

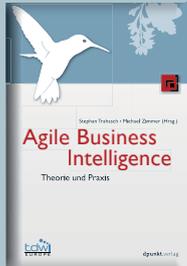
Carsten Dittmar, Carsten Felden,
Ralf Finger, Rolf Scheuch, Lars Tams

Big Data

Ein Überblick

tdwi
EUROPE

dpunkt.verlag



Stephan Trahasch, Michael Zimmer (Hrsg.)

Agile Business Intelligence

Theorie und Praxis

2016, 300 Seiten, Festeinband
€ 59,90 (D) / ISBN 978-3-86490-312-0



Tom Gansor, Andreas Totok

Von der Strategie zum Business Intelligence Competency Center (BICC)

Konzeption – Betrieb – Praxis

2., überarbeitete und aktualisierte Auflage

2015, 446 Seiten, Festeinband
€ 69,90 (D) / ISBN 978-3-86490-043-3



Detlef Apel, Wolfgang Behme, Rüdiger Eberlein,
Christian Merighi

Datenqualität erfolgreich steuern

Praxislösungen für Business-Intelligence-Projekte

3., überarbeitete und erweiterte Auflage

2015, 390 Seiten, Festeinband
€ 69,90 (D) / ISBN 978-3-86490-042-6



Rolf Scheuch, Tom Gansor, Colette Ziller

Master Data Management

Strategie, Organisation, Architektur

2012, 492 Seiten, Festeinband
€ 79,90 (D)
ISBN 978-3-89864-823-3



dpunkt.verlag

Wieblinger Weg 17 · 69123 Heidelberg
fon: 0 62 21 / 14 83 40 · fax: 0 62 21 / 14 83 99
e-mail: bestellung@dpunkt.de · www.dpunkt.de

**Carsten Dittmar · Carsten Felden · Ralf Finger ·
Rolf Scheuch · Lars Tams**

Big Data

Ein Überblick

 **dpunkt.verlag**

Dr. Carsten Dittmar
NTT DATA Deutschland GmbH
Carsten.Dittmar@nttdata.com

Univ.-Prof. Dr. Carsten Felden
Technische Universität Bergakademie Freiberg
Carsten.Felden@bwl.tu-freiberg.de

Dr. Ralf Finger
INFORMATION WORKS Unternehmensberatung & Informationssysteme GmbH
R.Finger@information-works.de

Rolf Scheuch
OPITZ CONSULTING Deutschland GmbH
Rolf.Scheuch@opitz-consulting.com

Lars Tams
OPITZ CONSULTING Deutschland GmbH
Lars.Tams@opitz-consulting.com

TDWI Germany e.V.
c/o SIGS-DATACOM GmbH
Lindlaustraße 2c
53842 Troisdorf
www.tdwi.eu

1. Auflage 2016
Copy Editing: Ursula Zimpfer, Herrenberg
Satz und Herstellung: Frank Heidt
Umschlaggestaltung: Anna Diechtierow, Heidelberg
Druck: www.wp-consult.eu
Artikel-Nr. 077.95739

Copyright © 2016 dpunkt.verlag GmbH
Wieblinger Weg 17
69123 Heidelberg

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten.
Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen. Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autor noch Verlag können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

5 4 3 2 1 0

Vorwort

Mit dem Ziel, umfassender die Anforderungen und Konsequenzen des Big-Data-Themas – und damit gebündelt – darzustellen, initiierte der TDWI Germany e.V. in 2015 den Big-Data-Arbeitskreis. Der TDWI Germany e.V. unterstützt hierbei die Arbeitsgruppe, in der Mitglieder der TDWI-Community, bestehend aus Vertretern von Dienstleistern und Universität, verschiedene Themen rund um Big Data bearbeiten.

Ein Ergebnis dieses Projekts ist der vorliegende Leitfaden, der im Kern grundlegende Aspekte des Big-Data-Themas beinhaltet und somit dem Leser eine Unterstützung bei der eigenen Fassung des Themas bietet. Die Mitglieder der Arbeitsgruppe verfassten themenbezogene Kapitel, die jeweils eine relevante Facette des Big-Data-Themenkomplexes betrachten und dafür konkrete Handlungsempfehlungen geben.

Für die Aufstellung des Big-Data-Leitfadens sei an dieser Stelle der TDWI-Community sowie allen Projektmitgliedern gedankt, die dieses Vorhaben unterstützt haben.

Die Mitglieder der Arbeitsgruppe waren: Dr. Carsten Dittmar (NTT DATA Deutschland GmbH), Univ.-Prof. Dr. Carsten Felden (Technische Universität Bergakademie Freiberg), Dr. Ralf Finger (INFORMATION WORKS Unternehmensberatung & Informationssysteme GmbH), Rolf Scheuch und Lars Tams (beide OPITZ CONSULTING Deutschland GmbH).

Carsten Dittmar, Carsten Felden, Ralf Finger, Rolf Scheuch
und Lars Tams

Köln, Freiberg und Gummersbach im Juni 2016

1 Big Data – Einführung

Carsten Dittmar und Carsten Felden

Big Data ist nicht neu, sondern nur die Fortschreibung bekannter Konzepte in Kombination mit neuen Technologien. Big Data ist insofern keine Revolution, sondern vielmehr eine Stufe in der Evolution zu einer stärker datenorientierten Unternehmensrealität.

Big Data setzt einen analytischen Fokus auf Basis von großen und vielfältig strukturierten Datenmengen aus unterschiedlichen Quellen, die in der Regel unmittelbar nach Entstehung der Daten erzeugt werden und aktuelle Maßnahmen und zukunftsorientierte Handlungen in Bezug setzen. Damit grenzt sich Big Data tendenziell von der gelebten Realität älterer Konzepte wie Business Intelligence ab.

Big Data und Business Intelligence verschmelzen in der breiten Diskussion mehr und mehr unter dem neuen Oberbegriff Analytics.

Big Data ist derzeit als einer der sogenannten Megatrends in aller Munde. Big Data wird häufig stellvertretend für das Phänomen der zunehmenden Digitalisierung aller Lebensbereiche angeführt, die tiefgreifend auch Struktur und Wertschöpfung der Unternehmen verändert und Daten zu einer dominierenden Ressource kürt – dabei ist der Begriff Big Data aber auch immer in seinem Verständnis perspektivgetrieben. In der breiten öffentlichen Diskussion zu Big Data löst die damit verbundene Idee, aus der Vielzahl und Vielfalt der verfügbaren Daten schnell werthaltige Informationsschätze zu heben, sehr viele positive, aber durchaus auch negative Reaktionen aus: Während sich auf der einen Seite ein schier unbegrenztes Spektrum an Anwendungsfeldern mit zum Teil komplett neuen Geschäftsmodellen eröffnet, wird auf der anderen Seite eindringlich auf potenzielle Gefahren hingewiesen, die sich aus der breiten Verknüpfung von Dateninhalten und der damit potenziell vollständigen Transparenz über das Verhalten und die Vorlieben Einzelner ergeben.

