



# Big Data: Vom Kundenwissen zum Marktwissen

**Die „Big-Data-Herausforderung“: Immer mehr Nutzer wollen in nahezu Echtzeit die Daten aus der immer mehr ausufernden Datenflut und aus immer mehr und unterschiedlichsten Quellen analysieren.**

Diese Big-Data-Herausforderung beschreibt recht gut, was Big Data ist und was es bedeutet:

- Extremes und anhaltendes Wachstum des Datenvolumens. Allein im Jahr 2011 hat die Menschheit 1,8 ZB neue Daten produziert. (1 Zetta Byte = 1 Milliarde Tera Bytes). Diese Daten sind zudem mehrheitlich unstrukturiert.
- Die Anzahl der Datenquellen nimmt ebenfalls massiv zu. Es sind nicht nur die Social Media, sondern auch maschinell erzeugte Daten wie die Lokalisierungsdaten aus dem mobilen Internet oder Messdaten aus intelligenten Ablesegeräten (Telephonie, Strom, Gas, Wasser, RFID etc.).
- Die in dieser extremen Menge von Daten verborgene Information und das darin verborgene Wissen wollen sich mehr und mehr Unternehmen und innerhalb der Unternehmen mehr und mehr Fachbereiche zu nutzen machen. Die Anzahl der Mitarbeiter, die Information brauchen und wollen steigt so rasant an.
- Information hat den größten Wert, wenn sie neu und aktuell ist. Denn Dinge in der digitalisierten Welt passieren jetzt und überall. Daher brauchen wir Information in Echtzeit, hier und jetzt.

Jeder dieser vier Trends an sich ist eine große Herausforderung an die Informationstechnologie. Jetzt gilt es aber diese vier Anforderungen gemeinsam zu meistern, sonst gibt es keine Antworten! Damit ist klar, dass die traditionellen Business Intelligence (BI)-Technologien zum Analysieren von Daten nicht mehr ausreichen, ja, man kann sogar sagen, obsolet geworden sind. Nicht mehr geeignet sind relationale Datenbanken, die am Datenvolumen scheitern. Nicht mehr geeignet sind traditionelle ETL-Prozesse zum Laden von Data Warehouses und Datamarts. Sie sind nicht nur zu langsam, sondern kommen auch mit der Vielfalt und Beschaffenheit der Quellen nicht klar. Nicht mehr geeignet sind traditionelle BI-Werkzeuge, die auch viel zu langsam sind, in der Regel nicht zur Analyse unstrukturierter Daten ausgelegt sind und bis zum Aufkommen von Tablets auch immer noch nicht die Benutzerfreundlichkeit hatten, um wirklich in den Fachabteilungen akzeptiert zu werden. Jetzt sind Innovationen auf allen Ebenen gefragt.

Bevor wir uns solche Innovationen anschauen, fragen wir uns, was denn Big Data mit Kundenorientierung und Kundenbeziehungsmanagement (CRM) zu tun hat und wie wir vom Big Data beim Kundenwissen profitieren können. Beginnen wir dazu mit zwei Beispielen.

## Beispiel: Big Data im Handel.

Im Handel kämpft man schon lange mit sehr großen Datenmengen, den Kassensbon-Daten beispielsweise. In den Kassensbon-Daten steckt viel Kundenwissen, denn sie eignen sich gut, um die Produktprofitabilität pro Kunden auszurechnen. Das ist nämlich eine wichtige Kennzahl zur Steuerung von personalisierten Kampagnen und Echtzeit-Produkttempfehlungen, also eine im analytischen CRM typische Kennzahl im Handel für die Outbound- und Inbound-Kundenkommunikation. Nur konnte man mit den traditionellen BI-Werkzeugen diese Kunden/Produkt-Profitabilität-Kennzahl nicht ausrechnen, da das zugrunde liegende Datenvolumen zu groß war und die Analysen einfach zu lange dauerten. Mit Big-Data-Technologien lässt sich das Problem lösen.

