



openHPI Big Data Einführung

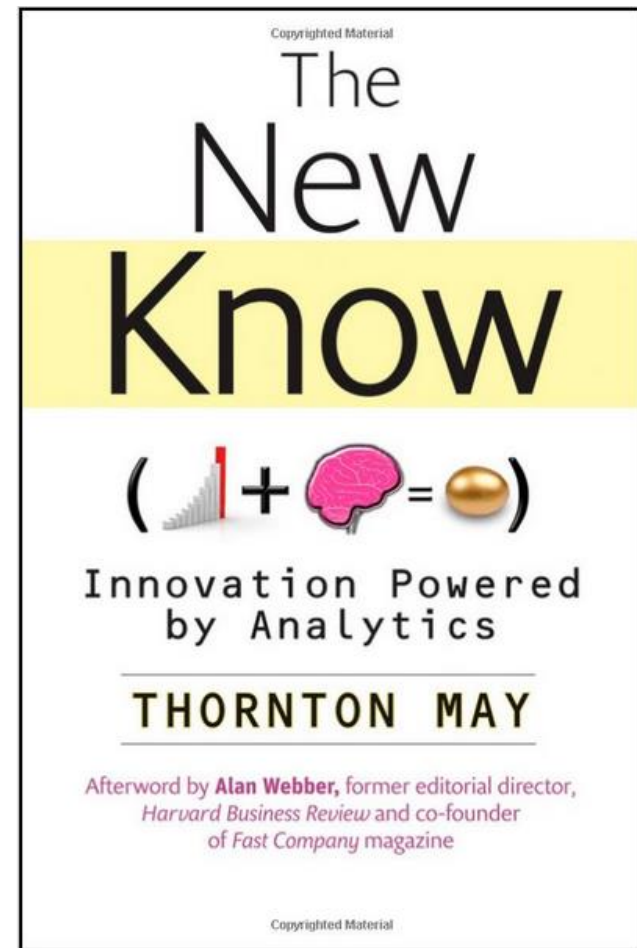
Wolfram Greis
European Mainframe Academy
Stein am Rhein (CH)

Warum sich um Big Data kümmern?

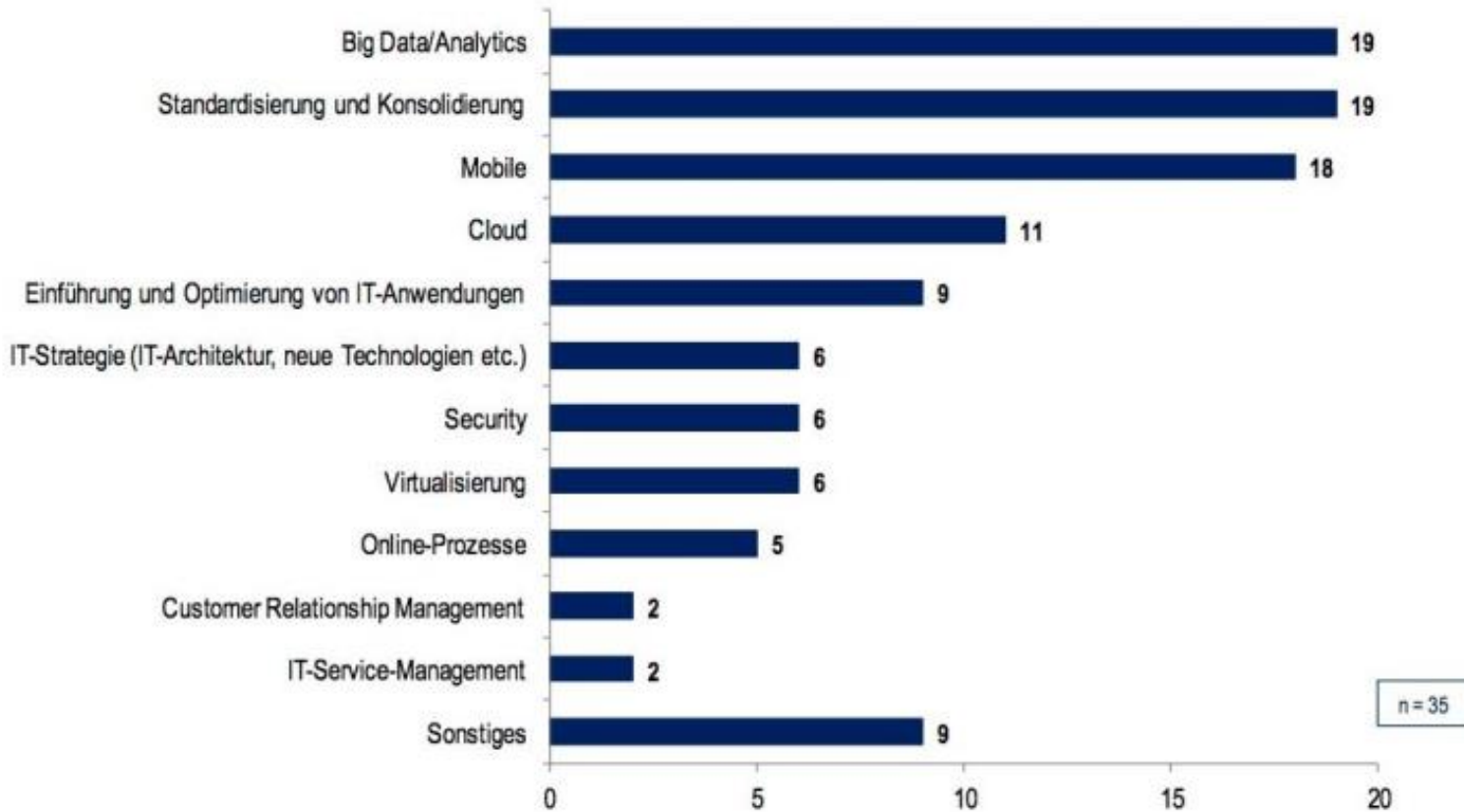
“The bosses discovered big data, and they expect you to do something about it.”



