

# *Big Data – Bedeutung Nutzen Mehrwert*

*Wir zeigen Ihnen, wie Sie Big Data durch gezielte Datenanalyse gewinnbringend nutzen können und unterstützen Sie bei der Entwicklung einer digitalen Wertschöpfungsstrategie, die perfekt auf die Anforderungen Ihres Unternehmens für das Zeitalter der Digitalisierung zugeschnitten ist.*





---

## Vorwort

Über große Institutionen heißt es gelegentlich, dass sie ungeahnte Chancen nutzen könnten, wenn sie nur wüssten, was sie wissen. Waren damit früher die in Köpfen und Ordnern gespeicherten Informationen gemeint, so lässt sich die Wendung heute ausweiten auf die Menge der in einem Unternehmen vorhandenen Daten – sowie derjenigen, die täglich neu hinzukommen. Unvorstellbar groß sind die Datenmengen, die mittlerweile in den Unternehmen und anderen Einrichtungen vorliegen – und oft fehlen ihnen die Kapazitäten für einen durchdachten Umgang mit dieser Herausforderung.

Solche Datenvolumen werden als Big Data bezeichnet und oft sind ihre äußerst attraktiven Potenziale bislang ungenutzt. So bergen sie etwa neue Erkenntnisse über Optimierungsmöglichkeiten betrieblicher Abläufe, über Marktlücken, Kundenwünsche und vieles mehr. Doch wie können die Unternehmen die Daten für ihren spezifischen Bedarf erfassen und auswerten und so wichtige Informationen erschließen, die ihnen bisher nicht zugänglich waren?

Die effiziente Nutzung von Big Data erfordert einerseits geeignete technologische Werkzeuge und andererseits „intelligente“ Algorithmen, mit denen sich die gewünschten Daten aus dem vorhandenen Pool herauslösen, darstellen und miteinander in Beziehung setzen lassen. Und vor allem bedarf es einer Strategie, die die Richtung eines Big-Data-Projekts definiert und aus der sich die Kernfragen ableiten lassen, die das Projekt beantworten soll.

Big-Data-Projekte sind Aufgabe der Unternehmensleitung. Vorausschauende Führungskräfte haben das erkannt und in ihren Unternehmen damit begonnen, den in der Datenflut verborgenen Schatz zu heben. Dafür benötigen sie jedoch auch entsprechend qualifizierte Experten.

Gemeinsam mit Ihnen möchten wir daran arbeiten, das Potenzial von Big Data auch für Ihr Unternehmen zugänglich zu machen – damit Sie im globalen Markt mit innovativen Lösungen punkten und sich von den Mitbewerbern abheben können. Wir nennen das die „digitale Wertschöpfung“. Wir wünschen Ihnen eine anregende Lektüre!

**Barbara Lix**      **Jan Stüben**

---

# *Inhaltsverzeichnis*

Abbildungsverzeichnis .....	5
A Die Datenwelt verändert sich – und mit ihr die Situation der Unternehmen.....	6
B Big-Data-Projekte gezielt planen und Weichen stellen .....	11
C Umsetzung von Big-Data-Projekten – strategisches Vorgehen.....	13
D Die Vorteile von Big Data im Überblick .....	18
E Exkurs 1: Umsetzung von Big-Data-Projekten – technologische Werkzeuge.....	21
F Exkurs 2: Neue Erkenntnisse – in jeder Branche .....	25
G Exkurs 3: Datenwissenschaftler als neue Akteure im Unternehmen .....	27
Ihre Ansprechpartner.....	30

---

# Abbildungsverzeichnis

Abb. 1	Big-Data-Dimensionen .....	9
Abb. 2	Strategiekreislauf zu einer nachhaltigen, digitalen Wertschöpfungsstrategie .....	14
Abb. 3	Klassifizierung unterschiedlicher Analyseziele .....	16
Abb. 4	Entscheidungsgrundlage als Zieldefinition .....	17
Abb. 5	Vorteile durch die Nutzung von Big Data .....	19
Abb. 6	Beispiele technologischer Möglichkeiten.....	21

## ***A Die Datenwelt verändert sich – und mit ihr die Situation der Unternehmen***



Die Zahl der Daten, die wir in unserem Alltag verbreiten, steigt rapide an, da eine wachsende Zahl von Geräten, mit denen wir täglich in Berührung kommen, Daten erzeugt und übermittelt. Weltweit nehmen die Datenmenge und die Vielfalt der Datenstrukturen beständig zu: 2,5 Trillionen Bytes pro Tag werden rund um den Globus erzeugt – rund 90 Prozent der vorhandenen Daten entstanden allein in den letzten zwei Jahren. In Zukunft ist damit zu rechnen, dass sich die weltweit erzeugte Datenmenge alle zwei Jahre mindestens verdoppeln wird.

Die Quellen der Daten, die in diesen immensen Pool neu einfließen, sind zahlreich und Daten, die bisher – gezwungenermaßen – oft als irrelevant betrachtet wurden, da man zu ihrer Verarbeitung nicht imstande war, gewinnen jetzt an Bedeutung. Diese Daten haben ihren Ursprung in den Technologien der letzten fünf Jahre.

---

Nachfolgend einige Beispiele für Daten aus unterschiedlichen Quellen:

- **Daten von mobilen Endgeräten**

Sie erlauben Einblicke in die Aufenthaltsorte unzähliger Menschen weltweit und haben etwa für ortsbasiertes Marketing einen immensen Wert.

- **Daten aus sozialen Netzwerken**

Twitter, Facebook und andere Netzwerke sind zu allgegenwärtigen Kommunikations- und Interaktionsplattformen geworden. Sie beherbergen eine unvorstellbare Datenmenge. Momentan ist daher ein deutlicher Wille zur Verlagerung von der Transaktionsverarbeitung hin zur Kommunikations- und Interaktionsverarbeitung zu bemerken.

- **Finanzmarktdaten**

Finanztransaktionen, Handel oder Zahlungen laufen heute mit einer derartigen Geschwindigkeit und in solchen Umfängen ab, dass von Regierungsseite bereits der Wunsch nach einer Drosselung geäußert wurde. Nur mit leistungsfähigen IT-Lösungen können diese Datenvolumen noch verarbeitet und betrachtet werden.

- **Multimediatdaten**

Digitale Daten wie Videos und Bilder bilden eine der größten Datenquellen, mit einem riesigen Volumen. Eine Videominute enthält mehr Daten als ein Text mit 120.000 Seiten. Multimediatdaten machen mittlerweile den Großteil des Datenflusses im Internet aus. Sie bergen einen unschätzbaren Wert und ihre Auswertung eröffnet völlig neue Möglichkeiten. So etwa wäre es in der Medizin möglich, durch Implementierung automatisierter Algorithmen Röntgenaufnahmen systematisch zu speichern, auszuwerten und zu vergleichen, damit künftig Anomalien besser erkannt werden können.

- **Textdateien**

Sie bilden womöglich die kostbarste Form unstrukturierter Daten. Man ist zunehmend in der Lage, Textdateien in Daten umzuwandeln, die von Computern ebenso gut verstanden, analysiert und verändert werden können wie von Menschen.



- **ERP-Systeme von Unternehmen**

Diese Systeme mit ihren Schnittstellen zu mobilen Endgeräten sowie die Vernetzungen mit anderen Systemen können wertvolle Informationen enthalten, die über ein standardisiertes Reporting und Controlling hinausgehen. Beispielsweise durch die Aufdeckung von Regelverstößen oder eine Überwachung des Verhaltens hinsichtlich der Zahlungsmoral. Ebenso sind Aussagen zur Kundenfluktuation oder zur Erstellung von Ratingprofilen der Kundschaft möglich.

- **Internet der Dinge**

Sensoren verfolgen heute eine Fülle von physischen Objekten – Pakete, Lagerbestände (RFID) oder Sensordaten und GPS-Signale im Flottenmanagement – und liefern dabei riesige Datenmengen, die ausgewertet werden wollen. Hier können durch Big-Data-Anwendungen neue Dienstleistungen erkannt und angeboten werden.

