



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Big Data

Techniken, Analysemethoden und Visualisierungs-Tools

Nicola Heintze

Matrikel-Nr.: 2180929

Nicola Heintze

Big Data – Techniken, Analysemethoden und
Visualisierungs-Tools

Ausarbeitung eingereicht im Rahmen der Ringvorlesung „Content & Technology“

im Studiengang Next Media (M.A.)
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer : Prof. Kai von Luck

Abgegeben am 20.02.2014

Nicola Heintze

Big Data - Techniken, Analysemethoden und Visualisierungs-Tools

Stichworte

Big Data, Automatische Datenerfassung, Predictive Analysis, Große Datenmengen, Big Data Visualization, In-Memory, Hadoop, Tableau, Spotfire.

Kurzzusammenfassung

In dieser Ausarbeitung werden aktuelle Methoden und Techniken zur Erfassung, Speicherung und Analysierung von Big Data im Unternehmenskontext untersucht und gängige Visualisierungstools vorgestellt.

Inhaltsverzeichnis

1	Einleitung	1
2	Definition Big Data	2
3	Der Trend Big Data	3
3.1	Aktuelle Entwicklungen.....	4
4	Technologien zur Datenverarbeitung	5
4.1	In-Memory Lösungen	5
4.2	Hadoop.....	5
5	Analysemethoden	6
5.1.1	Kopplungsanalyse	6
5.1.2	Predictive Analytics	7
5.1.3	Data Mining	7
6	Visualisierungstools	8
6.1	Tableau Software	8
6.2	TIBCO Spotfire	9
7	Zusammenfassung / Ausblick	10
8	Quellenverzeichnis	12
9	Versicherung über Selbstständigkeit	14

1 Einleitung

Big Data ist momentan das Topthema der IT Branche. „Bei Big Data handelt es sich um eine Revolution“, sagt der Sprecher des Fraunhofer Instituts für Intelligente Analyse- und Informationssysteme Professor Stefan Wrobel (Buttong 2013). Und damit könnte er Recht haben, denn laut IBM lassen sich mit Big Data gerade für Unternehmen komplexe Entscheidungen ad-hoc und auf Basis zuverlässiger Prognosen treffen – so werden Geschäftsmodelle und -prozesse modernisiert und ein spürbarer Mehrwert für ihre Anleger, Kunden, Partner, Mitarbeiter und die Gesellschaft geschaffen (vgl. CeBIT 2014).

Unternehmen sammeln seit Jahren unzählige Daten unterschiedlicher Ausprägung aus verschiedenen Quellen – nun sollen diese sinnvoll analysiert, ausgewertet und visualisiert werden, um Kosten einzusparen und die Umsätze zu steigern (vgl. Buttong 2013).

Doch auch in unserem Alltag werden fast durchgängig Daten gesammelt – ohne dass wir es überhaupt mitbekommen. Durch smarte Sensortechniken unserer Smartphones werden Daten erfasst und gespeichert.

Big Data befasst sich ausschließlich damit, jegliche Daten unterschiedlicher Ausprägung und Format aus verschiedenen Quellen sinnvoll zu verknüpfen und in Zusammenhang mit anderen Datenquellen auszuwerten. Unternehmen arbeiten mit Hochdruck daran ihre Systeme für eine schnellere und effizientere Datenverarbeitung zu rüsten, um Daten bestenfalls in Echtzeit verarbeiten zu können (vgl. ebd.).

In dieser Ausarbeitung soll zunächst der Begriff Big Data im Zusammenhang mit der Nutzung in Unternehmen näher beleuchtet werden.

Außerdem werden aktuelle Methoden und Techniken zur Erfassung, Speicherung und Analysierung von großen Datenmengen untersucht, sowie gängige Visualisierungstools vorgestellt.