Thornton May, Futurist
IT Leadership Academy



WICHTIGSTE TECHNOLOGIETHEMEN DER IT-ANWENDERUNTERNEHMEN MEHRFACHNENNUNGEN MÖGLICH (MAXIMAL 3) – ABSOLUTE ZAHLEN



BizTech

ENTER

INDUSTRIES ▼ TOPICS ▼ TIPS & TACTICS FEATURES VOICES C-SUITE VIDEO

JAN 14 2015



The Mainframe Strikes Back: IBM's New z13 System Can Handle 2.5 Billion Transactions a Day

This is the new technology IBM promises will support the Big Data and mobile future our economy is headed toward.



Roger Kay
Contributor

[FOLLOW](#)

TECH 1/13/2015 @ 5:06PM | 3.893 views

With Z Systems Refresh, IBM Reminds Us That Mainframes Remain Relevant

[+ Comment Now](#) [+ Follow Comments](#)

[IBM](#) IBM +1.41% introduced its latest round of mainframes today. Who cares, you ask? Maybe you should. The mainframe has been declared dead a thousand times already, and yet it keeps on keeping on. While [Google](#) GOOGL +0.13% has proven a point about the power of scale-out computing — quantities of cheap servers internetworked to share tasks — there are some types of computing that just plain work better on Big Iron. Mainframes are good at tasks like transaction processing, which has performance needs well beyond what scale-out computing can deliver. That's why the

- 1960's
 - Hierarchische Datenbank IMS → Hierarchisches Datenmodell
- 1970's
 - Relationale DBs → Daten in Tabellen
 - Daten werden normalisiert
- 1980's
 - RDBMSs werden kommerzialisiert
 - SQL wird Standard
 - Datenbank Transaktionen mit **ACID** Eigenschaften

Atomicity

Consistency

Isolation

Duration

■ 1990's

- 3-Tier Client/Server
- Verteilte Datenbanken
 - Two Phase Commit (2PC)
- Internet
- Web 2.0
- Social Media

■ 2000's

- Big Data
- NoSQL (Keine Transaktionsunterstützung und **BASE** Eigenschaften)