- **M2M**

Machine to Machine Kommunikation wird ebenfalls immer mehr Daten erzeugen und damit die Datenflut weiter vergrößern. Dieser Prozess findet unsichtbar statt. Er bietet allerdings ein enormes Erkenntnispotenzial, das aktuell in seiner Tragweite noch gar nicht erkannt wird. Als Beispiel dient hier die Interaktion mit Bewegungsdatenmustern über Smartphones und Messpunkte in Stores.

---

Derzeit findet eine enorme Verschiebung statt – nicht nur der Daten selbst, sondern auch hinsichtlich des Ortes und der Art und Weise ihrer Speicherung sowie Verwendung: von standortbasierten hin zu Cloud-basierten Daten. Dies hat Auswirkungen auf die Wirtschaftlichkeit der EDV; zudem entsteht eine neue Welle der Datenfragmentierung. Gleichzeitig findet eine Verschiebung vom Desktop hin zum mobilen Gerät statt, das den User befähigt, kontextsensibel und standortbasierte Dienste zu nutzen. In der Folge kommt es zu einem rasanten Anstieg der Geräte- und Sensorinteraktionsdaten. Diese aktuellen Veränderungen heizen das explosive Wachstum an Daten zusätzlich an; es kommt zu größeren Datenmengen, größerer Datenvielfalt und höheren Übertragungs- und Verarbeitungsgeschwindigkeiten.

Die Fähigkeit, diese Daten zu speichern, sie zu sammeln, miteinander zu kombinieren und die aus diesen Prozessen resultierenden Ergebnisse für detaillierte Analysen zu nutzen bietet immenses Potenzial – und wird nun für Unternehmen aller Branchen immer greifbarer.

**90 % der vorhandenen Daten  
entstand in den letzten 2 Jahren.**



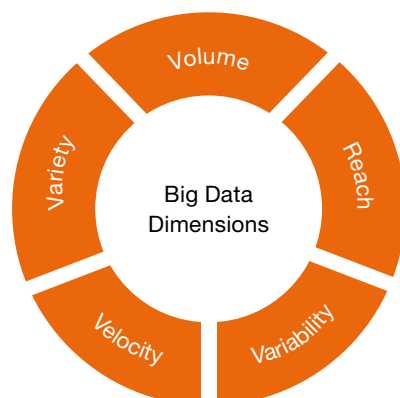
**Definition „Big Data“**

Der Begriff „Big Data“ hat sich innerhalb kürzester Zeit zu einer Art Trendwort des Jahres 2013 entwickelt. Was versteht man darunter?

Er bezeichnet Datensätze, deren Größe die Kapazität und Fähigkeit klassischer Datenbank-Softwaretools zur Erfassung, Speicherung, Verarbeitung und Analyse übersteigt. Diese Beschreibung ist subjektiv und beinhaltet, da es hier keine genauen Messgrößen gibt, keine genaue Angabe dazu, wie groß ein Datensatz sein muss, um als Big Data zu gelten. Die Frage, welches Datenvolumen als Big Data gilt, kann je nach Branche, Software sowie den in einem bestimmten Industriezweig verfügbaren oder verwendeten Tools variieren.

Gemeint ist jedoch, dass der Begriff „Big Data“ das Phänomen wachsender Datenmengen, -quellen und -strukturen bezeichnet. Angesichts zunehmender Komplexität gelangen klassische Business-Intelligence-Strukturen an ihre Grenzen und Unternehmen stehen vor ganz neuen Herausforderungen.

**Abb. 1 Big-Data-Dimensionen**



**Allgemein kann man Big Data anhand folgender fünf Dimensionen beschreiben:**

**1. Datenmenge (Volume)**

Neben den Daten, die von außen in ein Unternehmen fließen, wächst auch die von ihm selbst erzeugte Datenmenge. Dabei ist unerheblich, ob es um Tera-, Peta-, Exa-, oder Zetabytes geht, um Datenbanken oder Streamingdaten, um Speicherung oder In-Memory-Lösungen. Entscheidend ist vielmehr, wie zielgerichtet und effektiv die Daten genutzt werden können. Als Big Data werden Datensätze bezeichnet, deren Größe die Fähigkeiten typischer Datenbanksoftware zur Speicherung, Verarbeitung und Analyse übersteigt.

**2. Datenvielfalt (Variety)**

Die heute zu verarbeitenden Daten stammen aus unterschiedlichsten Quellsystemen (Sprache, Bild, Text), Eingabesystemen (z. B. Notebooks, mobile Endgeräte, digitale Bezahlssysteme, soziale Medien) und Applikationen. Die Datenstrukturen und -formate sind somit sehr unterschiedlich.

### 3. **Geschwindigkeit und Datenauswertung (Velocity)**

Bei der Auswertung von Daten ist Schnelligkeit oft noch wichtiger als die Menge. Durch Auswertungen in Echtzeit oder annähernder Echtzeit können Unternehmen schneller und flexibler agieren und sich so strategische Wettbewerbsvorteile verschaffen. Sie können Trends zeitnah erkennen: zum Beispiel in E-Shops durch Suchanfragen nach einem bestimmten Produkt, das bislang nur die Konkurrenz anbietet.

### 4. **Datenquellen (Reach)**

Ein zentraler Aspekt von Big Data ist die Datenherkunft. Es gilt zu prüfen, welche Daten im Unternehmen vorhanden sind, welche davon bereits aktiv genutzt werden und welche noch nicht bzw. weshalb nicht. Es kann auch sinnvoll sein, externe Daten einzubeziehen oder diese innerhalb der Organisation – etwa mithilfe einer Datenlandkarte – neu zu erheben. Zudem ist zu klären, welche Daten für eine bestimmte Entscheidung jeweils erforderlich sind.

### 5. **Komplexität (Variability)**

Die Analyse polystrukturierter Daten erfordert enorme Rechnerleistungen, da komplexe Algorithmen zur Anwendung kommen. Mit entsprechender Technologie und richtiger Anwendung lassen sich jedoch völlig neue Erkenntnisse generieren (Data Mining), die es ermöglichen, Zusammenhänge, Abhängigkeiten und Widersprüche zu erkennen oder sogar Vorhersagen und somit effizientere und effektivere Managemententscheidungen zu treffen.

Der gezielte Umgang mit Big Data ist eine Aufgabe, der sich die Unternehmen stellen müssen und zugleich eine Chance, die sie nutzen sollten, um Einblicke in neue Datentypen und -inhalte gewinnen und Antworten auf Fragen zu finden, die bisher außerhalb ihrer Reichweite lagen. Denn die effiziente Nutzung von Big Data ermöglicht es, neue Erkenntnisse zu gewinnen, mit denen die strategische und operative Steuerung, etwa in Hinblick auf Risikominimierung, Kundenbindung, Marktpositionierung oder Profitsteigerung verbessert werden kann. Die Betriebseffizienz kann somit durch Big Data optimiert werden.

***Die effiziente Nutzung von Big Data ist Aufgabe und Chance zugleich.***

## ***B Big-Data-Projekte gezielt planen und Weichen stellen***



Um das Potenzial von Big Data maximal nutzen zu können, bedarf es einer durchdachten digitalen Wertschöpfungsstrategie.

- Insbesondere sind Hard-, Soft- und Freeware-Anwendungen intelligent mit analytischen Verfahren zu kombinieren. Im Zuge dieser Kombination wandeln sich reine Datenanalysen, auch aufgrund enormer Rechnerleistungen, zu komplexen Algorithmen.
- Herkömmliche relationale und zeilenorientierte Datenbanken benötigen für die aufwendigen Berechnungen zu viel Zeit und zu viele Ressourcen. Deshalb sollte eine produktive Nutzung von Big Data auch eine Umstellung auf analytische Datenbanken sowie Parallelisierungstechniken einschließen.
- Die Kosten für die Speicherung und Verarbeitung riesiger Datenmengen fallen mittlerweile kaum mehr ins Gewicht. Dennoch nutzen längst nicht alle Unternehmen ihre Datenbestände für zielgerichtete Analysen. Eine große Hürde stellen hierbei die Heterogenität der Daten und das Fehlen geeigneter Methoden dar.