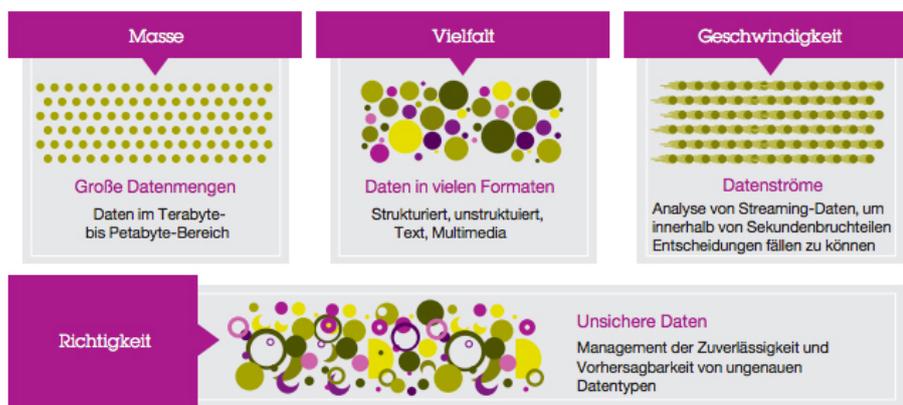
Ziel der vorliegenden Arbeit ist es, sowohl einen Überblick über die technischen Voraussetzungen eines Systems zur Datenverarbeitung als auch deren Analysemethoden zu geben. Außerdem sollen die Vorteile einer Visualisierung komplexer Datenstrukturen aufgeführt werden.

2 Definition Big Data

Big Data bezeichnet die Erfassung, Speicherung und Analyse großer Datenmengen in Form unterschiedlich strukturierter Daten auf Basis spezifischer Methoden und Technologien (vgl. Horvath 2013, S. 1). Welcher Art die Struktur der Daten sein können, beschreibt diese Definition des Intel IT Centers: „Big data is characterized by huge data sets and varied data types, both semistructured and unstructured (videos, images, audio, clickstreams, weblogs, text, and e-mail)“. Weit verbreitet ist auch die Charakterisierung von Big Data durch die sogenannte V3-Formel, welche die Komponenten von Big Data in folgende Segmente gliedert: Volume, Variety und Velocity (zu deutsch Masse, Vielfalt und Geschwindigkeit) (vgl. Schroeck/Shockley/Smart/Romero-Morales/Tufano 2012, S. 4).

Der Begriff „Masse“ in Bezug auf Big Data geht ganz klar auf die Menge der zu verarbeitenden Daten zurück, die Vielfalt bezeichnet in diesem Zusammenhang die unterschiedliche Struktur vorhandener Daten und die dritte Komponente der V3-Formel steht für die Tatsache, dass die Geschwindigkeit der Daten, die erzeugt, verarbeitet und analysiert werden ständig zunimmt (vgl. ebd.). Die folgende Grafik stellt diese 3 Charakteristika von Big Data nebeneinander dar. Hier wird die 3V-Formel noch durch den Aspekt der Richtigkeit erweitert, der die Zuverlässigkeit von Informationen, die mit bestimmten Datentypen verbunden wird, bezeichnet (vgl. ebd.).

Abb. 1: Die 4 Dimensionen von Big Data



Quelle: IBM Global Business Services

3 Der Trend Big Data

Im Zusammenhang mit Big Data wird oft auch der Begriff Datability genutzt. Dieser Begriff ist eine Kombination aus Big Data und den Möglichkeiten seiner nachhaltigen und verantwortungsvollen Nutzung (Englisch: ability, sustainability, responsibility) (vgl. CeBIT 2014).

Als Top Thema auf der CeBIT 2014 wirbt der Messeveranstalter mit „Datability – Wie Big Data das Leben vereinfacht“. So gelingt es beispielsweise den Unternehmen Wettbewerbsvorteile zu erzielen, deren Segmentierung und Datenverarbeitung Irrelevantes von Relevantem trennen und letzteres nutzbar machen kann (vgl. Buttong 2013).

Die ganze Welt spricht von dem Mega-Trend rund um die Erfassung, Speicherung und Verarbeitung von Datensätzen und wie diese sinnvoll miteinander verknüpft werden können, um beispielsweise Aussagen über zukünftige Verhaltensweisen von Konsumenten treffen zu können.

