

ZT IT-KOLUMNE

Was ist eigentlich Big Data?

„Big Data“ kommt aus dem Englischen und bedeutet übersetzt „große bzw. umfangreiche Daten“. Dieser Artikel gibt einen Überblick in die Welt der „Big Data“ und versucht Klarheit zu verschaffen.

Bereits in der Antike wurden eifrig Daten gesammelt. Man denke nur an die römische Volkszählung unter Kaiser Augustus, die ja bekanntlich dazu führte, dass Maria und Josef nach Bethlehem kamen. Freilich waren die Datenmengen bzw. Zahlen sehr überschaubar gegenüber den heutigen, jedoch bedeutete das Einsammeln und Verarbeiten der Daten damals einen sehr großen menschlichen (Zeit-)Aufwand, denn alles musste manuell geschehen und automatisierte Schritte gab es noch lange nicht. Erst mit der Erfindung des Computers im letzten Jahrhundert sollte sich das alles grundlegend ändern. Mit der Erfindung der Mobilfunktechnik kam dann nochmals ein enormer Schub in das Produzieren von digitalen Massendaten, deren Speicherung und anschließender Verarbeitung.

Wie ist „Big Data“ definiert?

Eines sei an dieser Stelle klargestellt. Es gibt derzeit unterschiedliche Auffassungen, welches Merkmal bei „Big Data“ besonders hervorsteicht. Einerseits sind es wohl zweifelsfrei große Datenmengen, andererseits sind es neue Daten- und Analysetypen oder anwachsender Bedarf an sogenannten Echtzeitanalysen. Echtzeitanalyse ist eine Methode, um Daten sofort, also in der Entstehungszeit ohne Zeitverzögerung zu analysieren. Aber auch andere Bedeutungen werden derzeit oft genannt bzw. geschrieben: Das Anwachsen von Daten aus neuen Technologien, Daten aus sozialen Netzwerken. Einige behaupten sogar, dass es sich bei dem Begriff „Big Data“ um ein aktuelles Modewort handelt. Es sei hier deutlich gesagt: Der Begriff ist kein aktuelles Modewort. Bei „Big Data“ handelt es sich um das immer schneller anwachsende weltweite Datenvolumen, das leider unstrukturiert gespeichert wird. Unstrukturiert bedeutet: Die Daten liegen nicht in einer schön lesbaren z.B. Tabellenform mit Zeilen und Spalten vor. Die Daten müssen für eine Analyse bzw. Auswertung erst aufwendig strukturiert werden. Genau hier liegt eine der großen Herausforderungen, komplexe und astronomisch große Datenmengen in Zukunft so zu strukturieren (auch in Echtzeit), dass Organisationen, Unternehmen und die Bürger einen Mehrwert bekommen. Unstrukturierte Daten liegen in unterschiedlichsten Datenformaten vor, z.B. als Text, als Dokumente (z.B. Office-Dateien), Bilder oder Videos. Prinzipiell kann man sagen, dass unstrukturierte Daten schneller wachsen als strukturierte Daten.

Aber auch der normale Anwender steht immer mehr vor einem Problem, massenhaft generierte digitale Daten einfach zu verwalten. Als Beispiel sei hier die digitale Fotografie genannt. Ein Zahnarzt mit einer modernen Digitalkamera in der Zahnarztpraxis muss die generierten digitalen Fotodateien auch einfach verwalten können. Hierfür müssen Softwareunternehmen auch entsprechende Softwareprodukte an den Anwender liefern, sodass z.B. eine einfache digitale Fotoverwaltung auch möglich ist. Ein einfaches Abspeichern der digitalen Fotodateien ist zwar möglich, jedoch im professionellen Bereich völlig unzureichend.

Merkmale von Big Data

Der Begriff „Big Data“ bezeichnet wie ja bereits beschrieben große und komplexe Datensätze, die besonders durch drei Hauptkriterien charakterisiert sind:

- ihre unglaublich große Menge, die immer schneller anwächst,
- ihre Verschiedenheit und Komplexität, die weiter ansteigt,
- und die Geschwindigkeit, mit der die Daten anfallen und gespeichert werden.

Man spricht auch von den drei V (Volume, Variety, Velocity) der Big Data.

Aktuell verdoppeln sich die zur Verfügung stehenden Daten alle zwei Jahre. Wer generiert denn eigentlich die Datenmengen? Ich möchte hier einige Bereiche aufzählen, die große Datenmengen generieren. Die Wirtschaft, Wissenschaft und Forschung, und im privaten Bereich werden heute große digitale Datenmengen generiert, z.B. immer mehr Sensoren erfassen in immer mehr Lebensbereichen eine Unmenge von Daten:

- Sensoren in Autos (z.B. Sensordaten für die Abstandshaltung bei der Einparkhilfe)