Angesichts der rasant wachsenden Datenmengen und zunehmenden Leistungsfähigkeit immer preisgünstigerer Systeme wird die digitale Wertschöpfungsstrategie künftig eine wachsende Rolle für die Unternehmen spielen. Hierzu sollen folgende Leitlinien dienen.

**Mit einer digitalen Wertschöpfungsstrategie gehen Sie vorausschauend in das digitale Zeitalter.**

#### **Unternehmensstrategie als Ausgangspunkt für die richtigen Fragen**

Die digitale Wertschöpfungsstrategie muss aus der Unternehmensstrategie abgeleitet sein und mit ihr im Einklang stehen. Nur dann können die wesentlichen Fragen formuliert werden, auf die das Big-Data-Projekt Antworten liefern soll. Diese Fragen können sowohl prozess- und produkt- als auch unternehmensspezifisch sein. Für ihre Auswahl ist vor allem der Zeitpunkt entscheidend, zu dem die Daten bzw. die Ergebnisse der Datenanalyse bereitstehen und verwendet werden sollen. Je nach Zielsetzung kann es erforderlich sein, die Daten entweder in Echtzeit, annähernder Echtzeit oder in einer anderen zeitlichen Taktung zu analysieren.

#### **Wie kann das Potenzial von Daten erkannt und erschlossen werden?**

In den IT-Abteilungen sind zwar zahlreiche Daten vorhanden, doch werden diese auch dort eingesetzt, wo sie hilfreich wären? Wird ihr Potenzial richtig eingeschätzt? Nur wer bei der Nutzung von Daten eine Führungsrolle übernimmt, kann sich von seinen Mitbewerbern abheben. Deshalb stehen die Unternehmen vor der Aufgabe, aus vielen unterschiedlichen Daten auf effiziente Weise hilfreiches Wissen zu generieren.

#### **Einsatz intelligenter Algorithmen**

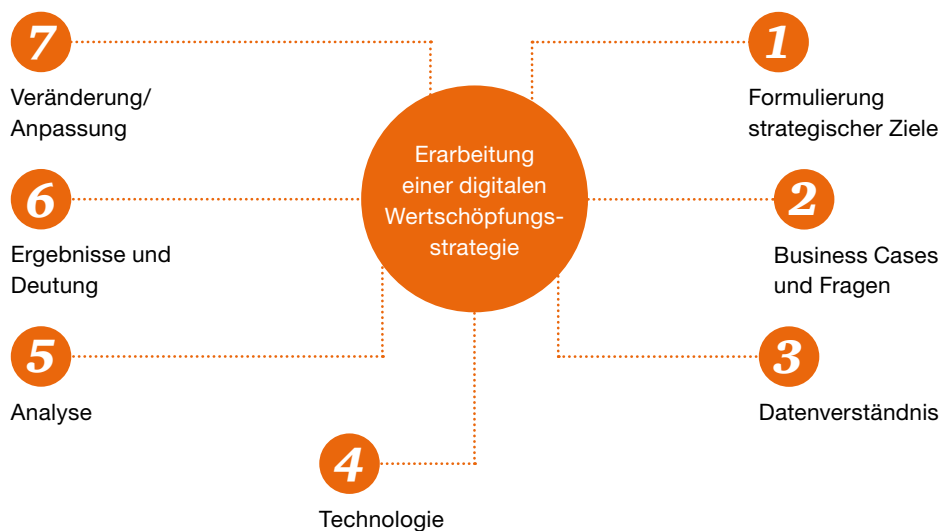
Erst der Einsatz intelligenter Algorithmen und der guten Kenntnis über statistische und mathematische Verfahren sowie passender Methoden der Datenanalyse ermöglicht ein vollständiges Ausschöpfen der Potenziale von Big Data. Zunächst sind aus dem Datenpool die für einen bestimmten Informationsgewinn erforderlichen Daten herauszufiltern. Im nächsten Schritt muss der Anspruch an die technischen Qualitätsmerkmale definiert werden, damit das optimale Datenanalyseverfahren bestimmt werden kann. Ziel dabei ist, die vorhandenen Daten so zu kombinieren, dass eventuelle Abhängigkeiten und Zusammenhänge aufgedeckt werden können. Das mittels der Verwendung intelligenter Algorithmen gewonnene Wissen kann dann im Unternehmen strategisch genutzt werden.

## ***C Umsetzung von Big-Data-Projekten – strategisches Vorgehen***



PwC begleitet Sie ganz konkret bei der Umsetzung Ihres Big-Data-Projekts in sieben Schritten, die sich zu einem Strategiekreislauf zusammenfügen. Ziel dabei ist die Erarbeitung einer digitalen Wertschöpfungsstrategie, die perfekt auf die Bedürfnisse, Kenntnisse und Möglichkeiten Ihres Unternehmens abgestimmt ist.

**Abb. 2 Strategiekreislauf zu einer nachhaltigen, digitalen Wertschöpfungsstrategie**



**Die sieben Schritte zur digitalen Wertschöpfungsstrategie –  
Schritt 1: Wir formulieren mit Ihnen die strategischen Ziele der  
Big-Data-Initiative**

**Leitfragen (Beispiele)**

- Welches Ertragswachstum (Segment) streben Sie innerhalb der nächsten Jahre an?
- Welche Chancen und Möglichkeiten sehen Sie für Ihr Unternehmen im Rahmen der Digitalisierung?
- Streben Sie im Kontext der alltäglichen Geschäftsprozesse nach einer Risikominimierung?

Wir verschaffen uns ein vertieftes Verständnis Ihres Geschäftsmodells und erstellen aus dem Ist-Zustand Ihres Unternehmens Geschäftsprognosen, die Ihnen eine Hilfestellung zur Formulierung der strategischen Ziele bieten.

Dieser Schritt ist erforderlich, da die digitale Wertschöpfungsstrategie mit Ihrer Unternehmensstrategie im Einklang stehen muss. Nur so können im nächsten Schritt die für Ihr Unternehmen relevanten Fragen formuliert werden.

### **Schritt 2: Wir leiten die Business Cases ab und formulieren relevante Fragen**

Gemeinsam mit uns formulieren Sie die Fragen, auf die Sie durch die Nutzung von Big Data Antworten erwarten. Wir unterstützen Sie dabei, diese Fragen aus Ihrer Unternehmensstrategie abzuleiten und entsprechend zu formulieren. Insbesondere soll deutlich werden, welche Informationen Sie aus den Ergebnissen der Datenanalysen ziehen möchten und welche Effekte Sie erwarten.

- Wünschen Sie detailliertere Informationen über Ihre Kunden?
- Möchten Sie Ihre Position innerhalb des Wettbewerbs genau erkennen?
- Möchten Sie Risikofaktoren aufdecken?
- Welche Geschäftsprozesse und -abläufe sind unmittelbar betroffen?
- Wo besteht Transformationsbedarf?
- Möchten Sie Trends hinsichtlich Markt und Kunde zeitnah erkennen?

### **Schritt 3: Gemeinsam mit Ihnen gewinnen wir ein umfassendes Verständnis Ihres Daten- und Informationsbedarfes**

Die zuvor formulierten Erwartungen werden nun genauer für die von Ihnen benötigten Daten definiert.

- Welche Daten sind verfügbar?
- Welche Daten sind erforderlich?
- Sind alle diese Daten vorhanden?
- Müssen durch Datenanreicherungsverfahren/Datenerhebungsverfahren zusätzliche Daten beschafft werden?
- Wie sollen die Daten und Ergebnisse strukturiert sein?
- In welcher zeitlichen Taktung möchten Sie die Ergebnisse sehen?
- Welche Qualität haben die im Unternehmen vorhandenen Daten?