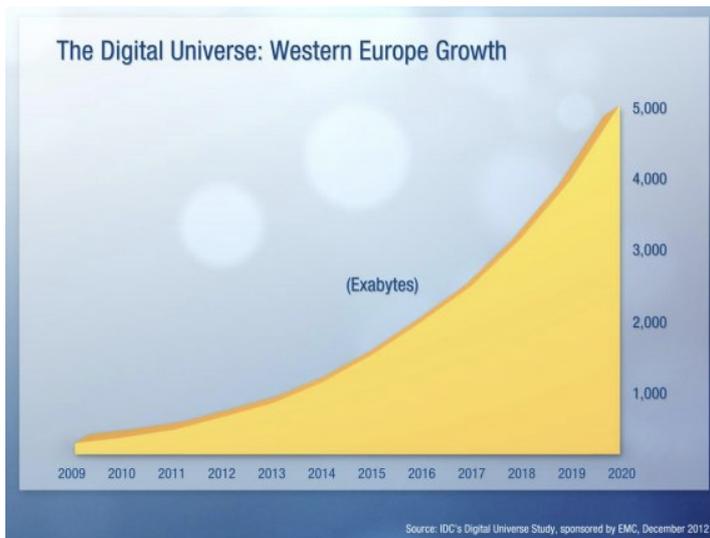
Die große Herausforderung dabei ist es, unterschiedliche Systeme in ein für das Unternehmen einheitliches System zu migrieren, um aus der Vielzahl an großen Datenmengen einen Nutzen ziehen zu können (vgl. Buttong 2013). Vor allem für Unternehmen sollen Big Data-Anwendungen immer mehr an Bedeutung gewinnen, weil durch das Anreichern des aktuellen Datenbestandes mit zusätzlichen Informationen, sowie deren Auswertung laut einer Studie des Fraunhofer Instituts unternehmerische Ziele, wie Steigerung der Umsätze und Einsparung von Kosten ermöglicht würden (vgl. ebd.).

IBM sieht in Big Data den größten Wachstumsbereich der nächsten 10 Jahre und auch IDC schätzt, dass der Markt für Big-Data-Lösungen in den nächsten fünf Jahren im Durchschnitt um 32 Prozent wachsen wird (vgl. FOCUS 2014). Man kann schon fast sagen, dass ein Umbruch in der IT Branche stattgefunden hat. Während noch vor ein paar Jahren vor allem die Hardware im Mittelpunkt aller Entwicklungen stand, ist nun seit langem die Software der dominierende Faktor - in der die Daten selbst zum entscheidenden Erfolgsfaktor werden (vgl. BITKOM 2012).

3.1 Aktuelle Entwicklungen

„Aktuelle Prognosen besagen, dass sich das weltweite Datenvolumen jährlich verdoppeln werde“ (Buttong 2013).

Abb. 2: Wachstum des weltweiten Datenvolumens bis 2020



Quelle: International Data Corporation (IDC) (2013), <http://tinyurl.com/oqta3c4>, Stand: Februar 2013

Die nachfolgende Grafik zeigt die Reduzierung der Dauer der Datenverarbeitung für 5 Exabyte (das entspricht etwa 5 Mrd. Gigabyte) Datenvolumen. Während im Jahr 2003 die Verarbeitung dieser Datenmenge noch mehr als 1000 Jahre gedauert hätte, waren es im Jahr 2011 nur noch 2 Tage und im vergangenen Jahr nur noch unschlagbare 10 Minuten.

Abb. 3: Entwicklung der Dauer der Datenverarbeitung



Quelle: CeBIT 2014, <http://www.cebit.de/de/news-trends/trends/>, Stand: 08.02.2014

Daraus lässt sich ablesen, dass die Daten auf Systemebene in Zukunft immer schneller verarbeitet werden müssen, wenn gleichzeitig die Menge der produzierten Daten stetig ansteigt.

4 Technologien zur Datenverarbeitung

Wachsende Datenmengen, die immer schneller verarbeitet werden müssen - um im besten Fall Echtzeitinformationen bereitzustellen - stellen auch höhere Anforderungen an Soft- und Hardwaresysteme. Im Folgenden werden aktuell verbreitete Technologien vorgestellt, die große Datenmengen in relativ kurzer Zeit verarbeiten können.