Eine intensivere Beschäftigung mit Big Data setzt eine definitorische Einordnung voraus. Die folgenden Abschnitte widmen sich dieser Aufgabe. Darüber hinaus stellen sie sich auch der Herausforderung, Big Data in den Kontext von bekannten Konzepten und Technologien der Business Intelligence einzureihen.

1.1 Definitoriale Einordnung von Big Data

Mit Big Data werden einerseits ganz speziell zu charakterisierende Datenbestände bezeichnet, andererseits aber auch Methoden und Technologien adressiert, die gerade diese speziell zu charakterisierenden Daten verarbeiten können. In einem ersten Schritt der Begriffsfestlegung gilt es insofern, die Besonderheiten von Big Data als Betrachtungsgegenstand aufzugreifen.

In frühen Veröffentlichungen zum Thema wird das Phänomen Big Data häufig anhand einer Negativabgrenzung bezüglich der Leistungsfähigkeit klassischer IT beschrieben – Big Data lagen dann vor, wenn die Möglichkeiten der klassischen Datenhaltung, -verarbeitung und -analyse auf konventioneller Hardware überstiegen wurden [Merv 2011; Plattner 2013].

Weiterentwickelt und geschärft wird diese Definition anhand von charakteristischen Eigenschaften der Big Data. Die erste Eigenschaft bezieht sich auf die exponentiell gestiegene Menge auswertungsrelevanter Daten (Volume). Der Anstieg des Datenvolumens wird u.a. durch den zunehmenden Einsatz von permanent datenerzeugender Sensorik, die zunehmende Protokollierung von IT-Nutzung durch Logdateien sowie den ständigen Zuwachs an Transaktionsdaten begründet. Die steigende Informationsintensität in der Wertschöpfung, aber auch die Informationsintensität in den Produkten/Dienstleistungen treiben die Menge an auswertungsrelevanten Daten aus Sicht der Unternehmen deutlich stärker als die häufig benannte Datenflut in unterschiedlichen Social-Media-Kanälen.

Neben dem wachsenden Datenvolumen steht aber auch die Komplexität und Schnelligkeit der Verknüpfung, Integration und Analyse von Daten unterschiedlichster Struktur und Herkunft im Fokus von Big Data. Als weitere Eigenschaften von Big Data werden daher sowohl die Vielfalt an Datenformaten (Variety) als auch die Geschwindigkeit (Velocity), mit der neue Daten entstehen und auszuwerten sind, angeführt.

Auch wenn die exakten Grenzen hinsichtlich dieser drei Vs nicht eindeutig quantifiziert werden, sind somit die wesentlichen Eigenschaften der Big Data als Gegensatz zu den Small Data benannt, obwohl zuweilen noch zusätzliche Eigenschaften angeführt werden. Konsequenterweise beschreibt dann unter Berücksichtigung der skizzierten Eigenschaften das Phänomen Big Data insofern auch Methoden und Technologien für die hochskalierbare Integration, Speicherung und Analyse polystrukturierter Daten. Dabei bezieht sich der Skalierungsanspruch nicht ausschließlich

auf ein hohes Datenvolumen (Volume), sondern auch auf eine hohe zeitliche Skalierung hinsichtlich hoher Datenaktualität und hoher Performance-Ansprüche bei der Durchführung komplexer Analysen (Velocity) auf Basis von Daten, die in vielfältiger Strukturierung – polystrukturiert – vorliegen können (Variety).

1.2 Big Data und Business Intelligence

Die Integration, Speicherung und Analyse von Daten aus unterschiedlichen Quellen mit dem Ziel eines tieferen Geschäftsverständnisses ist dabei keinesfalls neu und wurde bis dato unter dem Konzept Business Intelligence (BI) subsumiert. Die meisten Unternehmen haben eine langjährige Historie, dispositive Datenbestände aus unterschiedlichen IT-Systemen zusammenzuführen und darauf aufbauend Reporting- und Analyseanwendungen zu erstellen. Die begriffliche Einordnung von BI gibt dabei keinerlei Einschränkungen weder hinsichtlich der Datengewinnung auf Anzahl, Lokalisierung oder Quellenart noch hinsichtlich der Datenauswertung auf den Einsatzzweck und die Schnelligkeit der Analyse und Folgeentscheidungen vor.

Vor diesem Hintergrund stellt sich die Frage, in welcher Beziehung das Phänomen Big Data zum BI-Konzept steht. Hier werden im Moment im Wesentlichen zwei Auffassungen vertreten. Zum einen wird BI als Oberbegriff zu Big Data vor dem Hintergrund der Eigenschaften von Big Data als Betrachtungsgegenstand gesehen. Zum anderen mehrten sich die Stimmen, die Big Data als chronologische nächste Stufe von BI mit erweitertem Fokus verstehen und insofern Big Data als neuen Oberbegriff sehen [Baars/Felden/Gluchowski/Hilbert/Kemper/Olbrich 2014].

Aus der ersten Auffassung heraus stellt Big Data ein spezielles BI-orientiertes Gebiet dar, das den Fokus auf die Skalierbarkeit hinsichtlich der drei Vs legt. Ein Anwendungsfeld wird heute in der Regel zu Big Data gezählt, wenn mindestens zwei der drei Vs zutreffen. Big Data steht damit nicht in Konkurrenz zu klassischen BI-Methoden und BI-Technologien, sondern stellt eher einen speziellen Ausschnitt aus dem breiten Anwendungsfeld von BI dar.

Auch wenn begrifflich mit BI ein breiter Anwendungsbereich adressiert wird, zeigt sich jedoch in der Praxis häufig ein ernüchterndes Bild. Hier wird de facto unter BI das klassische vergangenheitsorientierte Reporting verstanden, das auf den strukturierten Daten des Data Warehouse

aufsetzt. Dieses gelebte Begriffsverständnis führt zu der Auffassung, dass die meist auf die Fokussierung von zukünftigen Handlungen und Aktionen ausgerichteten Big-Data-Anwendungen nicht als Unterbegriff von BI einzuordnen sind, sondern vielmehr Big Data als chronologische nächste Stufe von BI aufzufassen ist. Damit wäre Big Data eher als übergeordnet zu klassifizieren.