### **Schritt 4: Wir finden für Sie maßgeschneiderte Technologien**

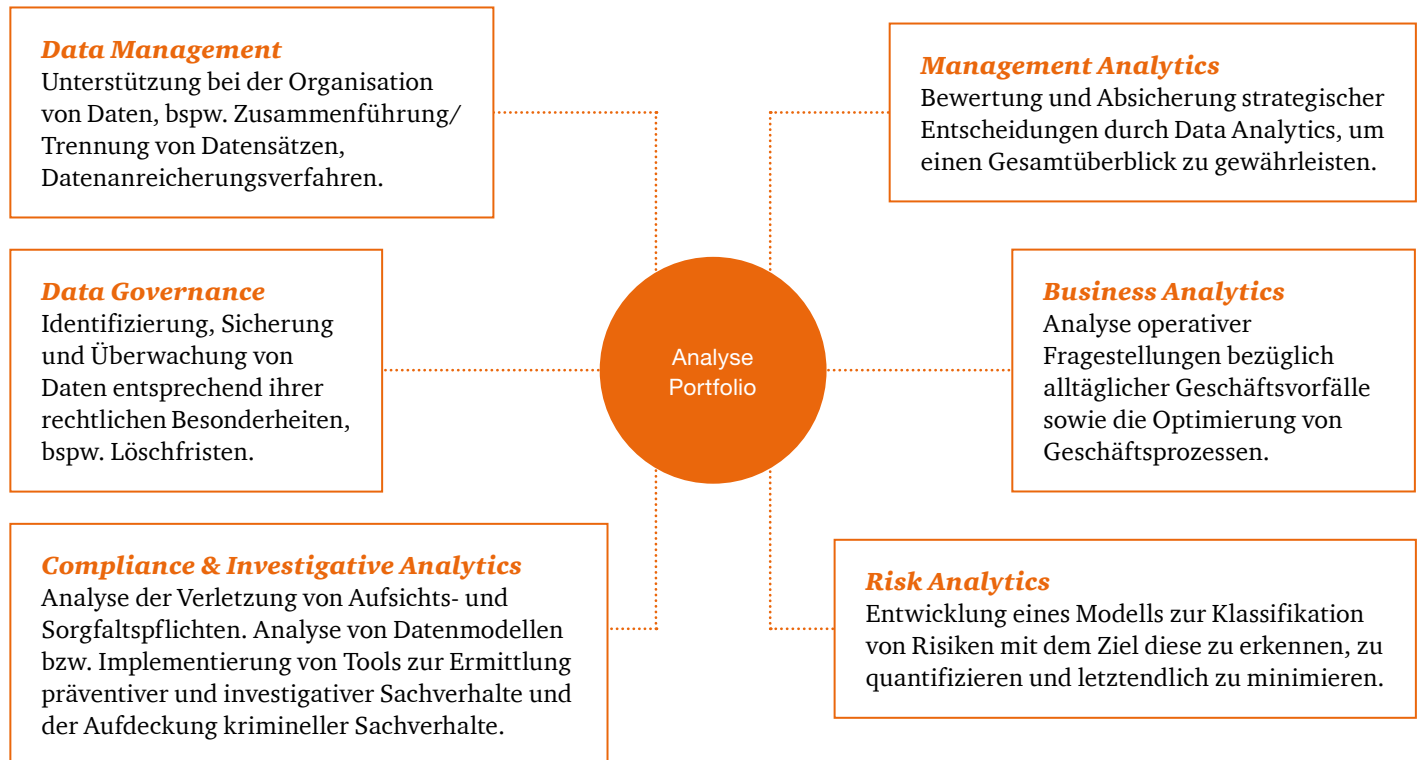
Auf der Basis der zuvor definierten Anforderungen unterstützen wir Sie bei der Auswahl derjenigen Technologien, die optimal zu Ihrem Unternehmen passen. Dabei zeigen wir Ihnen ein umfassendes Repertoire an anwendbaren Technologien auf.

**Aus einem breiten Spektrum von möglichen Analysen erarbeiten wir die richtigen Algorithmen speziell für Ihre Anforderungen.**

**Schritt 5: Wir führen die notwendigen Spezialanalysen mit Ihnen durch oder implementieren diese**

Hier greifen wir auf ein breites Spektrum von Analysen zurück, die in der folgenden Abbildung dargestellt sind.

**Abb. 3 Klassifizierung unterschiedlicher Analyseziele**



Auf Basis dieser Analysen generieren wir speziell für Ihre Anforderungen konzipierte, komplexe und zukunftsweisende Analysealgorithmen. Diese implementieren wir in Ihre Systeme und stellen so sicher, dass die Analysen automatisiert ablaufen und Sie Zugriff auf die gewünschten Informationen und Ergebnisse haben.



**Schritt 6: Ergebnisinterpretation**

Wir erläutern Ihnen das Vorgehen bei der Interpretation der neuen Informationen und Ergebnisse. Damit schaffen wir die Voraussetzung, um die Informationen künftig als Entscheidungsgrundlage zu verwenden und einen ganzheitlichen Überblick zu erhalten. So wird es möglich, neue Erkenntnisse zu gewinnen, Vorhersagen zu treffen, die zuvor nicht machbar waren und eine bisher nicht erzielbare Transparenz herzustellen.

**Abb. 4 Entscheidungsgrundlage als Zieldefinition**

**Schritt 7: Veränderung und Anpassung der Geschäftsprozesse und -strukturen**

Anknüpfend an die Unternehmensstrategie und auf der Grundlage der neu erlangten Erkenntnisse unterstützen wir Sie bei der Umsetzung der notwendigen Veränderungen in Ihrem Unternehmen. Die Gewährleistung eines Gesamtüberblicks und einer umfassenden Transparenz sowohl Ihrer Geschäftsprozesse als auch der Marktsituation bieten Ihnen eine fundierte Entscheidungsgrundlage. Auf ihrer Grundlage können Sie mit unserer Unterstützung die Änderungen formulieren, die vorgenommen werden müssen, um Ihre Geschäftsprozesse effizienter zu gestalten, sie zu optimieren und Vorteile gegenüber Ihren Mitwettbewerbern zu erlangen.

Möglicherweise werden die neu gewonnenen Erkenntnisse auch zu einer nachträglichen Anpassung der Unternehmensstrategie führen oder einen Bedarf an zusätzlichen Daten sichtbar machen. Eine solche erneute Anpassung ist als positiver Schritt hin zu einer noch präziser abgestimmten Strategie und Vorgehensweise zu werten. Somit schließt sich der Kreis der digitalen Wertschöpfung.

***Die Gewährleistung eines Gesamtüberblicks und einer umfassenden Transparenz sowohl Ihrer Geschäftsprozesse als auch der Marktsituation bieten Ihnen eine fundierte Entscheidungsgrundlage.***

## ***D Die Vorteile von Big Data im Überblick***



Durch die Analyse von Big Data können Sie für Ihr Unternehmen zahlreiche Vorteile erschließen:

**Abb. 5 Vorteile durch die Nutzung von Big Data**



### ***Klare Entscheidungsgrundlagen schaffen***

Die effiziente Nutzung von Big Data erzeugt einen umfassenden Überblick über betriebliche Szenarien. Sie erhöht die Transparenz und den Informationsgehalt der Daten und vermittelt Einsichten, die zuvor nicht möglich waren. Auch wichtige Zusammenhänge und Abhängigkeiten können besser erkannt werden. Im Ergebnis führt dies zu deutlich fundierteren Entscheidungsgrundlagen und effizienteren Geschäftsprozessen.

### ***Geschäftsprozesse optimieren***

Durch genaue Analysen der unternehmenseigenen Daten und ihre Anreicherung mit Daten aus ergänzenden Quellen können die Prozesse innerhalb Ihres Unternehmens dargestellt und optimiert werden. Abstrahierte, mathematische oder visuelle Darstellungen sind dabei besonders geeignet, Schwachstellen und Fehler aufzudecken. Ihre Prozesse werden effizienter und die Kosten sinken.

### ***Risiken kalkulieren***

Auch Risiken können durch die Analyse von Big Data wirksam minimiert werden, da die Entscheidungen somit auf belegbaren Erkenntnissen basieren. Zudem ist es möglich, Risiken vorausschauend einzukalkulieren.

Eintrittswahrscheinlichkeiten oder auch die Relevanz der Risiken lassen sich exakt bewerten. Zudem können Zusammenhänge zwischen unterschiedlichen Risikofaktoren erkannt werden. So ist es möglich, sich diejenigen Optionen aufzeigen zu lassen, die trotz eines bestehenden Risikos gerade noch profitabel sind.