4.1 In-Memory Lösungen

In-Memory Systeme zeichnen sich dadurch aus, dass die zu verarbeitenden Daten sich im Arbeitsspeicher des Servers befinden, während bei anderen Systemen zeitaufwändig über den Festplattenspeicher auf die zu verarbeitenden Daten zugegriffen werden muss (vgl. T-Systems International GmbH 2012). „Im Unterschied zu herkömmlichen Datenbanken müssen In-Memory-Datenbanken die zu verarbeitenden Daten nicht bei jeder Anfrage von einer Festplatte abrufen und in den Arbeitsspeicher laden, da die Daten sich schon dort befinden“ (Heise 2014).

4.2 Hadoop

Hadoop bezeichnet eine Technik, bei der die Datenmengen und ihre Bearbeitung auf mehrere Rechner gleichzeitig aufgeteilt werden – dieses parallele Ablaufen von Prozessen auf mehreren Rechnern ist auch unter dem Begriff Map Reduce bekannt (vgl. Buttong 2013).

Neben der Verteilung von Aufgaben fangen dabei intelligente Algorithmen für den Nutzer unbemerkt Server-Ausfälle ab und sorgen dafür, dass das System immer verfügbar ist (vgl. Buttong 2013). „Die Software-Bibliothek Apache™ Hadoop® ist ein Framework, das für die verteilte Verarbeitung großer Datenmengen über Cluster von Computern mit einfachen Programmiermodellen ermöglicht“ (The Apache Software Foundation 2014).

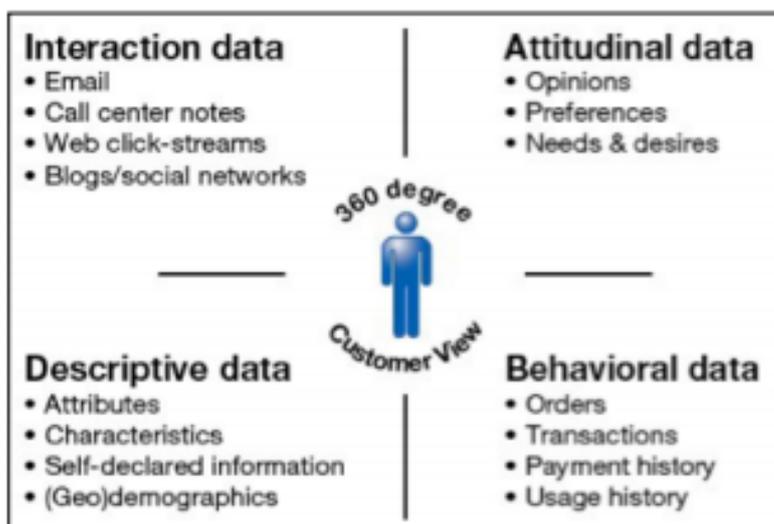
5 Analysemethoden

Wo viele Daten gesammelt werden wird es umso wichtiger, diese Daten sinnvoll zu strukturieren und auszuwerten. Im Folgenden werden Analyse- und Visualisierungswerkzeuge vorgestellt, die den Herausforderungen zunehmend komplexerer Datensätze und immer größer werdender Datenmengen gewachsen sind (vgl. Parthier 2013). Diese Werkzeuge werden in den folgenden Beispielen vornehmlich zur Analyse von Unternehmensdaten verwendet, können aber genauso gut generell für alle möglichen Datenbestände genutzt werden.

5.1.1 Kopplungsanalyse

Bei Kopplungsanalysen wird untersucht, wie viele separate Datenelemente sich miteinander kombinieren lassen, um so ein Maximum an Erkenntnisgewinn aus den analysierten Daten ziehen zu können (vgl. ebd.). Laut IT Research werden die Datenmengen künftig in einem Lifecycle gesammelt, verwaltet, verarbeitet und analysiert. Die folgende Grafik aus einer Studie des IT Verlags für Informationstechnik zeigt den Lifecycle gesammelter Daten, bei dem die Daten aus der Sicht des Kunden in 4 Haupttrubriken eingegliedert werden können.

Abb. 4: Lifecycle Datenmengen



Quelle: Parthier 2013, S. 3

Diese Methode wird auch als Kopplungsanalyse bezeichnet, bei der aussagekräftige Beziehungen zwischen den kategorisierten Datenbeständen ermittelt werden sollen (vgl. Parthier 2013).