Der Hype um Big Data stellt insofern aus Sicht von BI-Professionals Fluch und Segen zugleich dar. Auf der einen Seite führt die breite Diskussion dazu, dass auch außerhalb der IT über innovative Ideen und neue Geschäftsmodelle nachgedacht wird, die auf dem Rohstoff des 21. Jahrhunderts – den Daten – aufsetzen. Dadurch wird die Bedeutung und Bewertung des Assets Daten noch einmal deutlich verstärkt. Daten sind ein technisch und methodisch vollständig nutzbarer, wichtiger Produktionsfaktor. Daten treiben und verändern Geschäftsprozesse und Geschäftsmodelle und lassen das *data-driven enterprise* Realität werden. Zudem werden analog zum einschränkenden BI-Verständnis in der Praxis bisher vernachlässigte BI-Themen wieder neu belebt (z.B. Data Mining, Visualisierung, Endbenutzer-Autonomie). Diesen positiven Aspekten steht gegenüber, dass viele Diskussionen um Big Data losgelöst von der bestehenden BI-Erfahrung geführt werden. Schließlich sind die Herausforderungen aus der Small-Data-Welt (z.B. mangelnde Datenqualität, inadäquate Organisationsstrukturen, unzureichende strategische BI-Ausrichtung) nicht endgültig gelöst und sie sind auch für Big Data nicht nur weiterhin relevant, sondern treten auch wieder deutlicher hervor. Insofern bringt Big Data viele Ansätze in die Unternehmensrealität ein, die konträr zum Anspruch an einen hohen Integrationsgrad der BI stehen.

1.3 Führen BI und Big Data zu Analytics?

Folgt man der historischen Entwicklung im Umfeld dispositiver Informationssysteme, so scheint regelmäßig (losgelöst von der Tatsache, ob BI und Big Data als Themenstellungen oder Konzepte mit einer definierten Unter- oder Oberordnung definiert werden) eine neue Terminologie eingeführt zu werden, die die Diskussion um Management-Information-Systeme, Decision-Support-Systeme, Executive-Information-Systeme, Executive-Support-Systeme, Data Warehousing oder BI neu zusammenbringt. Hier scheint der Begriff Analytics im Moment ein geeigneter Kandidat zu sein, der die traditionellen Aspekte der BI und die neuen Aspekte

von Big Data miteinander verbinden kann. Unter Business Analytics wird die kontinuierliche Erforschung und Untersuchung von vergangenheitsorientierten Geschäftsdaten mittels analytischer Methoden verstanden, um daraus Erkenntnisse sowohl über die abgelaufene als auch die kommende Geschäftstätigkeit zu erlangen, die wiederum in die einzelnen zu planenden Geschäftsaktivitäten einfließen. Die Kontinuität entsteht durch die regelmäßige Ausführung von Analysetätigkeiten, die sich entsprechend in einer Ablauforganisation implementieren lassen. Iterativ sind derartige Aktivitäten, weil im Analyseprozess häufig eher neue Fragen als abschließende Antworten entstehen, die letztlich zu untersuchen sind. So kann die bisherige Geschäftstätigkeit nachvollzogen werden, um Verbesserungen bei neuen Handlungen zu ermöglichen. Dies entspricht dem integrativen Gedanken der Business Intelligence, allerdings verbunden mit dem intensiveren Analyseverständnis auf und aus den Big Data.

Der Trend zu Analytics wird anhalten, denn in vielen Themengebieten, wie beispielsweise der Industrie 4.0, erhält die Analysekomponente eine zentrale Bedeutung. Jedoch muss ein Schritt nach dem anderen gegangen werden – und das bedeutet, mit Big Data zurechtzukommen. Wir fokussieren weiterhin auf den Begriff Big Data, da es die Kernfrage ist, welche Daten zu erfassen sind und welche Ergebnisse zu erzielen sind. Und dabei steht das Big Picture im Vordergrund anstelle der *Out-of-the-Box*-Insellösung.

1.4 Literatur

[Baars/Felden/Gluchowski/Hilbert/Kemper/Olbrich 2014] Baars, H.; Felden, C.; Gluchowski, P.; Hilbert, A.; Kempert, H.-G.; Olbrich, S.: Shaping the Next Incarnation of Business Intelligence. In: Business & Information Systems Engineering 01/2014. DOI 10.1007/s12599-013-0307-z.

[Merv 2011] Merv, A.: It's going mainstream, and it's your next opportunity. In: Teradata Magazine Online, Q1, 2011, <http://www.teradata-magazine.com/v11n01/Features/Big-Data/> (letzter Abruf: 09.05.2016).

[Plattner 2013] Plattner, H.: Big Data. In: Kurbel, K.; Becker, J.; Gronau, N.; Sinz, E.; Suhl, L. (Hrsg.): Enzyklopädie der Wirtschaftsinformatik – Online Lexikon. <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Datenmanagement/Datenmanagement--Konzepte-des/Big-Data/> (letzter Abruf 09.05.2016).

2 Big-Data-Anwendungsfallklassifikation zur Projekt- und Managementunterstützung

Carsten Felden

Das Konzept der Referenzmodelle bietet Unterstützung bei der Themenidentifikation im diffusen Umfeld der Big Data.

Durch deren lobbyistische Prägungen sind viele existierende Referenzmodelle zur ausschließlichen Nutzung nicht geeignet, sondern dienen lediglich zur spezifischen Visualisierung.

Open Innovation unterstützt die integrierte und damit multiperspektivische Betrachtung mehrerer Referenzmodelle.

Mit der Beschreibung des Big-Data-Phänomens und damit der Abgrenzung des Themas selbst besteht die Grundlage zur Nutzung der Referenzfälle als Blaupause.

Open Innovation, als hier genutzter Ansatz zur Projekt- und Managementunterstützung, entstand als Teil einer in vielen Bereichen nachweisbaren Bewegung der Öffnung von Inhalten zur allgemeinen Nutzung (Open Data, Open Access, Open Science). Es beschreibt den Ansatz, Inhalte der Allgemeinheit zu öffnen und nutzerseitig sich nicht nur unternehmensintern durch Ideen zu Innovationen führen zu lassen, sondern auch von außen, beispielsweise durch den Markt, den Innovationsprozess inspirieren zu lassen. Innovationen wirken in Architekturen und Systemen, die durch das jeweilige Geschäftsmodell vorgegeben sind [Chesbrough 2006]. Neuartige Technologien wirken dabei nicht als direkter Innovationstreiber. Technologie wirkt zum Beispiel bei Big Data eher als Mittel zum Zweck, damit ein Unternehmen sein Geschäftsmodell im Rahmen unternehmensinterner und -externer Aktivitäten anpassen oder unterstützen kann

Um eine Projekt- und Managementunterstützung im Sinne der Open Innovation durch die Big-Data-Projektbeispiele zu bieten, wird Vergleichbarkeit durch den Aufbau eines Klassifikationsschemas von Big-Data-Anwendungsfällen hergestellt. Dies erfolgt unter der Maßgabe, dass im Rahmen des jeweiligen Geschäftsmodells agiert wird bzw. Geschäftsmodellanpassungen notwendig sind. Hierzu stellen wir einen morphologischen Kasten [Zwicky 1959] zur Anwendungsgruppierung für Big Data vor, da dieser mehrere Strukturierungsdimensionen abbilden kann.

