

# Big Data für den Tourismus

Armin Hofer

18.01.2016



---

# Agenda

- Definitionsversuch „Big Data“
- Einflussfaktoren – die vier „V“s
- Anwendungsfälle
- Ausblick
- Diskussion

## Big Data – Definition

- „Big Data stellt **Konzepte, Technologien und Methoden** zur Verfügung, um die geradezu exponentiell steigenden Volumina **vielfältiger Informationen** noch besser als fundierte und zeitnahe **Entscheidungsgrundlage** verwenden zu können und so **Innovations- und Wettbewerbsfähigkeit** von Unternehmen weiter zu steigern.“ (Bitkom in BITKOM 2012)
- “Big Data” is a term encompassing the use **of techniques to capture, process, analyse and visualize** potentially **large datasets** in a reasonable timeframe not accessible to standard IT technologies. By extension, the platform, tools and software used for this purpose are collectively called “Big Data technologies.” (Networked European Software and Service Initiative [NESSI], 2012)
- **Extraktion von neuem Wissen** für die Unterstützung von Entscheidungen für unterschiedlichste Fragestellungen auf Basis des steigenden heterogenen **Datenvolumens.** (Köhler; Meir-Huber, #Big Data in #Austria; 2014)

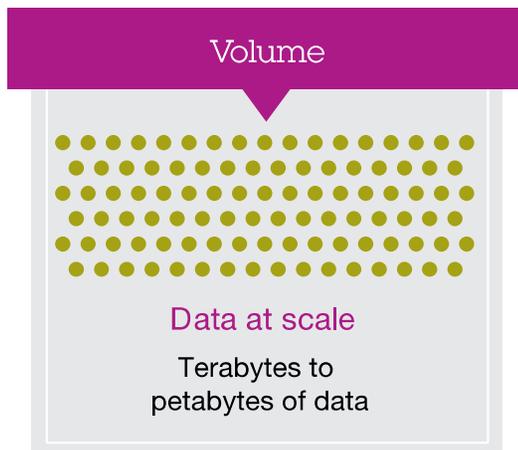
## Big Data – Definition

„**Big Data** [ 'big ' deitə] (von englisch big „groß“ und data „Daten“) bezeichnet Datenmengen, die zu groß oder zu komplex sind oder sich zu schnell ändern, um sie mit händischen und klassischen Methoden der Datenverarbeitung auszuwerten.

Der Begriff "Big Data" unterliegt als Schlagwort derzeit einem kontinuierlichen Wandel; so wird mit Big Data ergänzend auch oft der Komplex der Technologien beschrieben, die zum Sammeln und Auswerten dieser Datenmengen verwendet werden.“

[https://de.wikipedia.org/wiki/Big\\_Data](https://de.wikipedia.org/wiki/Big_Data) abgerufen am 15. Jänner 2016

## Die vier "V"



**Volume:** enormer Anstieg der vorhandenen Datenmenge in den letzten Jahren

## 40 ZETTABYTES

[ 43 TRILLION GIGABYTES ]

of data will be created by 2020, an increase of 300 times from 2005

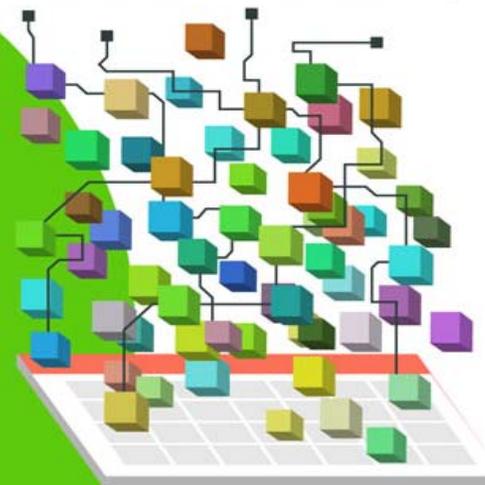


It's estimated that

## 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

## Die vier "V"

### Volume



#### Data at scale

Terabytes to  
petabytes of data

**Volume:** enormer Anstieg der vorhandenen Datenmenge in den letzten Jahren

**Variety:** verschiedenste Formate von Datenquellen müssen verarbeitet und flexibel in unterschiedlichste Formate integriert werden

### Variety



#### Data in many forms

Structured, unstructured,  
text, multimedia

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

**150 EXABYTES**

[ 161 BILLION GIGABYTES ]