5.1.2 Predictive Analytics

Predictive Analytics ist eine Sammlung von Technologien, die es Unternehmen ermöglichen, sowohl gespeicherte als auch in Echtzeit verarbeitete Daten zu verwenden, um Aussagen über zukünftige Handlungsoptionen machen zu können (vgl. Intel IT Center 2013).

Während bei traditionellen Business Intelligence Methoden vornehmlich mit bereits strukturierten Datensätzen, die sich in existierende Muster einordnen lassen, gearbeitet wird, soll bei der Anwendung von Big Data vor allem auf unstrukturierte Rohdaten zurückgegriffen werden. Bei Predictive Analytics sollen über weiterführende Informationen eines bestehenden Datensatzes bestimmte Muster und die Beziehungen der Daten zueinander erfasst werden. Dabei ist es das Ziel, vor allem in Echtzeit verarbeitete Datensätze für zukünftige Analysen zu verwerten – je größer die Datenmenge hierbei ist, desto genauer können Zusammenhänge zwischen den aufgezeichneten Daten aufgezeigt werden (vgl. ebd.).

5.1.3 Data Mining

„Data Mining erlaubt tief gehende Analysen von großen Mengen verschiedenartiger Daten. Dank Automatisierung kann auf diese Weise der gesamte Datenbestand kategorisiert werden“ (Parthier 2013). Bei Data Mining können verschieden strukturierte Daten miteinander verglichen werden und wiederkehrende Ereignisse ausgemacht werden (vgl. ebd.). Über Data Mining können also selbständig neue Muster und Zusammenhänge in vorgegebenen Daten erfasst und ausgewertet werden (vgl. Bubeck/Kleine Büning 2014).

6 Visualisierungstools

Die Analysemethoden Kopplungsanalyse, Predictive Analysis und Data Mining haben verschiedene Techniken gezeigt, wie große Datenmengen unstrukturierter Art analysiert, in Relation zueinander gesetzt und in eine einheitliche, interpretierbare Form gebracht werden können. Damit diese Daten für jedermann verständlich dargestellt werden können, bedarf es spezifischer Visualisierungstools, um die verarbeiteten Daten visuell darzustellen. Im Folgenden werden die beiden gängigen Visualisierungsprogramme Tableau und TIBO Spotfire vorgestellt.

6.1 Tableau Software

Tableau ist eine Anwendung zur Visualisierung und Verarbeitung von Daten jeglicher Art. Die richtige Visualisierung der Daten ist, nach eigenen Angaben des Softwareunternehmens, der Schlüssel, Zusammenhänge zwischen den Daten zu erfassen und die richtigen Schlüsse daraus zu ziehen (vgl. Bayer 2013). Im Folgenden wird eine Übersicht über einige technische Funktionalitäten des Programms gegeben.

Tableau setzt bei der Datenanalyse und visuellen Verarbeitung auf die In-Memory Lösung. Während bei anderen In-Memory Lösungen alle Daten im RAM des PCs verarbeitet werden, nutzt Tableau mit seiner speziellen In-Memory Methode die gesamten Speicherressourcen eines Laptops oder PCs zur Verarbeitung (vgl. Tableau Software 2014).