Wesensgebendes Merkmal eines morphologischen Kastens ist eine mehrdimensionale Matrix, die nach verschiedenen Gesichtspunkten die objektive Erfassung auch komplexer Themengebiete mittels Dimensionen ermöglicht. Im gewählten Themenfeld Big Data bietet der morphologische Kasten Unterstützung für Big-Data-Innovationen durch eine Einordnung bestehender und den Vergleich mit eigenen Anwendungsfällen. Wir detaillieren in diesem Beitrag einige ausgewählte existierende Klassifikationsansätze von Big-Data-Anwendungsfällen und führen diese Ansätze im morphologischen Kasten zusammen.

Im folgenden Abschnitt werden die existierenden Kategorisierungen von Big-Data-Anwendungsfällen vorgestellt. Initial der fachlich orientierte Use-Case-Cluster [Dittmar 2016, S. 55 ff.], da dieser die Ausgangsbasis darstellt. Des Weiteren werden die Kategorisierung nach Bitkom und der Ansatz nach Carter erläutert. Darauf aufbauend wird eine Detaillierung der Perspektiven der bestehenden Ansätze vorgenommen und abschließend in Abschnitt 2.3 im morphologischen Kasten zusammengeführt.

2.1 Existierende Klassifikationen von Big-Data-Anwendungsfällen

»Der Nutzen von Big-Data-Analysen lässt sich in einigen Funktionsbereichen besonders belegen« [Dorschel 2015, S. 104]. Nur ist das Wiedererkennen in einem Anwendungsszenario [Querg 2010] nicht immer nachvollziehbar, sodass Anwendungscluster einen Ordnungsrahmen bieten.

2.1.1 Use-Case-Cluster nach Dittmar

Anwendungscluster unterstützen die Identifikation und Analyse von industriespezifischen Big Data Use Cases [Dittmar 2016]. Zur Analyse dieser Anwendungsszenarios und als Basis für den hier benannten morphologischen Kasten bietet Dittmar ein Vier-Quadranten-Schema, das

zum einen den Analysegegenstand und zum anderen den Geschäftsmodellbezug beschreibt. Die vertikale Achse wird hierbei in einen internen Analysegegenstand (*Unternehmen*) und einen externen Analysegegenstand (*Markt*) unterteilt. Die horizontale Achse wird in *bestehendes Geschäftsmodell* und *neues Geschäftsmodell* differenziert. Die Betrachtung des Geschäftsmodellbezuges unterstellt implizit, dass der beschriebene Anwendungsfall ökonomischer Natur ist und es um ein Unternehmen oder dessen Branche [Engelhard 2015] geht. Auch wenn die Kategorien bestehendes/neues Geschäftsmodell denen der Open Innovation ähneln, geht es nicht um die interne/externe Inspiration, sondern um die nach innen gerichtete oder marktorientierte Wirkung der Aktivitäten. Diesen Ansatz begreifen wir aufgrund seiner fachlichen Ausrichtung als Ausgangspunkt zu Konzeption des morphologischen Kastens, indem er die ersten beiden Dimensionen enthält.

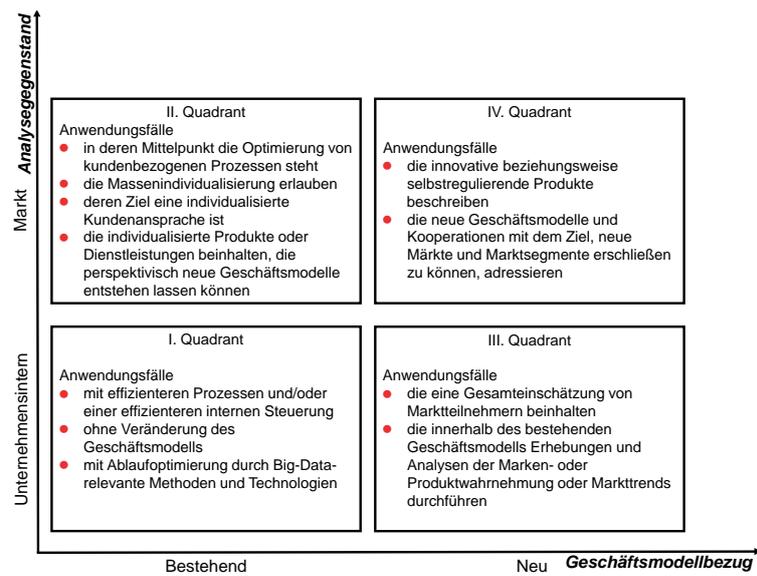


Abb. 2-1 Use-Case-Cluster für Big Data [Dittmar 2016]

Zwischen den Quadranten ergibt sich ein fließender Übergang. Die innerhalb der Abbildung benannten Merkmale sind eher abstrakt. Sie lassen anhand dieser vier Kategorien noch kein ausreichendes Urteil über eine Vergleichbarkeit von Anwendungsfällen zu. Somit ist auch die Zuordnung eines Anwendungsfalls selbst zu einem Quadranten zunächst noch nicht ausreichend, um eine Übertragungsfähigkeit auf einen eigenen Fall vornehmen zu können.

2.1.2 Kategorisierung des Bitkom-Arbeitskreises

Nach dem Bitkom-Arbeitskreis für Big Data lassen sich die heutigen Geschäftsmodelle und Business Cases in der Datenwirtschaft ebenfalls in vier Kategorien einteilen. Dafür sind die Dimensionen Geschäftsmodellbezug (Business) und Datenquelle gewählt worden.

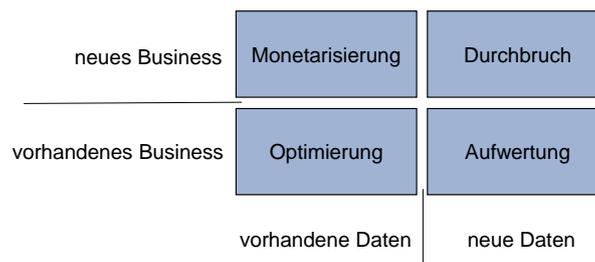


Abb. 2-2 Big-Data-Kategorisierung [Bitkom 2015]

Der Geschäftsmodellbezug wird dabei in *neues Business* und *vorhandenes Business* unterteilt. Die Dimension *Datenquelle* bezieht sich in dieser Kategorisierung auf *vorhandene Daten* und *neue Daten*. Die Anwendungsfälle werden in diesem Modell über die vier Klassen in Optimierung, Aufwertung, Monetarisierung und Durchbruch kategorisiert.