## Beispiel:

### Stimmungsanalysen in den Social Media.

Insbesondere die Konsumgüterhersteller interessieren sich für die Meinungen aller Marktteilnehmer zu den eigenen Produkten und Marken ebenso im Vergleich dazu zu den Produkten und Marken der Mitbewerber. Hier bieten die Social Media neue und ergiebige Quellen. Neben der Chance, Zielgruppen mit chirurgischer Präzision zu bearbeiten, bringen Social Media aber auch Risiken: In Blogs, Foren und Tweets wird schlicht und einfach alles über Produkte und Unternehmen gesagt – inklusive echter Lügen: Expertenforen können schnell und nachhaltig Werbeversprüche entzaubern. Zur Auswertung all dieser Kommunikation in den Social Media beginnt man mit dem Identifizieren und Extrahieren der relevanten Quellen im Big Data. Dann gilt es, diese Quellen auszuwerten. Das liefert nicht nur statistische Information, wo und wie viele Spuren sich im Web und in den Social Media befinden, sondern mit Hilfe von Stimmungsanalysen (sentiment analysis) lässt sich auch die Tonalität der Beiträge bestimmen. Auf Basis eines **Social Media Monitoring** kann im nächsten Schritt eine **Social-Media-Interaktion** aufgebaut werden. Das Unternehmen kann jetzt auf relevante Beiträge sofort reagieren und intervenieren. Das bringt Vorteile vor allem im Kundenservice oder bei der Einführung neuer Produkte im Markt, da sich sofort eine Kommunikation mit Communities im Web aufbauen und unterhalten lässt. So sind beispielsweise schon in verschiedenen Service-Call-Centern die Agenten auch zu Social-Media-Agenten geworden, die jetzt eine Multikanal-Kommunikation mit den Kunden über die traditionellen und die Social-Media-Kanäle führen können. Das ist der Schritt von Outbound- und Inbound-Kundenkommunikation zur Unbound-Kundenkommunikation. So schafft man eine gesteigerte Time-to-Market und eine höhere Kundenbindung bei einer vergleichsweise überschaubaren Investition.

Der Nutzen dieser Transparenz durch Big Data liegt auf der Hand: Ein Hersteller von Konsumgütern will beispielsweise wissen, wie Konsumenten seine Marken und Produkte und/oder die seiner Mitbewerber in den einschlägigen Blogs diskutieren und bewerten. Oder eine Hotelkette interessiert sich für das elektronische Feedback ihrer Gäste und/oder für die Bewertungen der Mitbewerber. Alles wird möglich, wenn all diese Daten zugreifbar und auswertbar werden. In der Verbindung mit den Unternehmenskundendaten erhält man so nicht nur eine 360°-Sicht auf den Kunden, wie immer im CRM gefordert, sondern sogar eine 360°-Sicht auf den gesamten Markt: Mitbewerber, Kunden der Mitbewerber, Presse, Marktmultiplikatoren etc. Denn im Big Data spiegelt sich ja der Markt mit allen Marktteilnehmern wieder.

Diese Beispiele und Überlegungen zeigen deutlich, wie man im Kundenbeziehungsmanagement aus Big-Data-Nutzen ziehen kann und wie Big-Data-Technologie das traditionelle analytische CRM weiterbringt. Schauen wir uns jetzt kurz die wesentlichen Innovationen an, die die Big-Data-Technologie ausmachen und so die traditionellen BI-Methoden, -Werkzeuge und -Technologien erweitern und fortsetzen.

**Textanalytik.** Textanalytik ist eine neue Klasse von Analytik, die linguistische Verfahren mit Suchmaschinen, Text Mining, Data Mining und Algorithmen des maschinellen Lernens verbindet. Mit Textanalytik lassen sich Daten aus dem Big Data systematisch auszuwerten, aber niemand spricht gerne darüber. Man will seinen Kunden nicht gerne sagen, dass man inzwischen eine nahezu gläserne Transparenz geschaffen hat. Daher ist es hier besonders wichtig, alles konform mit den Datenschutzgesetzen zu halten.