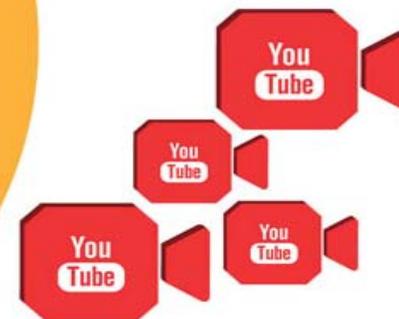
By 2014, it's anticipated there will be

**420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**



**4 BILLION+ HOURS OF VIDEO**

are watched on YouTube each month



# Variety

## DIFFERENT FORMS OF DATA

**30 BILLION PIECES OF CONTENT**

are shared on Facebook every month

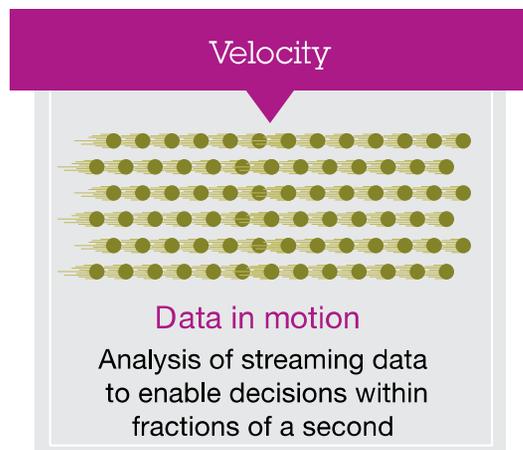


**400 MILLION TWEETS**

are sent per day by about 200 million monthly active users



## Die vier “V”



- **Velocity:** direkte bzw. zeitnahe Verarbeitung von Daten; auch Echtzeitanalysen; Analyse von immer mehr Daten in immer kürzerer Zeit

The New York Stock Exchange captures

**1 TB OF TRADE INFORMATION**

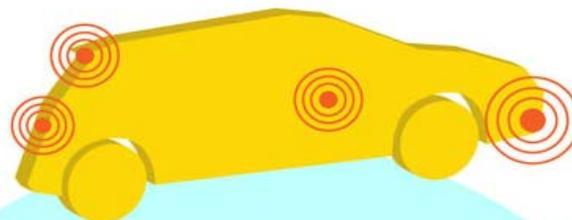
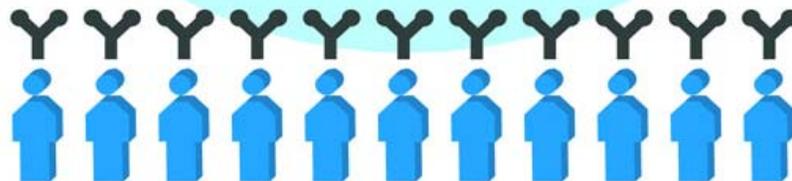
during each trading session



By 2016, it is projected there will be

**18.9 BILLION NETWORK CONNECTIONS**

– almost 2.5 connections per person on earth



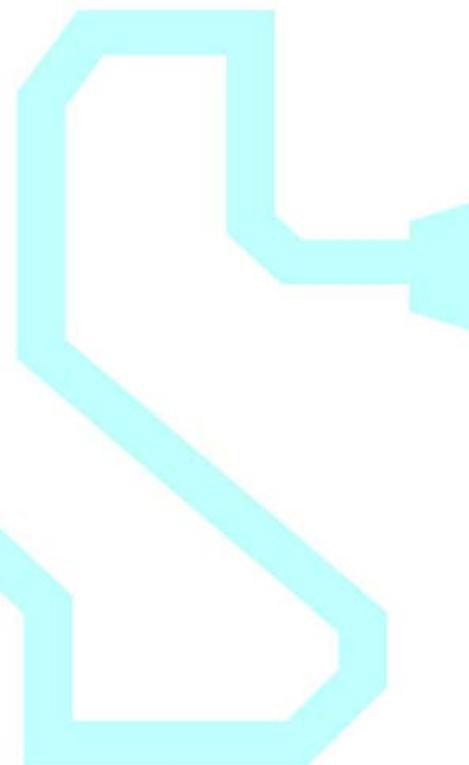
Modern cars have close to

**100 SENSORS**

that monitor items such as fuel level and tire pressure

# Velocity

## ANALYSIS OF STREAMING DATA



# Die vier "V"

Velocity

**Data in motion**  
Analysis of streaming data to enable decisions within fractions of a second

- **Velocity:** direkte bzw. zeitnahe Verarbeitung von Daten; auch Echtzeitanalysen; Analyse von immer mehr Daten in immer kürzerer Zeit

Veracity

**Data uncertainty**  
Managing the reliability and predictability of inherently imprecise data types

**Veracity:** Daten sind vielfach unpräzise und unvorhersehbar

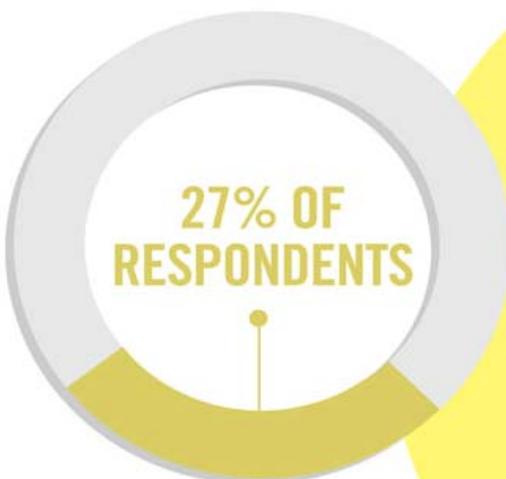
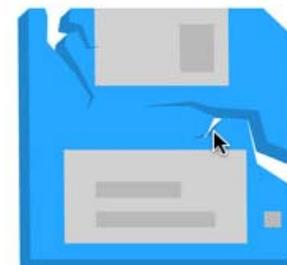
## 1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions



Poor data quality costs the US economy around

## \$3.1 TRILLION A YEAR



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

# Veracity

## UNCERTAINTY OF DATA