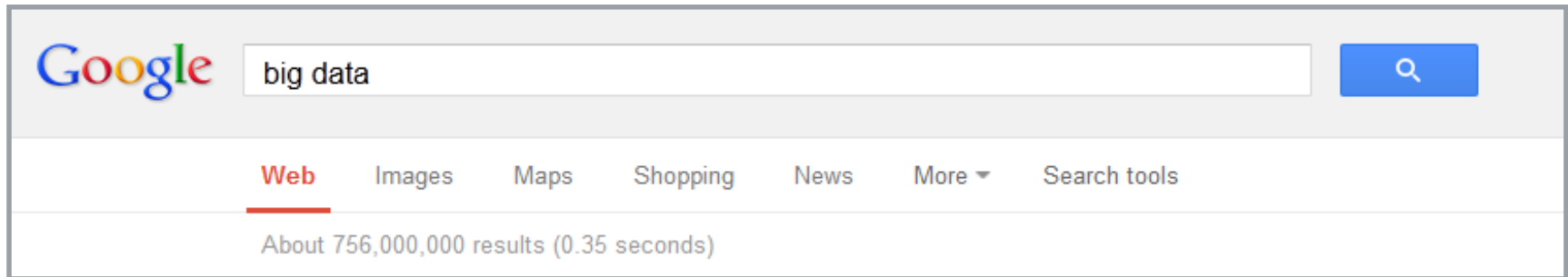
Basic Availability

Soft State

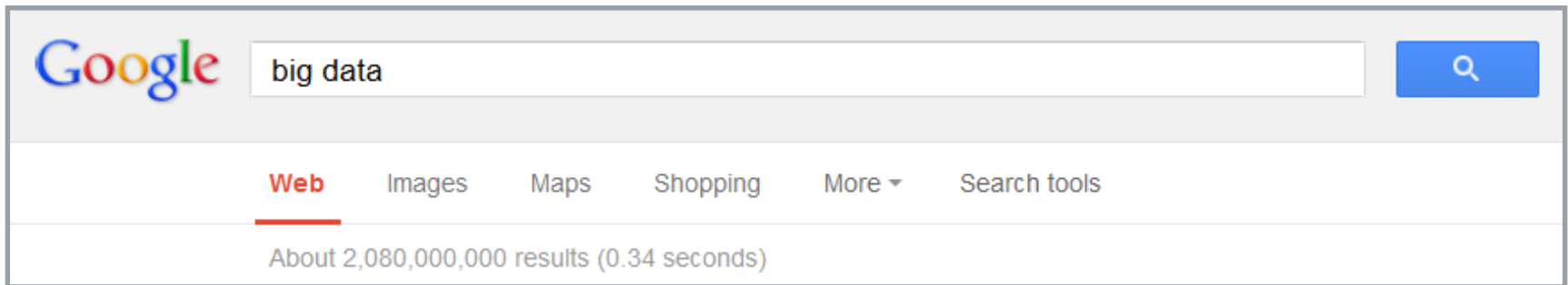
Eventual Consistent

- "Big Data"?
Heisst das, wir hatten "Small Data" bisher?
- Was heisst "Big Data" wirklich?
 - **Es geht nicht nur um die Menge!**
 - **Herausforderung sind die unstrukturierten Daten**
- Es gibt eine Menge neuer Technologien, die unter dem Begriff NoSQL zusammengefasst werden
- Google, Yahoo, Amazon, Twitter, eBay, LinkedIn ... sind Pioniere
 - **MapReduce Framework**
 - **Key-Value Stores**
- Tencent, Baidu (Chinesische Telekom Giganten)
- Open Source Community spielt eine grosse Rolle
 - **Hadoop, HBase, MongoDB, Cassandra, RabbitMQ ...**