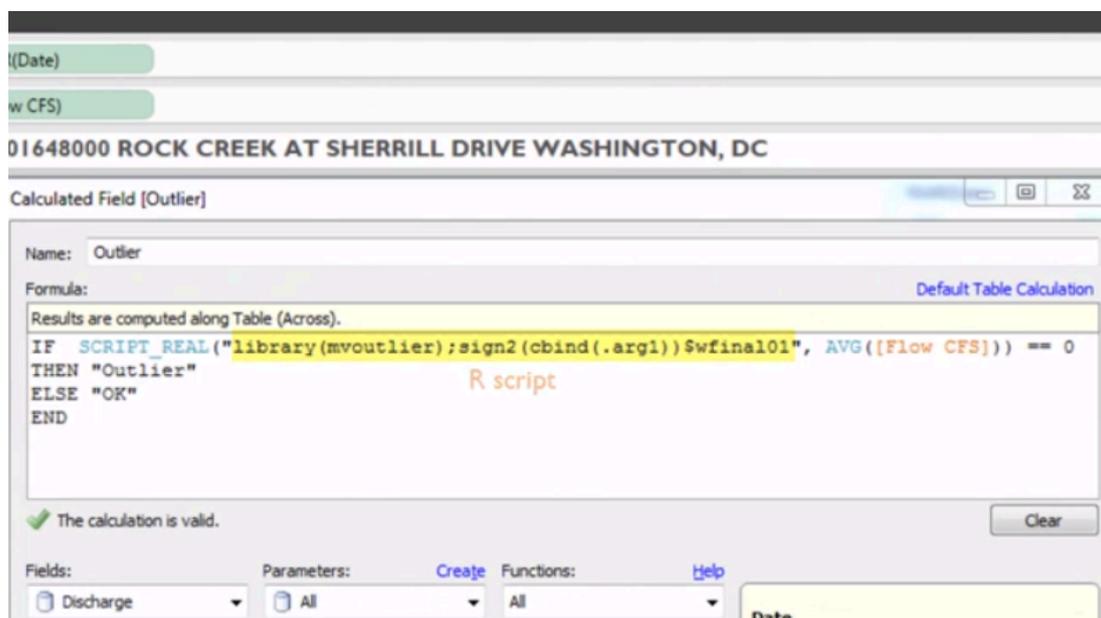
Anders als bei anderen Datenvisualisierungstools, wie beispielsweise Excel, lassen sich bei Tableau über die Schnittstelle VizQL™ pro Datentyp bestimmte Visualisierungen programmieren. VizQL™ ist als das visuelle Pendant zu SQL (verarbeitet Daten in Textform) zu verstehen (vgl. Tableau Software 2014). So können für verschiedene Arten von Daten spezifische Visualisierungsformen selbst programmiert werden.

Die Struktur der zu verarbeitenden Daten kann beliebig sein. Ist beispielsweise das Datum eines Datensatzes in einem unkonventionellen Format vorhanden, so lässt sich über eine

einfache Eingabe im Calculated Field das Datumsformat des gegebenen Datensatzes in Tableau angleichen (vgl. Tableau Software 2014).

Tableau kann sich mit einer Open Source Sammlung statischer Funktionen in der Programmiersprache R verbinden, um so gegebene Funktionen zur Verarbeitung und Darstellung von Daten zu verwenden. Diese Sammlung kann auch durch selbst hinzugefügte Funktionen erweitert werden – die Eingabe erfolgt ebenfalls über das Calculated Field (vgl. ebd.).

Abb. 5: Eingabe einer R-Funktion über das Calculated Field in Tableau



Quelle: Tableau Software 2014, <http://www.tableausoftware.com/de-de/node/25220>, Stand: 09.02.2014

6.2 TIBCO Spotfire

Ein weiteres Tool für Datenvisualisierung großer Datenmengen ist beispielsweise TIBCO Spotfire 6.0. Im Gegensatz zu Tableau wird bei Spotfire für jede Art von Datensatz eine vorgefertigte Visualisierungskomponente ausgewählt (vgl. TIBCO Spotfire 2013). Per Drag & Drop können verschiedene Datensätze, die einem Visualisierungsmuster zugeteilt worden sind, beliebig miteinander verglichen und kombiniert werden. Das ermöglicht dem Anwender eine Sicht auf die verarbeiteten Daten, die nach Möglichkeit eine andere und überraschende Sichtweise auf die Daten zulässt. Hierbei soll das Wertschöpfungstempo durch Automation erhöht werden (vgl. Parthier 2013). Ziel von automatisierten Analyse- und

Visualisierungsvorgängen soll es sein, Trends oder Ausreißer in Geschäftsprozessen einfach und schnell zu identifizieren und verfolgen zu können – und zwar am besten in dem Moment, in dem sie entstehen (vgl. ebd.).

7 Zusammenfassung / Ausblick

Diese Ausarbeitung hat einen Überblick über die aktuellen Techniken einer schnellen und effizienten Datenverarbeitung, sowie deren Analyse- und Visualisierungsmöglichkeiten gegeben. Gerade in Bezug auf große Datenmengen hat sich gezeigt, dass präzisere Prognosen gemacht werden können, je größer die Datenmenge der vorliegenden Daten ist. Dies wird der vierten Charaktereigenschaft von Big Data, der Richtigkeit, gerecht. Ebenfalls wurde gezeigt, dass sich selbst Daten verschiedener Art miteinander vergleichen lassen, weil es diverse Interpretationsmöglichkeiten und Wege gibt, die Daten auf einen gemeinsamen Nenner zu bringen.