### ***Profitabilität steigern***

Eine präzise Absatz- und Bedarfsplanung sichert den finanziellen Erfolg eines Unternehmens. Mit der effizienten Auswertung von Big Data basiert diese stets auf realistischen Annahmen und fundierten Werten. Verlässliche Prognosen von Bedarf und Nachfrage verhindern Lieferengpässe sowie hohe Abschreibungen. Umsatzverluste durch Überschuss oder Mangel an dem vom Unternehmen angebotenen Produkt werden ebenfalls vermieden und führen zu Kostensenkungen bzw. Profitsteigerungen. Zudem hilft eine langfristig vorausschauende Planung dabei, Transportkosten zu senken, denn erstens werden teure ad-hoc-Situationen vermieden und zweitens kann eine langsamere Transportart gewählt werden, die meist kostengünstiger und ökologischer ist. Ähnliche Erfolge lassen sich auch mit der Implementierung automatischer Dispositionsabläufe erzielen.

### ***Preisgestaltung dynamisch anpassen***

Die gezielte Nutzung von Big Data eröffnet die Möglichkeit, Preise in Echtzeit, automatisch und dynamisch an die aktuelle Markt- und Wettbewerbssituation anzupassen. Zudem können die vielfältigen Faktoren, die auf einen Preis einwirken, vollständig in die Preisgestaltung einbezogen werden. Es ist sogar möglich, kundenspezifische Preissegmente im Bereich des e-commerce darzustellen, um z. B. tageszeit- oder wochentagspezifische „Prime-Time“ Preise und Werbungen anzubieten und so gegebenenfalls die interne Logistik besser auszulasten. Dies birgt wesentliche Vorteile des e-commerce gegenüber dem stationären Handel.

### ***Optimale Kundenorientierung***

Die Kombination von Daten aus unterschiedlichen Quellen ermöglicht ganzheitliche Kundenanalysen. Jeder Kunde kann entsprechend seines Wertes (ClientValue) für das Unternehmen eingestuft werden. Speziell die Auswertung von Social-Media-Daten birgt für die Gewinnung unternehmensrelevanter Informationen ein großes Potenzial. So können zum Beispiel Verhaltensmuster erkannt, simuliert und für die Prognose des zukünftigen Kundenverhaltens genutzt werden. Ein genauerer Blick auf den Kunden schafft die Voraussetzungen für mehr individuelle Kundenorientierung. Auch Marketingkampagnen können gezielter auf bestimmte Kunden oder Kundengruppen zugeschnitten werden; gleiches gilt für den direkten Umgang mit den Kunden. Derartige Maßnahmen erhöhen die Kundenzufriedenheit, schaffen Kundenbindung und steigern die Profitabilität des Unternehmens.

### ***Ausschöpfen unentdeckter Marktpotenziale***

In der globalen Welt wird es für die Unternehmen immer wichtiger, sich durch das Aufspüren von Marktlücken sowie passgenaue Angebote von ihren Mitbewerbern zu unterscheiden. Es wird jedoch zunehmend schwierig oder gar unmöglich, einen umfassenden Überblick über regionale Märkte oder den Weltmarkt zu erhalten – der für Global Player jedoch von immenser Bedeutung ist. Durch die Nutzung von Big Data, vor allem in Bezug auf den Zugang von Daten unterschiedlicher Weltregionen und Absatzmärkte, kann ein solcher Überblick erlangt und können bislang verborgene Marktpotenziale identifiziert und ausgeschöpft werden.

***Durch eine klare Vorgehensweise innerhalb der digitalen Wertschöpfungsstrategie stellen Sie eine erfolgreiche Umsetzung sicher.***

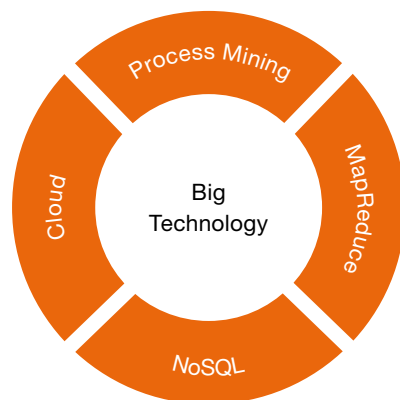
## E Exkurs 1: Umsetzung von Big-Data-Projekten – technologische Werkzeuge

Neue technologische Werkzeuge gewährleisten eine effiziente Auswertung von Big Data. Sie sind meist sogar wesentlich kostengünstiger als herkömmliche Technologien und sorgen somit dafür, dass Big Data in einem sehr breiten Anwendungsbereich eingesetzt werden können.

Insgesamt sind die Anpassungsmöglichkeiten der neuen Technologien äußerst vielfältig. Mit dem entsprechenden Know-how lassen sie sich auf sehr unterschiedliche Anforderungen und Fragestellungen ausrichten. Auch Kombinationen der neuen Technologien untereinander sowie Kombinationen mit bewährten, in den Unternehmen vorhandenen Anwendungen sind stets vorstellbar. Mittlerweile sind bereits vorgefertigte kombinierte Lösungen auf dem Markt, die mehrere dieser Technologien und ihre Vorzüge verbinden.

Wesentliche Elemente dieser neuen Technologien sind Process Mining, MapReduce, NoSQL-Datenbanken (NoSQL – Not only Structured Query Language) sowie Cloud Computing. Diese werden im Folgenden näher beschrieben.

**Abb. 6 Beispiele technologischer Möglichkeiten**



### **Process Mining**

Unternehmen, die täglich komplexe Prozesse bewältigen, bietet Process Mining ein beträchtliches Potenzial zur systematischen Auswertung der digitalen Spuren innerhalb dieser Prozesse.

Im Unterschied zur aufwendigen Optimierung von Prozessen durch die Verbesserung der zwischenmenschlichen Interaktionen werden beim Process Mining direkt aus den vorhandenen IT-Log-Daten grafische Prozessmodelle sowie strukturelle Prozessdetails generiert – etwa Zeitspannen, die in kürzester Zeit den Ist-Zustand eines Prozesses aufzeigen. Aufgrund der automatisierten Analyse kann Process Mining typischerweise folgende Fragen beantworten:

- Wie sieht der genaue Aufbau meines Prozesses aus?
- Liegen Engpässe vor?
- Entspricht der Ist-Prozess dem Soll-Prozess?

#### Die Vorteile von Process Mining

- Anhand der Ergebnisse der Process-Mining-Anwendung lassen sich Fehler oder Schwachstellen in Prozessen aufdecken.
- Die Anwendung liefert fundierte Grundlagen für eine anschließende Ursachenanalyse in den Fachabteilungen des Unternehmens, die zu einer Optimierung beitragen können.

#### MapReduce

Einer der wesentlichen Algorithmen für die Verarbeitung von Big Data ist MapReduce. Dabei findet die Bearbeitung der Daten in zwei Schritten statt:

- Beim **Map-Schritt** wird jeder Datensatz mittels einer Funktion gelesen und verarbeitet. Das geschieht innerhalb von verteilten Prozessen, also jeweils auf dem Server, auf dem der betreffende Datensatz gespeichert ist. Anschließend werden die Ergebnisse unabhängig voneinander und parallelisiert in einer Key-Value-Datenstruktur gespeichert. Dabei erhalten immer diejenigen Daten denselben Key, die anschließend im zweiten Schritt gemeinsam verarbeitet werden.
- Beim **Reduce-Schritt** werden die Daten aller Server gesammelt, gruppiert und weiterverarbeitet. Werte mit unterschiedlichen Keys werden unabhängig voneinander, gegebenenfalls parallel, weiterverarbeitet.

Durch die Einteilung in zwei getrennte Schritte laufen beide Phasen hochparallel und verteilt ab; zudem ist gewährleistet, dass auch innerhalb großer Datenmengen Zusammenhänge hergestellt werden können. Je nach Problemstellung können die Schritte zudem in ihrer Komplexität angepasst werden.