Die Klasse **Optimierung** beschreibt Big-Data-Anwendungsfälle, die einen großen Mehrwert durch Optimierung bestehender Geschäftsprozesse mithilfe der Auswertung bereits existierender Datenbestände erzeugt. Diese Klasse wird deshalb als Einstiegsklasse betrachtet, da hier eine Nutzung der bereits bestehenden Infrastruktur stattfindet. Die Analyse von existierenden und unternehmenseigenen Datenbeständen kann bereits durch die Optimierung von Prozessen einen Mehrwert generieren

[Bitkom 2013, S. 17]. In den Bereich der **Monetarisierung** lassen sich Anwendungsfälle einordnen, die durch die Nutzung vorhandener Datensätze neue Geschäftsmodelle bzw. Produkte erzeugen. Hierbei ist jedoch der Datenschutz zu beachten und es ist fraglich, ob die Nutzung der Daten zu diesem Zweck rechtlich zulässig ist. Es ist zum Beispiel möglich, anonymisierte Standortdaten von Telefonnutzern für ortsbezogene Werbung zu verwenden [Bitkom 2013, S. 17]. Die **Aufwertung** soll all jene Fälle kategorisieren, die bestehende Geschäftsmodelle und Dienstleistungen durch das Erschließen neuer Daten optimieren. Hier sind beispielsweise Reiseunternehmen zu finden, die detaillierte Wetterprognosen für Marketingaktivitäten und bessere Auslastung der Reiseziele integrieren [Bitkom 2013, S. 18]. Die vierte Klasse, **Durchbruch**, bezeichnet die Entwicklung neuer Geschäftsmodelle für die Datenwirtschaft. Neue Produkte und Dienstleistungen lassen sich auf der Basis der Sammlung und Digitalisierung neuer Datenbestände erschaffen. Als Beispiel ist hier die Kartierung von Google durch den Service *Streetview* zu nennen [Bitkom 2013, S. 18].

Während die Betrachtung des Geschäftsmodells redundant ist, erscheint der Einbezug der Datenherkunft als eine weitere relevante Beschreibungsdimension von Big-Data-Anwendungsfällen, da durch die Ergänzung das Ziel der Entscheidungsunterstützung bedient wird.

2.1.3 Carters Big Data Analytics

Häufig werden auch die drei zuvor genannten Vs zur Big-Data-Charakterisierung genutzt. Das Kriterium *Data Volume* skaliert zwar in der Regel unternehmensindividuell, jedoch ist es eine Grundlage von Dimensionierungsbetrachtungen bezüglich der Architektur (Netzwerk- oder Serverkapazitäten). Im Weiteren fokussieren wir auf die Dimensionen *Data Velocity* und *Data Variety*.

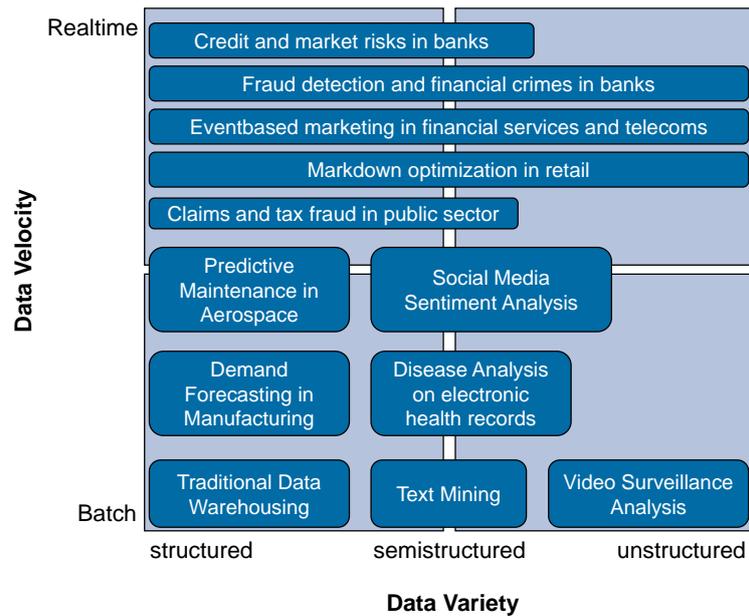


Abb. 2-3 Data-Velocity- und Data-Variety-Kategorisierung [Carter 2011]

Anders als zu den vorherigen Ansätzen sind hier keine Felder benannt, weshalb die Einteilung anhand dargestellter Kategorisierungen der Beispielanwendungsfälle vorgenommen wird. Betrugsdetektion im Bankensektor in Echtzeit ist ein dargestelltes Beispiel. Anwendungsfälle in dieser Kategorie müssen das komplette Spektrum der Datenvielfalt bewältigen und finden in Echtzeitanalysen statt [Carter 2011, S. 10]. Bei Echtzeitpreisanpassungen werden im Handel Informationen genutzt, um die Nachfrage der Kunden vorherzusagen und darüber die Marge der Produkte durch automatisierte Preisanpassungen zu steigern. Auch bei dieser Kategorie von Anwendungsfällen gilt es, eine hohe Datenvielfalt, von strukturierten bis hin zu unstrukturierten Daten, zu bewältigen [Carter 2011, S. 10]. Für Krankheitsanalysen durch Daten aus digitalen Patientenakten, wenn Krankheitsverläufe und Informationen zu Patienten digital gespeichert werden, entsteht ein Potenzial für Big-Data-Auswertungen.

Durch das Erfassen von Krankheitsverläufen und deren Behandlung lassen sich schnellere Diagnosen stellen sowie Analogien zwischen den Patienten erkennen, um Behandlungen zu verbessern [Carter 2011, S. 10].

2.2 Erweiterung der Dimensionen zur Kategorisierung von Big-Data-Anwendungsfällen

Die beschriebene Kategorisierung der Anwendungsfälle gemäß dem Use-Case-Cluster ist eher grob und ermöglicht noch keine umfassende Vergleichbarkeit der Anwendungsfälle eines Quadranten. Zudem ist die trennscharfe Zuordnung zwischen einer Veränderung bestehender Geschäftsmodelle und neuen Geschäftsmodellen schwierig und es müssen Veränderungen von Geschäftsmodellen in Betracht gezogen werden. Um die Vergleichbarkeit zu erhöhen, detaillieren wir zunächst den Analysegegenstand über intern/extern hinaus. Big-Data-Anwendungsfälle können dabei eine oder mehrere der später im morphologischen Kasten dargestellten Eigenschaften betreffen und ändern. Darauf aufbauend zeigen wir typische Kombinationen von gemeinsam von Big-Data-Anwendungsfällen betroffenen Eigenschaften des Analysegegenstands. Diese ermöglichen es, den eigenen Anwendungsfall auf weitere betroffene Eigenschaften des Geschäftsmodells hin zu untersuchen.