Textanalytik wie jede Analytik sollte stets mit einem Performance Management verbunden sein ganz im Sinne des bekannten Leitsatzes: Man kann nur managen, was man auch messen kann. Benötigt werden unter anderem Metriken zur Berechnung der Relevanz von Quellen und der Vernetzung von Quellen, Scorecards zum Visualisieren und Verdichten der Monitoring-Ergebnisse und schließlich auch ein Reporting, insbesondere ein Ausnahme-Reporting, um automatisch Auffälligkeiten in den Social Media wie ein Anstieg von Tags, von Autoren, von Threads etc. anzuzeigen.

Auch wenn Textanalytik-Lösungen dem Namen nach als Komplettlösungen daher kommen, heißt das nicht, dass diese „Produkte“ out of the box einsatzbereit sind. Textanalytik stellt im Moment ein arbeitsintensives und sehr lohnendes Feld

für Berater dar. Eine individuelle Beratung ist notwendig, bis einerseits die Unternehmen entsprechende Kenntnisse aufgebaut haben, und andererseits die Hersteller ihre Tools soweit standardisiert und parametrisiert haben, dass die Anwender damit aus vordefinierten Bausteinen bestimmte Klassen von Lösungen erstellen können. Zudem stellt sich auch in der Textanalytik ein ähnliches Problem wie in Data Mining: Die Interpretation der Ergebnisse erfordert ein tiefes Fachwissen. Denn mittels mathematischer Verfahren gefundene Strukturen und Beziehungen sind zwar faktisch richtig, aber solche Fakten müssen nicht unbedingt etwas mit der realen Welt zu tun haben. Das gilt ganz besonders für gefundene Fakten aus dem Big Data, denn Bewertungen können aus Freundschaft erfolgt sein, Meinungen in Blogs können manipuliert und Profile in sozialen Netzen auch frei erfunden sein. Daher ist es ganz wichtig, die durch Textanalytik gefundenen Fakten als Hypothesen auf Plausibilität zu testen. Das ist heute noch in den meisten Fällen dem Menschen/Berater vorbehalten.

**Agile Big-Data-Extraktionswerkzeuge.** Solche Werkzeuge können Big-Data-Quellen ohne vordefinierte Schnittstelle und ohne Programmierung visuell erschließen. Das ist ein wichtiger Vorteil, denn Schnittstellen im Big Data sind entweder in vielen Fällen nicht vorhanden, müssten also erst langwierig spezifiziert und programmiert werden, oder sie leisten nicht das, was man gerade braucht. Schnittstellen schränken vielfach den Zugriff auf Daten ein. Daten sind so dem Zugriff entzogen. Ein browser-basiertes Extraktionswerkzeug dagegen gibt Zugriff auf alle sichtbaren Daten und zwar sofort und auch in Echtzeit. Nicht nur Daten können so extrahiert werden, sondern jede Information aus Web-Applikationen. Die agilen Extraktionswerkzeuge können auf jede Schicht von Web-Applikationen zugreifen und Information extrahieren. Ein weiterer Vorteil dabei: Diese browserbasierte, visuelle Schnittstelle ist für alle Quellen im Big Data immer die gleiche. Bei der Extraktion arbeiten die Werkzeuge dann wie Mikro-Workflow-gesteuerte Roboter. In den Workflows werden Regeln und Schleifen eingesetzt. So kann eine umfangreiche Workflow-Logik aufgebaut werden, die Extraktionen jeder Komplexitätsstufe ohne Programmierung ermöglicht. Die Roboter können mittels einer Management-Konsole geplant und kontrolliert werden. Damit kein Webmaster das Extrahieren bemerkt und womöglich Gegenmaßnahmen ergreift, können sie auch bewusst langsam arbeiten, um einen menschlichen Leser vorzutäuschen. All das macht sie zuverlässig und sicher.

