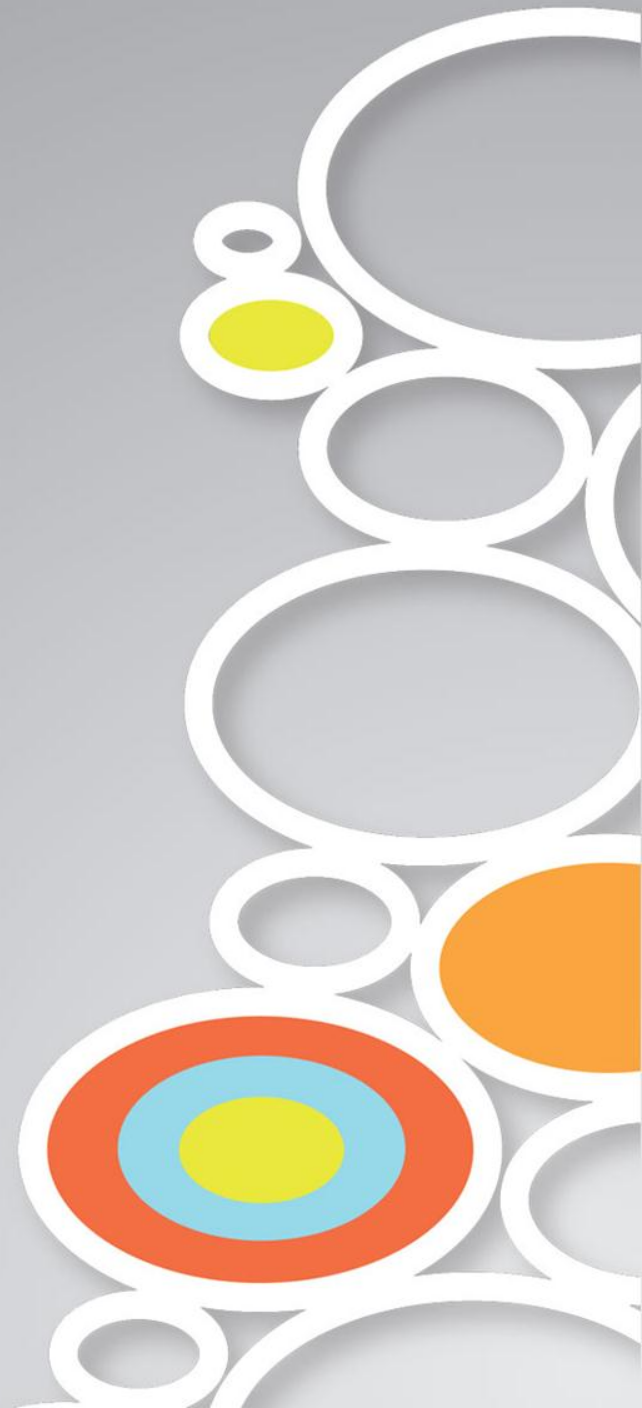
sas<sup>®</sup> forum

SWITZERLAND 2012

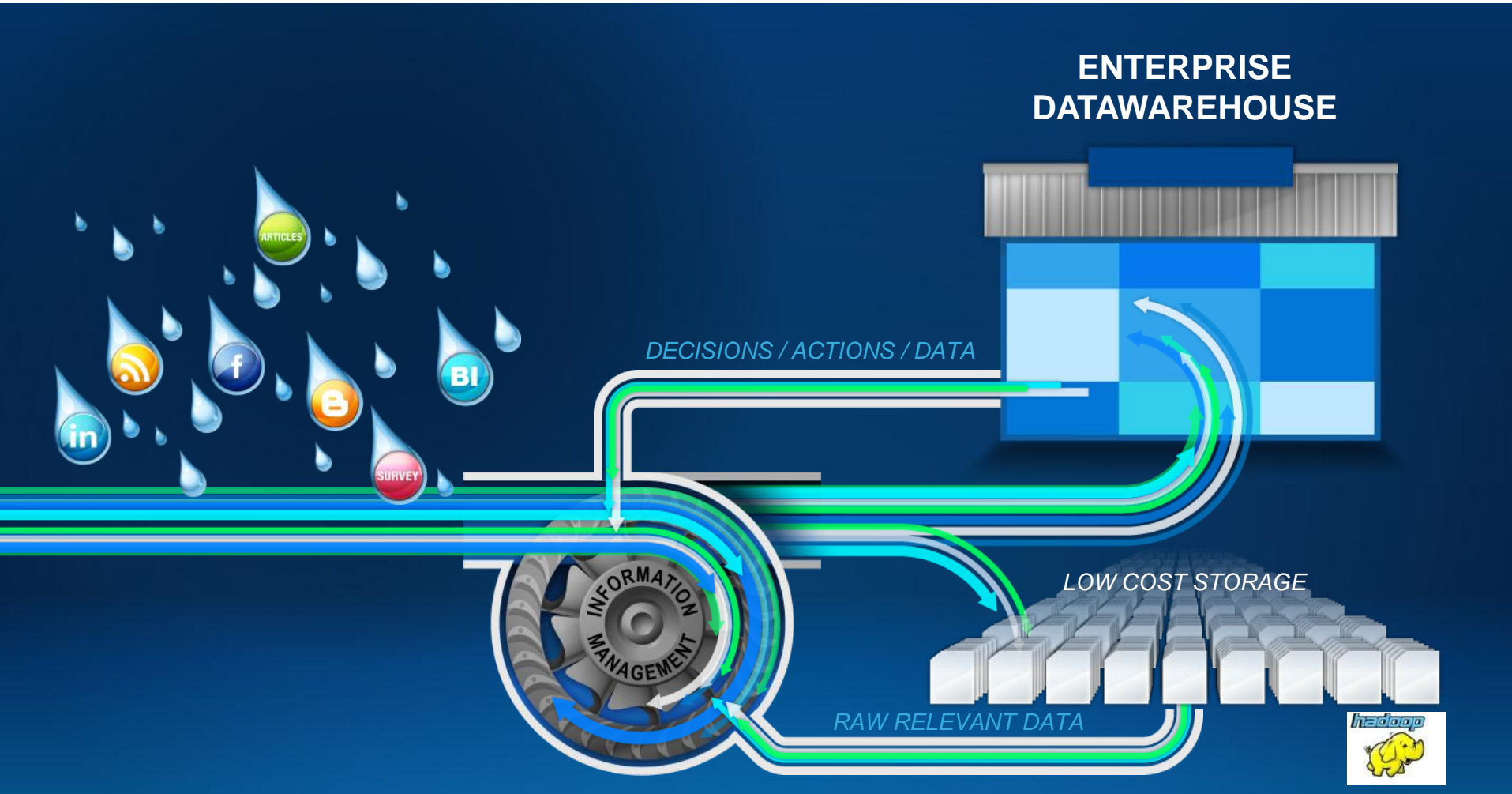
make connections • share ideas • be inspired

BACKUP SLIDES

---



# Welche Daten wohin?



# Business intelligence weiter denken

FORECASTING

TEXT ANALYTICS



DATA MINING

OPTIMIZATION

STATISTICS

Copyright © 2012, SAS Institute Inc. All rights reserved.