2.2.1 Detaillierung des Analysegegenstands

Die **externe Sicht** diskutiert Chancen und Risiken auf dem Markt, die zu den strategischen Erfolgsfaktoren führen [Mintzberg 2013, S. 36 f.]. Ein **Kundensegment** innerhalb des Marktes definiert Gruppen von Kunden eines Unternehmens oder einer Organisation mit jeweils gemeinsamen Bedürfnissen, Präferenzen [Kirchgeorg/Maier/Piekenbrock 2015] oder ähnlichem Verhalten, die das Unternehmen gezielt erreichen und bedienen will [Osterwalder/Pigneur 2010, S. 20]. Das Unternehmen muss die Frage beantworten, »für wen schaffen wir Wert und wer ist unser wichtigster Kunde?« [Osterwalder/Pigneur 2010, S. 21].

Bei der Auswahlentscheidung von Kunden kommt es auf den wahrgenommenen Wert eines Produktes durch den Kunden an. Aus Sicht des Unternehmens beschreibt das **Nutzenversprechen** ein Bündel von Produkten und Services, die einen Wert für ein spezifisches Kundensegment darstellen. Welchen Mehrwert ein Produkt für den Kunden hat, ist abhängig von der Neuheit und der Leistung des Produktes, der Anpassungsfähigkeit

an den Kunden, der Gestaltung und der Marke des Produktes sowie dem Preis und der Verlässlichkeit [Osterwalder/Pigneur 2010, S. 23 ff.].

Zwischen der reinen externen und der internen Sicht ist die **Schnittstelle** des Unternehmens zu seinen verschiedenen bereits genannten Märkten zu sehen, die ebenfalls von Big-Data-Anwendungsfällen adressiert werden können. **Vertriebskanäle** zeigen auf, wie das Unternehmen mit seinen Kunden kommuniziert und diese erreicht, um das Nutzenversprechen des Produktes zu vermitteln. Der Vertrieb unterstützt aber auch den Kunden vor dem Kauf mit Beratung/Informationen [Osterwalder/Pigneur 2010, S. 26 f.].

Der Baustein **Kundenbeziehung** umfasst die Arten der Beziehung zwischen Unternehmen und dem jeweiligen Kunden. Das Unternehmen muss innerhalb des Geschäftsmodells deutlich machen, welche Art von Beziehungen es zum Kunden pflegt [Osterwalder/Pigneur 2010, S. 28]. Mögliche Arten der Kundenbeziehung zwischen Kunden und Unternehmen können persönliche Betreuung, Selbstbedienung, Mitgestaltung und Kommunikation sein [Osterwalder/Pigneur 2010, S. 29].

Eine weitere wichtige Schnittstellenperspektive von Big-Data-Anwendungsfällen bilden **Schlüsselpartnerschaften**. Sie bilden das Netzwerk von Lieferanten und Geschäftspartnern, das für das Funktionieren des Geschäftsmodells unerlässlich ist [Osterwalder/Pigneur 2010, S. 38]. Die Motivation für Partnerschaften mit anderen Unternehmen können die Optimierung von Skaleneffekten sein, die Reduzierung von Unsicherheiten und Risiken sowie die Akquirierung von speziellen Ressourcen und Fähigkeiten [Osterwalder/Pigneur 2010, S. 39].

Die **interne Sicht** beschreibt die Quelle für Stärken und Schwächen eines Unternehmens, die zu den organisatorischen Fähigkeiten führen [Mintzberg 2013, S. 37]. **Kernressourcen** werden aus den wichtigsten Anlagen, Vermögenswerten und Ressourcen gebildet, die dafür sorgen, dass ein Geschäftsmodell funktionsfähig ist. Jedes Geschäftsmodell benötigt Schlüsselressourcen (physische, finanzielle, geistige oder Humanressourcen), damit es einen Wert schaffen kann sowie auf dem Markt agieren und die Kundenbeziehung aufrechterhalten kann. Aus dem Zusammenspiel der einzelnen Ressourcen entstehen dann die sogenannten organisatorischen Fähigkeiten [Grant/Nippa 2006, S. 183, S. 191], insbesondere aufgrund der Koordination unterschiedlicher Fertigungsfähigkeiten und der Integration vielfältiger Technologietrends [Grant/Nippa 2006, S. 193].

Kern- oder Schlüsselaktivitäten beschreiben die wichtigsten Tätigkeiten in einem Unternehmen, die notwendig sind, damit ein Geschäftsmodell funktioniert. Diese stehen in enger Verbindung mit den Schlüsselressourcen, denn auch die Kernaktivitäten sorgen dafür, dass Wert geschaffen wird. Gleichzeitig bewirken sie, dass ein Unternehmen auf dem Markt agiert und Kundenbeziehungen aufrechterhält [Osterwalder/Pigneur 2010, S. 37].

Der letzte Baustein **Kostenstruktur** betrachtet alle Kosten, die durch das Betreiben eines Geschäftsmodells anfallen. Sowohl das Schaffen als auch die Lieferung eines Kundennutzens, das Aufrechterhalten der Kundenbeziehung und das Generieren von Wert für den Kunden verursacht Kosten. Diese Kosten lassen sich jedoch relativ einfach kalkulieren, wenn die Bausteine Schlüsselressourcen, Schlüsselaktivitäten und Schlüsselpartnerschaften zuvor definiert wurden [Osterwalder/Pigneur 2010, S. 41].

2.2.2 Gruppen betroffener Eigenschaften von Geschäftsmodellen

Das Rahmenkriterium für die Klasse **Prozesseffizienzsteigerung** ist interner Natur, da es sich in den Funktionsbereichen eines Unternehmens befindet. Die Quelle für hier nutzbare Daten sind meist Maschinensensoren, Produktsensoren sowie Sensoren an Lkws, Wareneingängen und -ausgängen der Lieferanten und Kunden [Bitkom 2012, S. 38]. Mithilfe der Herkunft der Daten lassen sich der Klasse Prozessoptimierung die Funktionsbereiche Fertigung/Produktion sowie Vertrieb und Distribution als klassenspezifisches Kriterium zuordnen. Zusätzlich wird zur Klassifizierung von Anwendungsszenarios innerhalb der Prozesseffizienzsteigerung das zweite Rahmenkriterium **Struktur des Unternehmens** verwendet. Ohne eine Anpassung des Geschäftsmodells lassen sich die folgenden Bausteine als klassenspezifische Kriterien zuordnen: Kernaktivitäten, Kernressourcen, Kernpartnerschaften, Kostenstruktur und das Nutzenversprechen. Grund hierfür ist die Korrelation der Bausteine mit den Funktionsbereichen Fertigung/Produktion sowie Vertrieb und Distribution. Bei der Klassifizierung der Szenarios hat sich gezeigt, dass in Verbindung mit der Quelle der Daten immer einer oder mehrere Bausteine des Geschäftsmodells betroffen sind.