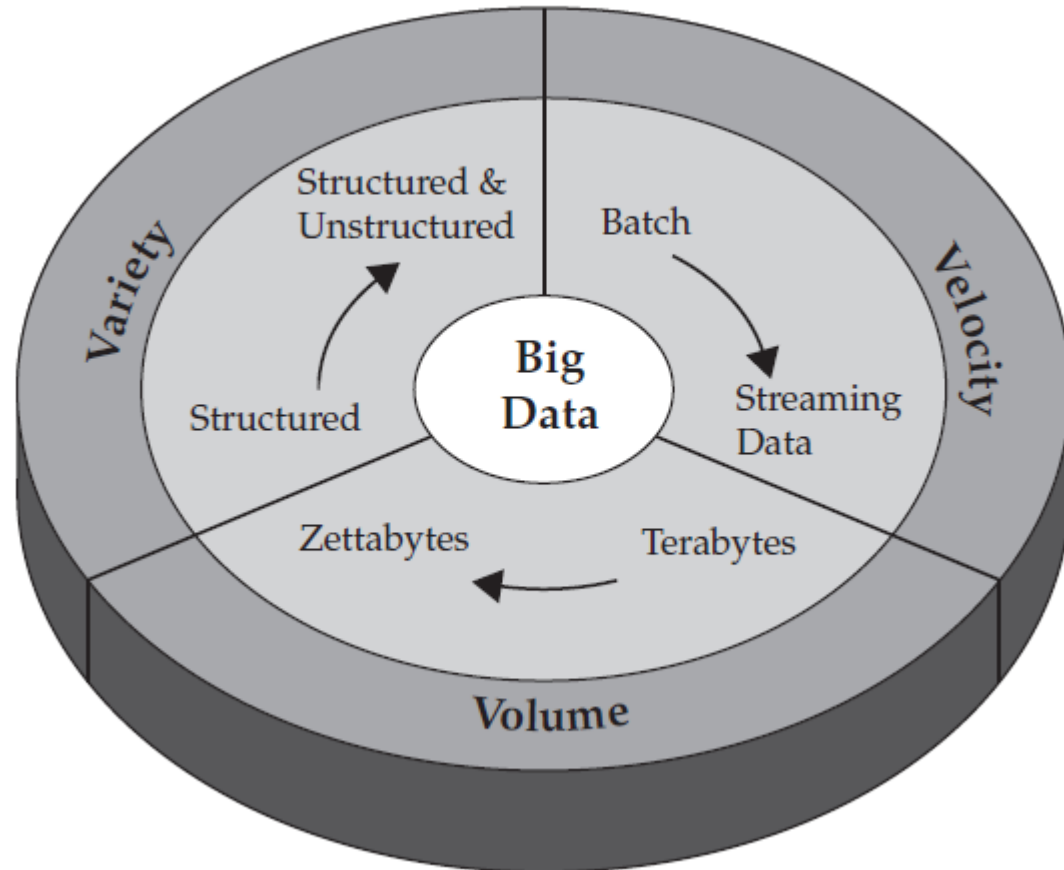
- Suche nach „Big Data“ in Google am 18.2.2013
- 756 millionen Hits in 0,35 Sekunden



- Gleiche Suche am 11.6.2013
- Mehr als 2 Milliarden Hits in 0,34 Sekunden
- Das ist Big Data! ... und die Antwortzeiten!

Big Data is Characterized by 3 Vs (aka V³)