#### Die Vorteile von MapReduce gegenüber herkömmlichen Verarbeitungs- algorithmen

- Aufgrund der Ausführung des Map-Schrittes auf dem jeweiligen Server kann die Rechenkapazität aller Server genutzt werden.
- Die Auswertung des Map-Schrittes erfolgt meist durch lineares Lesen aller Daten. Ein Ergebnis liegt rasch vor, da derartige Leseoptionen sehr schnell durchgeführt werden können.
- Da bei beiden Schritten eine hohe Parallelisierung gewährleistet ist, wird bei größeren Datenmengen nur eine größere Anzahl von Servern, nicht aber die Anschaffung leistungsstärkerer Server notwendig.

### **Implementierung in Hadoop**

Seit Mitte des vergangenen Jahrzehnts liegt MapReduce als Implementierung in Java unter dem Namen Apache Hadoop vor. Die Verarbeitung mit Hadoop gewährleistet gute Ergebnisse in folgenden Bereichen:

- **Datenlokalität**

Die Daten werden dort verarbeitet, wo sie gelesen werden, indem Datenspeicher und MapReduce-Algorithmus zusammengeführt werden. Der Code eines MapReduce-Befehls wird beim Start auf alle Datenknoten kopiert und dann auf alle dort befindlichen Daten angewendet. Diese Vorgehensweise ist ausgesprochen leistungsstark.

- **Datenredundanz**

Damit sichergestellt ist, dass bei keinem der beiden Schritte Teilergebnisse verloren gehen oder wegen möglicher Ausfälle nicht berechnet werden, liegen die Daten redundant vor. Jeder Schritt kann somit beliebig oft wiederholt werden.

- **Verarbeitungszeit**

Diese kann mit Hadoop weitgehend verkürzt werden. Die bereitstehenden Ressourcen lassen sich so optimal nutzen.

#### **Die Vorteile der Implementierung in Hadoop**

- Durch die Implementierung des Hadoop Distributed File System (HDFS) wird eine Reduzierung der Verarbeitungszeit erzielt. Die Datensätze einer Datei können dabei an einer beliebigen Stelle eines Clusters aus beliebig vielen Rechnern redundant abgelegt werden. Zugriffe auf das HDFS müssen somit nicht zur Konvertierung durch einen zentralen Rechner gehen. Nur die Struktur des Dateisystems wird über den NameNode zentral vorgegeben.
- Zudem existieren für die Auswertung von Daten mithilfe von Hadoop aufgrund der zusätzlichen Implementierung von Hive und Pig, SQL-ähnliche (SQL – Structured Query Language) Abfragesprachen. Dies gewährleistet, dass sich die Unternehmen bei einer Umstellung auf Hadoop sehr schnell einarbeiten können, da MapReduce-Anfragen wie klassische Datenbankabfragen eingelesen werden können, ohne dass eine aufwendige Programmierung notwendig wäre.
- Hadoop ist die erste Open-Source-Implementierung von MapReduce und dem Google File System (GFS) und bis heute eine der führenden Anwendungen. Allerdings existieren mittlerweile zahlreiche weitere Systeme, die MapReduce unterstützen, so zum Beispiel diverse NoSQL-Datenbanken.

### **NoSQL**

NoSQL-Datenbanken stellen in der Datenwelt in etwa das Gegenteil zu herkömmlichen, relationalen Datenbanken dar. NoSQL steht dabei für die Ergänzung von SQL mit dem Ziel, dieses zur Verwaltung großer, polystrukturierter Datenmengen zu befähigen.

NoSQL-Datenbanken sind auf bestimmte Anforderungen hin hochspezialisierte Strukturen und hinsichtlich ihrer Speicherstruktur weitestgehend vergleichbar mit Index Only Tables. NoSQL-Datenbanken arbeiten, indem sie Datensätze durch Sharding und Replikation skalieren:

- Beim Sharding werden die Daten in Shards, also Fragmente, aufgeteilt. Die Zuweisung eines bestimmten Wertes oder Attributs macht ersichtlich, welchem Shard eine Datei angehört. Auf jedem Server innerhalb eines Clusters können ein oder mehrere Shards gespeichert werden. Kommt ein neuer Server hinzu, so werden einige Shards auf diesen umgelagert. Diese Vorgehensweise erlaubt es, die Datenbank sowohl für Schreib- als auch für Lesezugriffe zu skalieren.
- Die Replikation, aufgrund derer die Daten auf mehreren Servern gleichzeitig verfügbar sind, macht die Datenbank zuverlässiger. Zudem vermindert sich die Zeitspanne eines Lesezugriffs, da auf eines der Replikate zugegriffen werden kann.

#### **Die Vorteile von NoSQL**

- NoSQL-Datenbanken können durch das einfache Hinzufügen neuer Server leistungsfähiger gemacht werden.
- Die verwendeten Server müssen nicht hochverfügbar sein, da diese Eigenschaft bereits durch die Replikation erfüllt ist. Somit sind die Kosten für einen Ausbau der Datenbankkapazität oder -leistung überschaubar.
- Die Entwicklung hin zum Cloud Computing macht es zudem möglich, Server jeweils nur nach Bedarf bzw. vorübergehender Auslastung anzumieten.

### **Cloud Computing**

Cloud Computing wird es auch kleineren Unternehmen ermöglichen, ohne hohe Kosten für die Implementierung neuer Hard- und Software die Vorteile von Big Data auszuschöpfen. Denn die Cloud bietet die Option, Speicher- und Rechenressourcen, die im Rahmen von Big Data anfallen, kostengünstig zu erwerben.

#### **Die Vorteile von Cloud Computing**

- **Dienstleistungscharakter**  
Der Zugriff erfolgt über eine zuvor definierte Schnittstelle. Der Nutzer hat somit vollen Zugriff auf die bereitgestellten Ressourcen, greift aber über ein Userinterface zu.
- **Flexibilität**  
Je nach Bedarf können die bereitgestellten Ressourcen über die Laufzeit der Cloud angepasst werden und somit auch die vom Nutzer zu tragenden Kosten. Manche Anbieter bieten mittlerweile neben den üblichen Abonnements oder Flatrates auch rein volumenabhängige Tarife an.
- **Standardisierung**  
Der Zugriff auf die Ressourcen erfolgt über Standardwerkzeuge und -programme, sodass keine langwierige Einarbeitung notwendig ist.



## **F Exkurs 2: Neue Erkenntnisse – in jeder Branche**

Im Rahmen eines Big-Data-Projekts lassen sich zahlreiche branchenspezifische und unternehmensindividuelle Analysen umsetzen, die vielfältige neue Erkenntnisse und Prognosen ermöglichen. Die folgende Auswahl branchenbezogener Anwendungen zeigt nur einen Bruchteil der vorstellbaren Möglichkeiten und soll das beträchtliche Potenzial von Big Data verdeutlichen.

### **Handel und Konsumgüter**

- Durch den Einsatz ganzheitlicher Kundenanalysen kann die Kundenbindung, vor allem im Onlinehandel und bei Onlinespielen, verbessert werden.
- Das Produktportfolio lässt sich an die Vorlieben und Bedürfnisse der Kunden anpassen und somit optimieren. Angebote können auf bestimmte Kundengruppen zugeschnitten und genau dann präsentiert werden, wenn die Nachfrage am größten ist – auch preissensitiv.
- Denkbar ist auch eine Verbesserung von Retourenvorgängen bzw. eine Reduktion der Anzahl der Retouren. Durch die Verknüpfung unterschiedlicher Daten können die Ursachen für Rücksendungen aufgedeckt und nach Möglichkeit behoben werden. Da zudem die Wahrscheinlichkeit einer Rücksendung berechnet werden kann, lassen sich kundenspezifische Vorkehrungen treffen und es kann präventiv reagiert werden.
- Detaillierte Analysen der Lieferanten erlauben spezifische Einstufungen, etwa im Hinblick auf deren Zuverlässigkeit.
- Mittels intelligenter Absatzplanung können (Lager-)Bestände und Bestellungen optimiert werden.
- Das Unternehmen versteht seine Kunden und Kundengruppen sowie deren Gewohnheiten noch besser und kann kundenorientiert Werbung betreiben und spezifische Angebote anbieten.