Innerhalb der Klasse **Massenindividualisierung** stammen die Themenkomplexe aus kundenbezogenen Prozessen, beispielsweise unternehmenseigenen Marktanalysen. Somit lassen sich diese Datenquellen den Funktionsbereichen eines Unternehmens zuordnen, die direkt mit dem

Kunden und den kundenbezogenen Prozessen in Verbindungen stehen. Daraus folgt, dass mithilfe der Einordnung der Klasse Massenindividualisierung die Funktionsbereiche Marketing sowie Vertrieb und Distribution als klassenspezifisches Kriterium zuzuordnen sind. Im Bereich Distribution und Vertrieb stehen vor allem die der Produktion nachgelagerten Prozesse im Vordergrund. Damit einhergehend lassen sich über die Struktur wiederum zusätzliche klassenspezifische Kriterien ermitteln. Aufgrund der kundenbezogenen Prozesse und mithilfe der Herkunft der Themenkomplexe findet vor allem eine Einordnung in die Bausteine Kundenbeziehung, Kundenversprechen und Vertriebskanal dieser Klasse statt. Abhängig vom Grad der Veränderungen der Bausteine kommt es in dieser Klasse zu einem teilweisen Austausch oder einer Verbesserung des Geschäftsmodells oder auch zu einem neuen Geschäftsmodell.

Marktanalyse und Marktbewertung bilden die Grundlage für Berichte über Märkte, Segmente, Wettbewerbsanalysen, aber auch für die Entwicklung von Produkt- und Marktstrategien [Bitkom 2012, S. 35 ff.]. Die in diesem Zusammenhang genutzten strukturierten und unstrukturierten Daten stammen in der Regel aus externen Quellen [Bitkom 2012, S. 32]. Somit können mithilfe der Herkunft der Daten die Kundenbedürfnisse und Kundenpräferenzen als klassenspezifisches Kriterium für die Marktanalyse und Bewertung zugeordnet werden. Damit einhergehend haben sich ebenfalls einzelne Bausteine als besonders wichtig für die Klasse Marktanalyse und Bewertung herausgestellt. Bei der Analyse von industriespezifischen Anwendungsszenarios haben sich bezüglich der Struktur des Unternehmens die Bausteine Kundensegmente, Ertragsströme, Vertriebskanäle und Nutzenversprechen als relevant herausgestellt. Diese Bausteine stehen in einem engen Zusammenhang mit den Bedürfnissen und Präferenzen des Kunden. Folglich lassen sich diese Bausteine der Klasse Marktanalyse und Bewertung zuordnen.

Die Inspiration für **intelligente/neue Produkte** liegt oftmals unternehmensextern und betrifft vor allem die Produkte des Unternehmens und die Positionierung auf dem Markt. Quellen der Daten hierfür sind beispielsweise soziale Medien, Ideenmarktplätze, Patentdatenbanken sowie Marktforschungsinstitute [Bitkom 2012, S. 37]. Somit lassen sich die Notwendigkeit von Produktinnovation, Produktverbesserung sowie Produkteliminierung als klassenspezifische Kriterien ansehen. Gleichzeitig sind diese Themen in Verbindung mit Nutzenversprechen, Kernpartnerschaften, Kernaktivitäten und Kernressourcen von besonderer

Bedeutung. Diese Bereiche sind für die Entstehung von Innovationen entscheidend und in Bezug auf die Struktur eines Unternehmens kann es zu einer teilweisen oder vollkommenen Veränderung des Geschäftsmodells aufgrund von Produktvarianten oder -verbesserungen kommen.

2.2.3 Veränderungsgrad Geschäftsmodell

In einem zweiten Schritt soll nun der **Grad der Veränderung** des Geschäftsmodells bestimmt werden. Mitchell und Coles unterscheiden drei Arten von Geschäftsmodell Anpassungen. Zwei davon beziehen sich auf das bestehende Geschäftsmodell: einerseits *Geschäftsmodellverbesserungen* und andererseits *Geschäftsmodell austausch*. Die dritte bezieht sich auf ein neues Geschäftsmodell, das als **Geschäftsmodellinnovation** bezeichnet wird. Wird nur ein Element eines Geschäftsmodells mit dem Ziel, Leistung, Profit und Verkäufe des Unternehmens zu steigern, verändert, so spricht man von einer **Geschäftsmodellverbesserung**. Werden darüber hinaus mehrere Elemente des Geschäftsmodells gegenüber dem Geschäftsmodell der Mitbewerber verbessert, spricht man von einem **Austausch des Geschäftsmodells** [Mitchell/Coles 2003, S. 16].

2.3 Morphologischer Kasten

Innerhalb eines Unternehmens sind die Bereiche mit größtem wirtschaftlichem Nutzen für Big Data: Forschung und Entwicklung, Service und Support, Marketing und Vertrieb, Produktion, Distribution, Logistik, Finanz und Risikocontrolling sowie Administration und Organisation [Bitkom 2012, S. 34]. Unterstützung bei der Identifikation eigener Entwicklungsbereiche für eigene Big-Data-Anwendungsfälle innerhalb eines Unternehmens bietet die Stärken-Schwächen-Chancen-Risiken-Analyse (SWOT) von Porter [Grant/Nippa 2006, S. 35].

Die im nachfolgenden morphologischen Kasten dargestellten Rahmenkriterien unterstützen Entscheidungsträger, mögliche Anwendungsszenarios im Unternehmen zu identifizieren und zu kategorisieren. Dies ermöglicht die Nutzung beschriebener Projekte als Blaupause zur Umsetzung eigener Big-Data-Anwendungsfälle.

Ausprägung	
Dimension	
Analyseperspektive	intern extern
Analysegegenstand	Schnittstelle Kundenbeziehung Vertriebskanäle Schlüssel-partnerschaften Kernressourcen
	Kostenstruktur Kern- und Schlüsselaktivitäten Kernressourcen
Gruppierung	Prozesseffizienzsteigerung Massenindividualisierung Intelligente / neue Produkte Marktanalyse / Marktbewertung
Geschäftsmodellveränderung	Geschäftsmodellverbesserung Geschäftsmodelltausch Geschäftsmodellinnovation
Datenquelle	Intern Intern/Extern Extern
Velocity	zeitversetzt Echtzeit
Variety	strukturiert semi-strukturiert unstrukturiert
Volume	wenig mittel viel sehr viel

Tab. 2-1 Morphologischer Kasten

2.4 Literatur

[Bitkom 2012] Bitkom-Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (Hrsg.): Big Data im Praxiseinsatz. Szenarien, Beispiel, Effekte. 2012.