ANZEIGE

LABOR-GEFÜHLE



Wir **LIEBEN** unsere Kunden. Nur so können wir 100% Einsatz bringen. Dazu ein umfangreiches Sortiment und Leistungen: Legierungen, Galvanotechnik, Lasersintern, Experten für CAD/CAM, Metoxit, 3Shape, GC, Roland u.w. Partner – Mit dem Plus an Service! Tel. 040/86 07 66 · www.flussfisch-dental.de

since 1911

FLUSSFISCH



- Sensoren im „Smart Home“ (z.B. Sensordaten für den Sonnenschutz am Fenster)
- Sensoren in der industriellen Produktion (z.B. Sensordaten für die Lackierroboter in der Autoindustrie)
- Daten im Gesundheitswesen (z.B. die neue Gesundheitskarte, digitales Röntgen)
- Ständige Datengenerierung durch die Verwendung von mobilen Geräten (z.B. Smartphones und Tablets)
- Daten im Internet (z.B. soziale Netzwerke wie Facebook)
- Daten in der Finanzindustrie (z.B. Daten von Kontotransaktionen)
- Daten in der Forschung und Wissenschaft (z.B. Daten in der Wettervorhersage, Geodaten, Daten für Modellberechnungen)

Bedeutung für Unternehmen

Gerade Wirtschaftsunternehmen erkennen derzeit den enormen betriebswirtschaftlichen Nutzen von „Big Data“. Informationen, die sich aus den Rohdaten der gespeicherten Datenmengen erschließen lassen, stellen oft einen Wettbewerbsvorteil dar. Mithilfe von sogenannten „Big Data-Analysen“ sollen Informationen in großen Datenmengen transparent für das Unternehmen aufbereitet werden. Dabei sind Umfang und Detailgrad der für die Segmentierung genutzten Daten von außerordentlicher Relevanz. Außerdem steigt der Wert von Informationen, je kontextbezogener und schneller sie den entsprechenden Anwendern zur Verfügung stehen. Gerade große Unternehmen und Konzerne, aber auch Versicherungen, Krankenhäuser, Ärzte, werden in Zukunft „Big Data“ immer mehr auswerten, um Produktvorteile, bessere Produktionsabläufe, bessere Zusammenhänge, Zukunftsprognosen, Service- und Supportmaßnahmen, Kundenverhalten,

Kundenwünsche, individuellere Patientenversorgung und bessere Therapiemöglichkeiten im Krankheitsfall, verbesserte medizinische Diagnostik, schneller Informationen zu erhalten. Derzeit lassen sich Unternehmen sehr viel einfallen, um der Lage Herr zu werden. Prinzipiell müssen die Unternehmen neue Strategien entwickeln, um „Big Data“ optimal zu nutzen.

Es ist zu erwähnen, dass gerade Kundendaten für viele Unternehmen sehr wichtig sind, um Produkte optimal in den Markt zu bringen. Eine Herausforderung besteht z.B. darin, die großen, komplexen und unstrukturierten Datenmengen kontextbezogen zu verbinden. Das Unternehmen „Google“ z.B. liegt im Bereich der Suchmaschinen nach wie vor an der Spitze, weil es in der Lage ist, die unterschiedlichen und komplexen Datenmengen kundenorientiert zu koppeln. Vielleicht wird gerade der Umgang mit „Big Data“ die Zukunft von Unternehmen entscheiden.

Technologien für „Big Data“

Nicht nur der Umgang mit den großen, meist unstrukturierten Datenmengen ist entscheidend und interessant, sondern auch die Technologie, mit der „Big Data“ verarbeitet werden kann. Was nutzen die großen Datenmengen, wenn man diese nicht durch fehlende Hardware und Software strukturieren, analysieren bzw. auswerten kann. Müssen die großen Datenmengen in Echtzeit ausgewertet werden, werden dafür noch zusätzlich besondere Technologien benötigt. Positiv ist hier zu erwähnen, dass die entsprechenden Technologien bereits verfügbar sind. Ich möchte hier ein bekanntes und weitverbreitetes Softwareprodukt vorstellen, das momentan im Bereich „Big Data“

führend ist. Das Open-Source-Framework „Hadoop“ kümmert sich um die verteilte Verarbeitung von großen Daten in einem Netzwerk. „Hadoop“ ist eine skalierfähige Speicher- und Verarbeitungs-umgebung, die u.a. unstrukturierte Daten in eine auswertbare Form bringt. Allgemein zu sagen ist, dass es zum Glück viele Open-Source-Produkte für „Big Data“ gibt.

Ausblick

Immer mehr Unternehmen arbeiten derzeit an „Big Data-Strategien“, um den wirtschaftlichen Erfolg zu erhöhen. Die Investitionen in diesem Bereich steigen kräftig an und zeigen die Wichtigkeit von „Big Data“ bzw. auch einen stattfindenden Wandel in Unternehmen und Organisationen beim Umgang mit den Daten.

Es wäre zu wünschen, dass die Datenflut in Zukunft nicht zu Datenmissbrauch und großflächiger Überwachung der Gesellschaft führt. Der Datenschutz muss dafür weiter verbessert und einheitliche Regelungen im Datenumgang müssen unbedingt geschaffen werden – auch grenzüberschreitend. Es wird sich zeigen, welche Zukunft „Big Data“ hat und wie die astronomischen Datenmengen zum Nutzen der Menschen verarbeitet werden. **ZT**

ZT Adresse

Thomas Burgard Dipl.-Ing. (FH)
Softwareentwicklung & Webdesign
Bavariastraße 18b
80336 München
Tel.: 089 540707-10
info@burgardsoft.de
www.burgardsoft.de
burgardsoft.blogspot.com
twitter.com/burgardsoft