### **Energiewirtschaft**

- Analysen der Kundenstruktur erlauben eine Segmentierung in Kundengruppen, die beispielsweise aufzeigt, welche Kunden besonders profitabel sind.
- Auch Kunden-Scoring-Modelle können abgeleitet werden.
- Auf Basis des generierten Wissens über Kunden, Kundengruppen und deren Verhalten sind zudem Clusteranalysen nach bestimmten gewünschten oder benötigten Tarifen denkbar.
- Durch eingehende Analysen der Abläufe in der Energieverteilung sind Optimierungen zu erzielen, die zu günstigeren Tarifen und folglich einer Gewinnung von Neukunden führen können.

### **Technologie, Medien und Telekommunikation**

- Aus ganzheitlichen Kundenanalysen können Erkenntnisse abgeleitet werden, die Optimierungsmöglichkeiten für Tarifarten und -strukturen eröffnen. Attraktivere Tarife lassen sich dann für die Kundenbindung und -gewinnung nutzen. Denkbar sind etwa Anpassungen in Abhängigkeit von der Tageszeit oder von geografischen Daten sowie Optimierungen von Gruppentarifen.
- Zudem sind auch Kundenabwanderungsanalysen – sogenannte Churn-Analysen – eine Option. Unzufriedene Kunden können identifiziert werden und es wird möglich, kundenseitigen Kündigungen vorzubeugen oder Schwachstellen des Unternehmens zu entdecken.

### **Transport und Logistik**

- Auch hier kann eine Optimierung der Retourvorgänge, wie weiter oben für den Handel beschrieben, sich positiv auswirken.
- Durch ganzheitliche Analysen können Risiken effizient und von einer fundierten Basis aus bewertet werden. Gerade bei komplexen Investitionen oder Projekten kann das für die Unternehmen sehr vorteilhaft sein.
- Denkbar ist zudem eine umfassende Optimierung der Prozessabläufe, des Personaleinsatzes und der Kapazitätsplanung.
- Ferner ist die Implementierung einer Track-and-Trace-Anwendung möglich, die den logistischen Prozess überschaubar darstellt und eine Echtzeitüberwachung rund um die Uhr gewährleistet.

### **Finanzdienstleistungen**

- Banken und Versicherungen profitieren davon, ihre Kunden und deren Lebensumstände und Vorlieben zu kennen. Ganzheitliche Kundenanalysen mit Daten aus neuen Quellen erlauben es, die Kunden auf einer persönlichen und dennoch abstrakten Ebene kennenzulernen.
- Angebote können dann gezielt auf die Lebenssituationen der Kunden zugeschnitten werden. Wo der gesamte Lebenszyklus eines Kunden abgebildet und die Angebotspalette entsprechend angepasst wird, kann eine lebenslange Kundenbindung erzielt werden.
- Risiken können individuell prognostiziert und Risikozuschläge flexibel definiert werden. Gerade für Banken ist dies im Hinblick auf Kreditausfallrisiken vorteilhaft.
- Maßgeschneiderte Angebote und flexibles Reagieren auf die wechselnden Lebensumstände eines Kunden tragen dazu bei, die Beratungsangebote zu verbessern.
- Marketingmaßnahmen können effektiver gestaltet werden, indem sie etwa auf die Bedürfnisse besonders profitabler Kundengruppen fokussieren.

### **Industrielle Produktion**

- Die neuen Analysen bieten die Möglichkeit, Fertigungsprozesse maßgeblich zu optimieren, indem sowohl die Maschinen und Anlagen als auch die an der Fertigung beteiligten Mitarbeiter sowie weitere Variablen (Energiekosten, Lastprofile, Revisions- und Wartungsarbeiten) optimal ausgelastet werden. Komplexe Produktionsabläufe lassen sich visuell oder mathematisch simulieren, sodass bislang unerkannte Schwachstellen oder Fehler sichtbar werden.
- Die Produktionskosten und die Qualität neuer Waren können präzise und unter Einbeziehung aller vorhandenen Variablen angegeben werden.
- Zudem ist es möglich, Vorhersagen über störungsbedingte Produktionsausfälle, etwa den Ausfall von Maschinen, zu treffen und präventiv zu reagieren, sodass teure Produktionsstillstände auf ein Minimum reduziert werden. Derartige Erkenntnisse ermöglichen es den Unternehmen, ihre Maschinen nicht mehr nach dem Kalender, sondern nach Bedarf zu warten, was zu immensen Kostenersparnissen führen kann.
- Automatische Prozesse der Disposition und des Bestandsmanagements vereinfachen und optimieren die Materialplanung. Hohe Lagerhaltungskosten können somit gesenkt werden.

## **G Exkurs 3: Datenwissenschaftler als neue Akteure im Unternehmen**

Die wirksame Nutzung von Big Data setzt unter anderem die passenden Werkzeuge, Techniken und Methoden voraus. Versteifen sich die Unternehmen hier zu früh auf eine bestimmte Technologie, so verliert das Big-Data-Projekt schnell seine Agilität. Genau diese ist jedoch – vor allem im frühen Entwicklungsstadium – überaus wichtig.

Der Fokus sollte daher zunächst auf der Benennung qualifizierter Spezialisten für das Projekt liegen. Denn die wertschöpfende Nutzung von Big Data stellt vorwiegend ein People Business dar und nur sekundär ein Product Business. Anwendungen wie MapReduce und Apache Hadoop vereinfachen zwar die Handhabung, ohne vertieftes Wissen um die mathematischen Hintergründe und die Prozesse können jedoch keine optimalen Ergebnisse erzielt werden.

Bei der Analyse großer Datenmengen wird auf statistische Berechnungen und auf Wahrscheinlichkeitstheorien zurückgegriffen. Bei den Ergebnissen geht es um Eintrittswahrscheinlichkeiten und Prognosen, nicht um exakte Kalkulationen.

Es gibt für die verschiedenen Anwendungsgebiete von Big Data keine Pauschallösungen. Daher ist es wichtig, ein großes Repertoire an Werkzeugen, etwa Programmiersprachen und Plattformen, zu kennen. Dies ist eine Aufgabe für eine neue Gruppe von Akteuren: die Datenwissenschaftler oder Data Scientists. Diese Berufsbezeichnung existiert zwar erst seit wenigen Jahren, doch ist die Zahl der entsprechenden Experten bereits stark gestiegen und die Unternehmen sehen sich zunehmend im Wettbewerb um die fähigsten Köpfe.

### **Anforderungsprofil eines Datenwissenschaftlers**

Das Problem bei der Identifikation dieser Experten ist, dass es bisher keine Studiengänge gibt, die Abschlüsse im Bereich Datenwissenschaft anbieten. Zudem besteht kaum Konsens darüber, welche Rolle den Datenwissenschaftlern im Unternehmen zukommt und wo im Unternehmen sie einzuordnen wären, damit ihr Einsatz den größtmöglichen Nutzen erzeugt. Auch ein Leistungsmaßstab existiert bislang nicht.

Zwei zentrale Fragen sind folglich:

- Wie sieht das Aufgabenfeld des Datenwissenschaftlers aus?
- Welche Fähigkeiten benötigt er?

***Es gibt für die verschiedenen Anwendungsgebiete von Big Data keine Pauschallösungen. Daher ist es wichtig, ein großes Repertoire an Werkzeugen zu kennen.***

**Ein Datenwissenschaftler benötigt sowohl soziale Kompetenzen, die Fähigkeit zu assoziativem Denken als auch ein umfassendes Verständnis der Geschäftsprozesse.**

Wesentliche Aufgabe eines Datenwissenschaftlers ist es, Daten aller Arten und ihre Strukturen zu erforschen. Dabei geht es vor allem darum, durch eine Strukturierung großer Datenmengen Analysen zu ermöglichen. Datenquellen müssen identifiziert und mit anderen kombiniert werden, zudem müssen eventuell unvollständige Datenquellen ergänzt werden.

Der Datenwissenschaftler muss durch eine intensive Analyse und Verknüpfung der Daten neue Fragen aufzeigen können, die zuvor nicht gestellt werden konnten oder zu beantworten waren. Er muss in der Lage sein, ein Problem in der Tiefe zu durchdringen, die Kernfrage aufzudecken und diese in Hypothesen zu verwandeln, die Ansatzpunkte für Analysen bieten. Seine Erkenntnisse muss er sprachlich vermitteln und in ihren potenziellen Auswirkungen auf das Unternehmen einschätzen können. Er benötigt somit auch entsprechende soziale Kompetenzen.