So beschafft man sich die notwendigen Web-Daten automatisiert und schnell. Dabei darf man allerdings gesetzliche Aspekte nicht vergessen. Selbst wenn Daten nicht geschützt und öffentlich sind, verstößt man unter Umständen bereits gegen die AGBs der Seiten, wenn man automatisch ausliest. Daten, die auf den Netzgemeinschaften als privat gekennzeichnet sind, darf man so natürlich unter keinen Umständen nutzen. Bei öffentlichen Daten ist ein opt-in der Kunden mehr als empfehlenswert. Die juristischen Details wollen wir aber hier nicht weiter vertiefen, die muss jeder in seinem Unternehmen beachten.

**Analytische Datenbanken.** Herkömmliche relationale Datenbanken sind zeilenorientiert. Sie sind zum Managen und Analysieren großer Datenmengen nicht gut geeignet, da sie nicht zum schnellen Lesen entwickelt und gebaut wurden. Daher nutzen analytische Datenbanken alternative Methoden, die sich auch miteinander kombinieren lassen: Spalten-Orientierung, Kompressionsverfahren, In-Memory-Verarbeitung, Appliances, Cluster-Verfahren und Parallelisierung, um die gängigsten zu nennen. Dazu kommen auch neue noSQL-Sprachen zur Abfrage wie das von Google vorgeschlagene MapReduce.

Analytische Datenbanken sind so konzipiert, dass Tuning-Aufwände, die in konventionellen Systemen notwendig sind und Zeit und Ressourcen kosten, nahezu entfallen. Die Datenbanklösung legt Indizes automatisch an, analysiert und komprimiert die Daten selbständig und verteilt sie optimal über die Knoten. Intelligente Algorithmen fangen Server-Ausfälle auf und sorgen dafür, dass das System für Nutzer innerhalb weniger Sekunden ohne dessen Zutun wieder zur Verfügung steht.

Analytische Datenbanken sind eine Voraussetzung, um aus Big Data Nutzen zu ziehen. Sie bieten wesentliche Vorteile wie:

- Informationen sind flexibler abrufbar und stehen bis zu 1.000mal schneller zur Verfügung.
- Die Nutzerzufriedenheit erhöht sich signifikant aufgrund des schnelleren und flexibleren Zugriffs auf Information. Es können jetzt Daten analysiert werden, die vorher ohne Nutzen, aber mit Kosten gespeichert wurden. Das unterstützt und schafft bessere Entscheidungen.
- Die IT wird entlastet, da die analytischen Datenbanken hoch automatisiert sind und ein spezielles Wissen über Datenbankdesign und Tuning deutlich weniger gefragt ist.

## Fazit

Die Nutzung von Big Data reichert das traditionelle Kundenwissen an und schafft darüber hinaus ein Marktwissen, denn in den Social Media und im Web findet man ja nicht nur seine Kunden, sondern auch seine Mitbewerber und deren Kunden ebenso wie auch Presse, Marktmultiplikatoren und alle anderen Marktteilnehmer. Nach der Outbound- und Inbound-Kommunikation über Kampagnen-Management sind wir nun definitiv im Unbound angekommen. Jeder kommuniziert mit jedem und alles ist – im Rahmen der gesetzlichen Vorschriften – transparent verfolgbar und erfassbar.

Die kritischen Erfolgsfaktoren – neben dem Beherrschen der Big-Data-Technologien – sind aber die gleichen wie in der Vergangenheit:

- Ohne Sponsor auf der Geschäftsführungsebene geht es nicht.
- Ohne eine Strategie für analytisches CRM im Rahmen von CRM und BI geht es nicht: analytisches CRM ist kein Selbstzweck, sondern macht die CRM-Prozesse intelligent.
- Ohne Governance (Organisation per Kompetenzzentrum und entsprechende Prozesse) geht es nicht. Die Konsequenz ist ein Alignment von Business und IT (gemeinsame Begriffe und Sprache, gemeinsames Verständnis).
- Ohne die Menschen geht es nicht: Die Mitarbeiter müssen motiviert und mitgenommen werden, auch ins Big Data.



**Dr. Wolfgang Martin**  
Wolfgang Martin Team  
Mitglied im CRM-Expertenrat  
E-Mail: [info@wolfgang-martin-team.net](mailto:info@wolfgang-martin-team.net)