Vor allem Unternehmen profitieren vom Einsatz von Big Data Lösungen, um beispielsweise Muster im Konsumentenverhalten zu erkennen oder um frühzeitig Trends zu identifizieren und die Geschäftsprozesse dahingehend auszurichten.

Weil aktuelle Technologien die Disposition schaffen, stetig wachsende Datenmengen immer schneller zu verarbeiten, kann in Zukunft der Großteil der gesammelten Daten gewinnbringend ausgewertet werden. Beispielsweise im Jahr 2012 wurden insgesamt 2,8 Zettabyte Daten erzeugt von denen bis heute nur ein paar Prozent gewinnbringend verarbeitet worden sind (vgl. Schulte 2013). Dank verbesserter Technologien kann dieser prozentuale Anteil in den nächsten Jahren gesteigert werden.

Die Analyse der komplexen Datenstrukturen, sowie deren Visualisierung sind wichtig, um die erfassten Daten miteinander in Verbindung zu bringen und auf Grundlage dessen neue Erkenntnisse zu gewinnen. Vor allem die Visualisierung dieser komplexen Datensystemen in bestimmten Kontexten zueinander wird immer wichtiger, um die sich ergebenden Strukturen verständlich darzustellen und neue Interpretationsmöglichkeiten ausschöpfen zu können.

In Zukunft könnte die Nutzung von Big Data Lösungen nicht nur für Unternehmen profitabel sein, sondern auch für den Endverbraucher. Momentan tritt der Begriff Big Data fast

ausschließlich im Zusammenhang mit der Optimierung von Geschäftsprozessen, Steigerung der Umsätze von Unternehmen und Senkung laufender Kosten für Unternehmen auf, doch auch für den Endverbraucher könnten Big Data basierte Anwendungen in naher Zukunft relevant werden. So könnten beispielsweise Verkehrsinformationen noch präziser in Echtzeit vorgenommen werden und das Unfallrisiko dank smarterer Sensortechnik und Datenauswertung verringert werden (vgl. Asche 2014). Dabei sollte im Fokus stehen, die Daten ausgewertet in einer verständlichen Visualisierung an den Endkunden zu bringen, um einen möglichst großen Nutzen für Einzelverbraucher gewährleisten zu können. Eins steht fest: In den nächsten Jahren wird Big Data auch für den Anwender an Relevanz zunehmen. „Hier sind am Horizont Möglichkeiten für intelligente Anwendungen zu erkennen, deren Potenzial kaum absehbar ist“, so der Informatiker Lars Schmidt-Thieme (Heise 2012).