Eine weitere wichtige Fähigkeit ist das assoziative und abstrahierende Denken, wenn es darum gehen soll, fruchtbare Analogien zwischen neu entdeckten Schemata oder Systemen und bereits wissenschaftlich belegten Konzepten zu bilden.

Wo marktspezifische Herausforderungen sich beständig wandeln, können Datenwissenschaftler den Entscheidungsträgern helfen, von ad-hoc-Analysen, die nur bei akutem Bedarf stattfinden, auf eine kontinuierliche und vorausschauende Nutzung der Daten umzusteigen. Sinnvoll ist ihr Einsatz somit vor allem in beratenden Positionen – etwa als rechte Hand von Führungskräften und Produktmanagern –, in denen sie die Bedeutung der Daten für Prozesse, Entscheidungen und Produkte darlegen können.

Neben Informatikkenntnissen benötigt ein Datenwissenschaftler somit auch ein Verständnis der Geschäftsprozesse, das ihm dabei hilft, die relevanten Fragen bei der Analyse seiner Daten zu berücksichtigen.

Ein Datenwissenschaftler kann auf mehreren Ebenen für das Unternehmen von Nutzen sein:

- Seine Arbeit kann neue Handlungsmöglichkeiten eröffnen, etwa indem mit seiner Hilfe bisher nicht sichtbare Muster in Datensätzen aufgedeckt werden, die Rückschlüsse etwa auf das Kundenverhalten, die Marktsituation oder die Geschäftsprozesse erlauben.
- Im Bereich der IT kann ein Datenwissenschaftler Speichercluster oder einzelne Server, die in naher Zukunft ausfallen könnten, frühzeitig identifizieren und so größere Schäden verhindern. Zudem ist er fähig, Sicherheitslücken aufzuspüren.
- In der Entwicklung können die Ergebnisse von Datenauswertungen dazu dienen, Produkte kundenfreundlicher zu gestalten. Ferner lässt sich eine Optimierung der Unternehmenswebsite erzielen, sodass potenzielle Neukunden länger dort verweilen.

Die Kenntnisse, die ein Datenwissenschaftler mitbringen muss, sind vielseitig und fächerübergreifend:

- **IT-Wissen**

Dieses sollte den Hauptpfeiler seiner Kenntnisse bilden. Basis ist hierbei die Fähigkeit zum Verfassen von Programmcodes in mehreren Programmiersprachen. Zudem sind ausgeprägte Kenntnisse über Datenstrukturen sowie Systemarchitektur erforderlich.

- **Mathematik**

Hier ist umfangreiches Wissen über Statistik sowie lineare Algebra vonnöten.

- **BWL und VWL**

Grundlegende Kenntnisse auf diesem Gebiet sind Voraussetzungen für ein Verständnis der Unternehmensstruktur und -führung.

- **Marketing**

Das grundlegende Verständnis sollte hier auch psychologische Kenntnisse im Hinblick auf Analysen mit Kundenbezug einschließen.

### **Unternehmensinterne Einbindung**

Das Wissen des Datenwissenschaftlers sollte sich innerhalb des Unternehmens möglichst ungehindert entfalten können. Seine Eingliederung in die IT-Abteilung erscheint wenig sinnvoll, da dann die direkte Verbindung zur Entscheidungsebene unterbrochen wäre – die aber für eine optimale Nutzung seiner Fähigkeiten maßgeblich ist.

Entsprechend der unklaren Berufsdefinition ist bislang auch die Frage der Einbindung innerhalb des Unternehmens noch nicht abschließend geklärt und muss von Fall zu Fall entschieden werden. Folgende Fragen lassen sich deshalb vermutlich erst im Laufe der Zeit beantworten:

- Ist es sinnvoll, neben traditionell vertretenen Abteilungen wie Controlling, Buchhaltung, Marketing oder Vertrieb eine zusätzlich Abteilung „Data Science“ oder „Big Data“, einzuführen und die neuen Rollen dort anzusiedeln? Auch bei dieser Option müssten die Verbindungen zwischen Datenwissenschaftler und Unternehmen jedoch erst noch ausgestaltet werden.
- Eine weitere Frage wäre daher, wo diese Abteilung im Unternehmen angesiedelt sein soll: innerhalb der IT und unter dem Weisungsrecht des Chief Information Officer, als eine dem Chief Operating Officer unterstellte Abteilung, als „rechte Hand“ des Chief Executive Officer oder als Nebenabteilung des Controllings?

Die zweite Frage ist auch deshalb schwer zu beantworten, weil die Bedarfsträger der Datenwissenschaft nicht ausreichend abgegrenzt werden können. Jede Abteilung innerhalb eines Unternehmens kann potenziell von den Erkenntnissen aus Big-Data-Projekten profitieren – die Marketingabteilung oder die Forschung ebenso wie die Logistik, die Buchhaltung und die Geschäftsführung.

Zu rasche Festlegungen erscheinen deshalb nicht zielführend. Es kommt jedoch darauf an, die Entwicklung von Big Data als Chance zu begreifen und sich bestmöglich darauf vorzubereiten – je eher, desto besser.

***Jede Abteilung kann potenziell von den Erkenntnissen aus Big Data profitieren; die Frage der Einbindung innerhalb des Unternehmens ist noch nicht geklärt.***

## Ihre Ansprechpartner

### Data Analytics

**Bernd Reimer**

Tel.: +49 711 25034-3571  
bernd.reimer@de.pwc.com

**Barbara Lix**

Tel.: +49 203 3175-013  
barbara.lix@de.pwc.com

### IT Consulting

**Marcus Messerschmidt**

Tel.: +49 211 981-4872  
marcus.messerschmidt@de.pwc.com

**Jan Stüben**

Tel.: +49 40 6378-1875  
jan.stueben@de.pwc.com

### Financial Services

**Michael Rasch**

Tel.: +49 40 6378-1806  
m.rasch@de.pwc.com

### Unsere Expertise

Wir sind ein führender Anbieter für analytische IT-Lösungen mit einem Team von mehr als 70 Wirtschaftsinformatikern, Mathematikern, Statistikern und Wirtschaftsingenieuren, die über fundiertes aktuelles Wissen über digitale Wertschöpfungsstrategien sowie über umfassende Branchenkenntnisse verfügen. Darüber hinaus können wir jederzeit auf das Know-how des weltweiten Kompetenznetzwerks von PwC zugreifen. Wir sind erfahren in der Entwicklung der bestmöglichen Lösungen zur Bewältigung Ihrer individuellen Big-Data-Aufgaben. Dafür erstellen wir intelligente Algorithmen zur Datenverarbeitung und implementieren diese vor Ort oder integrieren sie direkt in Ihre bereits bestehenden Prozesse. Mit dem Ziel, Ihre Daten zielgerichtet und effizient nutzbar zu machen – um aus Wissen Mehrwert zu schöpfen.

### Über uns

Unsere Mandanten stehen tagtäglich vor vielfältigen Aufgaben, möchten neue Ideen umsetzen und suchen Rat. Sie erwarten, dass wir sie ganzheitlich betreuen und praxisorientierte Lösungen mit größtmöglichem Nutzen entwickeln. Deshalb setzen wir für jeden Mandanten, ob Global Player, Familienunternehmen oder kommunaler Träger, unser gesamtes Potenzial ein: Erfahrung, Branchenkenntnis, Fachwissen, Qualitätsanspruch, Innovationskraft und die Ressourcen unseres Experten-netzwerks in 158 Ländern. Besonders wichtig ist uns die vertrauensvolle Zusammenarbeit mit unseren Mandanten, denn je besser wir sie kennen und verstehen, umso gezielter können wir sie unterstützen.

PwC. 9.300 engagierte Menschen an 28 Standorten. 1,49 Mrd. Euro Gesamtleistung. Führende Wirtschaftsprüfungs- und Beratungsgesellschaft in Deutschland.