- Source:
Understanding Big Data
IBM



The three defining characteristics of Big Data

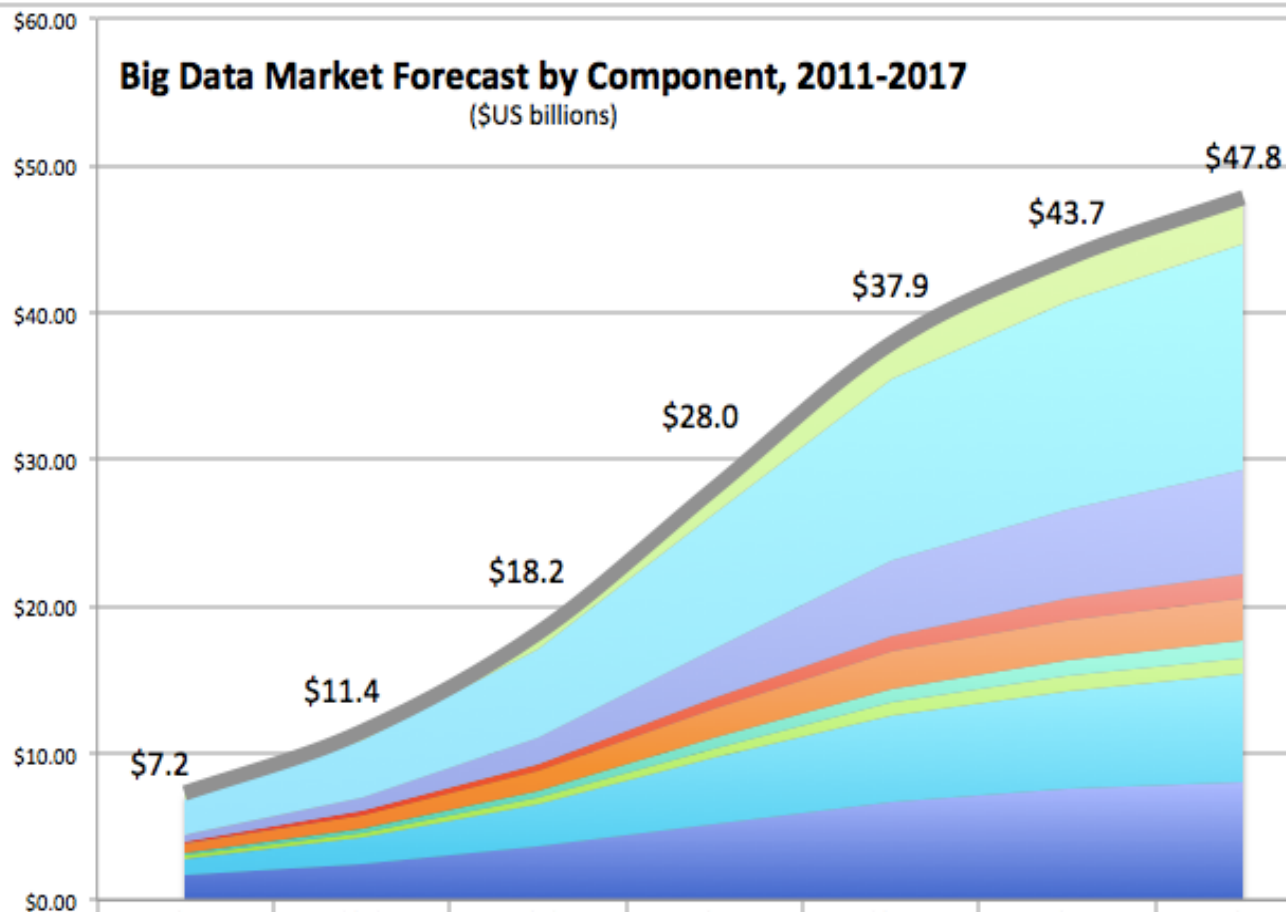
- Big Data ändert die Art der Datenanalyse
- Mainframes analysierten (bisher) strukturierte Daten mit Werkzeugen von
 - SAS
 - Cognos
 - Hyperion
 - etc.
- Mit Big Data wird die Analyse unstrukturierter Daten vorgenommen.
Dafür gibt es neue Werkzeuge:
 - MapReduce
 - Hadoop
 - NoSQL DBs
 - Hive
 - etc.



Yearly Revenue (\$US billions)

Big Data Market Forecast by Component, 2011-2017

(\$US billions)



	2011	2012	2013	2014	2015	2016	2017
Big Data XaaS Revenue	\$0.35	\$0.61	\$1.05	\$1.74	\$2.47	\$2.91	\$3.24
Big Data Professional Services Revenue	\$2.45	\$3.87	\$6.10	\$9.29	\$12.37	\$14.14	\$15.38
Big Data Application (Analytic and Transactional) Software	\$0.49	\$0.94	\$1.80	\$3.29	\$5.02	\$6.15	\$7.00
Big Data NoSQL Database Software	\$0.10	\$0.19	\$0.39	\$0.73	\$1.14	\$1.41	\$1.62
Big Data SQL Database Software	\$0.72	\$1.02	\$1.45	\$1.99	\$2.47	\$2.73	\$2.90
Big Data Infrastructure Software	\$0.16	\$0.26	\$0.43	\$0.70	\$0.96	\$1.12	\$1.24
Big Data Networking Revenue	\$0.18	\$0.28	\$0.44	\$0.67	\$0.89	\$1.02	\$1.11
Big Data Storage Revenue	\$1.16	\$1.83	\$2.89	\$4.40	\$5.86	\$6.70	\$7.28
Big Data Compute Revenue	\$1.64	\$2.45	\$3.64	\$5.23	\$6.70	\$7.50	\$8.06
Total Big Data Revenue	\$7.2	\$11.4	\$18.2	\$28.0	\$37.9	\$43.7	\$47.8



- In einem Report von 2001, definierte der Analyst Doug Laney (damals META Group, heute Gartner) die berühmten 3 Vs
- „Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization.“
Source: Gartner
- „In short, the term Big Data applies to information that can't be processed or analyzed using traditional processes or tools.“
Source: Understanding Big Data, McGraw-Hill, 2012

Source



Your Ultimate Guide to the
Non - Relational Universe!