8 Quellenverzeichnis

- ASCHE, Christoph; Huffingtonpost.de (2014): Diese 50 Big Data Unternehmen werden unseren Alltag verändern. URL: http://www.huffingtonpost.de/2014/01/21/big-data-alltag_n_4626083.html (Stand: 09.02.2014)
- BAYER, Martin (2013): Tableau Software bietet Einblicke in komplexe Zusammenhänge. URL: <http://www.computerwoche.de/a/tableau-software-oeffnet-einblicke-in-komplexe-zusammenhaenge.2547154> (Stand: 09.02.2014)
- BITKOM Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V. (2012): Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte. URL: [http://www.bitkom.org/files/documents/BITKOM_LF_big_data_2012_online\(1\).pdf](http://www.bitkom.org/files/documents/BITKOM_LF_big_data_2012_online(1).pdf) (Stand: 09.02.2014)
- BUBECK, Dr., Uwe; KLEINE BÜNING, Prof. Dr., Hans (2014): Maschinelles Lernen. URL: <http://www.cs.uni-paderborn.de/fachgebiete/fg-kleine-buening/lehre/ws13/maschinelles-lernen.html> (Stand: 09.02.2014)
- BUTTONG, Therese (2013): Big Data schafft neue Realitäten. In: Datability – der richtige Umgang mit unseren Daten. Berlin: Reflex Verlag. S. 9
- CeBIT (2014): Big Data erleichtert das Leben. URL: <http://www.cebit.de/de/news-trends/trends/datability/> (Stand: 08.02.2014)
- CeBIT (2014): Big Data gegen Shoppingfrust. URL: <http://www.cebit.de/de/news-trends/trends/big-data-datability/big-data-gegen-shoppingfrust/> (Stand: 09.02.2014)
- FOCUS (2014): Trendreport Zukunft Reloaded – Big Data und Daten-Visualisierung. URL: http://www.focus.de/finanzen/boerse/aktien/tid-33947/trendreport-zukunft-reloaded-big-data-und-daten-visualisierung_aid_1122117.html (Stand: 09.02.2014)
- GANTZ, John; REINSEL, David; AREND, Carla (2013): The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East - Western Europe. URL: <http://www.emc.com/collateral/analyst-reports/idc-digital-universe-western-europe.pdf> (Stand: 18.02.2014)
- Heise (2012): „Big Data“ – Rohstoff der Informationsgesellschaft. URL: <http://www.heise.de/newsticker/meldung/Big-Data-Rohstoff-der-Informationsgesellschaft-1659622.html> (Stand: 09.02.2014)
- Heise (2014): Big Data – so beherrschen Sie die Datenflut. URL: <http://www.heise.de/microsites/big-data-so-beherrschen-sie-die-datenflut/in-memory-bi-datenanalyse-auf-der-ueberholspur/150/304/917/> (Stand: 08.02.2014)
- HORVARTH, Sabine (2013): Aktueller Begriff – Big Data. Wissenschaftliche Dienste des Deutschen Bundestages Nr. 37/13 (6.November 2013), S.1

- Intel IT Center (2013): Predictive Analytics 101: Next-Generation Big Data Intelligence.
- PARTHIER, Ulrich (2013): Big Data Analysis – Ganzheitliche Sicht mit Visual Analytics. In: IT Research Note. Aying: IT Verlag für Informationstechnik GmbH, S. 1-5
- SCHROECK, Michael; SHOCKLEY, Rebecca, SMART, Dr., Janet, ROMERO-MOLARES, Prof., Dolores; TUFANO, Prof., Peter (2012): Analytics: Big Data in der Praxis - Wie innovative Unternehmen ihre Datenbestände effektiv nutzen. IBM Global Business Services. URL: <http://www-935.ibm.com/services/de/gbs/thoughtleadership/GBE03519-DEDE-00.pdf> (Stand: 08.02.2014)
- SCHULTE, Dagmar (2013): Kreativ dank Kunden-Input. In: Datability – der richtige Umgang mit unseren Daten. Berlin: Reflex Verlag. S. 10
- T-Systems International GmbH (2012): In-Memory Solutions. URL: http://www.t-systems.de/umn/uti/858658_1/blobBinary/Factsheet-In-Memory-Computing-ps.pdf?ts_layoutId=858130 (Stand: Februar 2012)
- Tableau Software (2014): Analyse großer Datenmengen. URL: <http://www.tableausoftware.com/de-de/solutions/big-data-analysis> (Stand: 09.02.2014)
- Tableau Software (2014): R-Integration. URL: <http://www.tableausoftware.com/de-de/node/25220> (Stand: 09.02.2014)
- Tableau Software (2014): Tableau Technology - VizQL™. URL: <http://www.tableausoftware.com/de-de/products/technology> (Stand: 09.02.2014)
- The Apache Software Foundation (2014): Apache™ Hadoop®. URL: <http://hadoop.apache.org/> (Stand: 08.02.2014)
- TIBCO Spotfire (2014): Configure Visualization Properties. URL: <http://spotfire.tibco.com/qrt/3ISMA/presentation.html> (Stand: 09.02.2014)
URL: <http://www.intel.com/content/dam/www/public/us/en/documents/best-practices/big-data-predictive-analytics-overview.pdf> (Stand: März 2013)
- Wikipedia (2014): Big Data. URL: http://de.wikipedia.org/wiki/Big_Data (Stand: 07.02.2014)

9 Versicherung über Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg, den 19.02.2014