[the **best selected** nosql link [Archive](#) in the
web]
...never miss a *conceptual* article again...
News Feed covering all changes [here](#) !

NoSQL DEFINITION:Next Generation Databases mostly addressing some of the points: being **non-relational, distributed, open-source** and **horizontally scalable**.

The original intention has been **modern web-scale databases**. The movement began early 2009 and is growing rapidly. Often more characteristics apply such as: **schema-free, easy replication support, simple API, eventually consistent** / BASE (not ACID), a **huge amount of data** and more. So the misleading term "*nosql*" (the community now translates it mostly with "**not only sql**") should be seen as an alias to something like the definition above. [based on 7 sources, 14 constructive feedback emails (thanks!) and 1 disliking comment . Agree / Disagree? [Tell](#) me so! By the way: this is a strong definition and it is out there here since 2009!]

LIST OF NOSQL DATABASES [currently 150]

Core NoSQL Systems: [Mostly originated out of a Web 2.0 need]

Wide Column Store / Column Families

[Hadoop / HBase](#) API: **Java** / **any writer**, Protocol: **any write call**, Query Method: **MapReduce Java** / **any exec**, Replication: **HDFS Replication**, Written in: **Java**, Concurrency: **?**, Misc: **Links**: 3 Books [\[1, 2, 3\]](#)

[Cassandra](#) massively scalable, partitioned row store, masterless architecture, linear scale performance, no single points of failure, read/write support across multiple data centers & cloud availability zones. API / Query Method: **CQL and Thrift**, replication: **peer-to-peer**, written in: **Java**, Concurrency: **tunable consistency**, Misc: built-in data compression, MapReduce support, primary/secondary indexes, security features. Links: [Documentation](#), [PlanetC*](#), [Company](#).

[Hypertable](#) API: **Thrift** (Java, PHP, Perl, Python, Ruby, etc.), Protocol: **Thrift**, Query Method: **HQL, native Thrift API**, Replication: **HDFS Replication**, Concurrency: **MVCC**, Consistency Model: **Fully consistent** Misc: High performance C++ implementation of Google's Bigtable. » [Commercial support](#)

[Accumulo](#) Accumulo is based on **BigTable** and is built on top of [Hadoop](#), [Zookeeper](#), and [Thrift](#). It

NoSQL EVENTS:

- 13th June **NoSQL Roadshow** Copenhagen »
- 30. Nov **NoSQL Matters Barcelona** »
Call for papers »

All past NoSQL Conferences »
register your event here! »

NoSQL Search



Only one sponsor place left!

NoSQL ARCHIVE



NoSQL NEWS

- Will be updated soon!

NoSQL FORUMS [\[3\]](#)



IMPRINT

<http://nosql-databases.org>
and
<http://nosql-database.org>
is owned by:



Prof. Dr. Stefan Edlich
Stubenrauchstr.9c
14167 Berlin
Germany
edlich@gmail.com
<http://edlich.de>
Skype: sedlich

- 1979: Ken Thompson entwickelt key/hash Datenbank
- 80`s: Lotus Notes, BerkeleyDB, GT.M ...
- 1998: NoSQL wird erstmals verwendet von Carlo Strozzi
- Ab 2000: Web 2.0 und grosse Datenmengen → Big Data
- Google, Yahoo, und Amazon entwickeln Big Data Systeme
- NoSQL-Systeme wie HBase/Hypertable, CouchDB, Cassandra, Voldemort, Dynamo/Dynomite, MongoDB, Redis, Riak u.v.m. kommen auf den Markt

RDBMS	NoSQL
Data fits into tables	Data use whatever structure (trees, key/value, graphs ...)
Scale up	Scale out
Declarative queries (SQL)	Functional programming and MapReduce algorithms
Schema-on-Write	Schema-on-Read (Hadoop)
ACID	BASE

Fragen / Diskussionen
gerne im Forum zu diesem MOOC