[Bitkom 2013] Bitkom-Arbeitskreis Big Data: Management von Big-Data-Projekten. Bitkom 2013.

[Bitkom 2015] Bitkom-Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (Hrsg.): Leitlinien für den Big Data Einsatz. Chancen und Verantwortung. 2015.

[Carter 2011] Carter, P.: Big Data Analytics: Future Architectures, Skills and Roadmaps for the CIO, September 2011.

[Chesbrough 2006] Chesbrough, H.: Open Innovation. Harvard Business School Press. Boston, Massachusetts, 2006.

[Dittmar 2016] Dittmar, C.: Die nächste Evolutionsstufe von AIS: Big Data. Erweiterung klassischer BI-Architekturen mit neuen Big Data Technologien. In: Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Business-Intelligence-Technologien und -Anwendungen. 5., vollständig überarbeitete Auflage. Unter Mitarbeit von Peter Chamoni, Springer-Verlag, 2016, S. 55-65.

[Dorschel 2015] Dorschel, J.: Praxishandbuch Big Data. Wirtschaft – Recht – Technik. Gabler (SpringerLink : Bücher), Wiesbaden, 2015.

[Engelhard 2015] Engelhard, J.: Stichwort: Branche. Hrsg. v. Springer Gabler Verlag. Gabler Wirtschaftslexikon. Online verfügbar unter <http://wirtschaftslexikon.gabler.de/Archiv/1210/branche-v9.html> (letzter Abruf: 29.01.2016).

[Grant/Nippa 2006] Grant, R. M.; Nippa, M.: Strategisches Management. Analyse, Entwicklung und Implementierung von Unternehmensstrategien. 5., aktualisierte Auflage [der engl. Ausgabe, 1. dt.-sprachige Ausgabe], Pearson Studium (Wirtschaft), München, 2006.

[Kirchgeorg/Maier/Pieckenbrock 2015] Kirchgeorg, M.; Maier, G.; Pieckenbrock, D.: Stichwort: Bedürfnis. Hrsg. vom Springer Gabler Verlag. Gabler Wirtschaftslexikon. Online verfügbar unter <http://wirtschaftslexikon.gabler.de/Archiv/57500/beduerfnis-v7.html> (letzter Abruf 29.01.2016).

[Mintzberg 2013] Mintzberg, H.: Rise and fall of strategic planning. Free Press, 2013.

[Mitchell/Coles 2003] Mitchell, D.; Coles, C.: The ultimate competitive advantage of continuing business model innovation. In: Journal of Business Strategy 24 (5), 2003, S. 15-21.

[Osterwalder/Pigneur 2010] Osterwalder, A.; Pigneur, Y.: Business model generation. A handbook for visionaries, game changers, and challengers. Wiley, Hoboken, NJ, 2010.

[Querg 2010] Querg, D.: Use Case, Anwendungsfall, Anwendungsszenario – wo ist da der Unterschied? Fachhochschule Aachen. <http://campusphere.de/experience/2010/12/01/use-case-anwendungsfall-anwendungsszenario-wo-ist-da-der-unterschied/> (letzter Abruf: 29.01.2016).

[Zwicky 1959] Zwicky, F.: Morphologische Forschung. Winterthur, 1959, Neuauflage Baeschlin, Glarus, 1989

3 Big-Data-Methodik und Vorgehen

Rolf Scheuch

Big-Data-Methoden und Vorgehen verändern sich im Lebenszyklus.
Das explorative Vorgehen in Data Labs ist nur der Startpunkt.
Operativer Erfolg bei Big Data erfordert Einbeziehung der Unternehmens-IT.
Nur ein Produktmanagement über den gesamten Lebenszyklus ermöglicht Innovation.

3.1 Einführung

Im diesem Kapitel beschränken wir uns auf Methoden und Vorgehen für die Projektdurchführung und vernachlässigen die Betrachtung von spezifischen Methoden und Vorgehensweisen bei der technischen Umsetzung bzw. der Ausgestaltung von Big Data. Generell umfassen Big-Data-Vorhaben eine große Spannweite an Projektarten und somit auch sinnvollen Vorgehensweisen. Wir werden im Folgenden unter einem *Big-Data-Projekt* sämtliche *projektähnliche* Big-Data-Vorhaben verstehen, die als Projekt oder Teilprojekt auf den in Kapitel 4 erwähnten Big-Data-Technologien bzw. Architekturen basieren. Explizit betrachten wir keine übergreifenden Initiativen, wie etwa eine Big-Data-Initiative, ein Big-Data-Programm, ein Big Data Competency Center oder die Big Data Governance.

3.2 Methodik und Vorgehen

Die *klassischen* Vorgehensmodelle und Methoden für die Projektdurchführung [Ferstl/Sinz 2008, S. 487 ff.] unterstützen die Lenkung bei der Entwicklung von IT-basierten Lösungen. Hierbei ist es die Aufgabe des *klassischen* Projektmanagements (PM), über ein geordnetes Vorgehen das im Vorfeld formulierte Sachziel mit den definierten Qualitätsanforderungen zu erreichen. Hierbei spielt das formale Ziel des PM zur Einhaltung eines gesetzten Budgetrahmens, der Laufzeit wie auch der Qualitätsanforderungen eine entscheidende Rolle. Seitens der Methodik des

PM ist es unerheblich, ob das Projekt in kleinere Einheiten zerlegt wird und man iterativ jeweils auf Veränderungen und Anpassungen reagieren kann oder ob man das *Projekt* als ein Programm führt [PMI 2006]. Auch diese Sicht auf das PM und das abgeleitete Vorgehen verändert nicht die grundlegenden Dimensionen: definiertes Sachziel, ein festgelegtes Budget und die Qualitätsansprüche, die im Vorfeld bestehen.

Wir verstehen unter den klassischen Vorgehensmodellen diejenigen Vorgehensweisen, die schon im Planungsprozess allen Beteiligten eines Projekts konkrete Arbeitsanweisungen zur Verfügung stellen und seitens Sachziel, Budget, Zeitrahmen und Qualitätsanforderungen wohldefiniert sind. Diese *klassischen* Modelle werden wir unter dem Begriff *Wasserfall* zusammenfassen.

Der Wasserfall liegt auch dem PM bei vielen BI-Vorhaben zugrunde. Gleichwohl haben sich in der BI-Welt alternative Vorgehensweisen zur Projektdurchführung und -steuerung erfolgreich etabliert: die Nutzung von agilen Vorgehensmodellen [Trahasch/Zimmer/Krawatzek/Gansor 2014]. Letztlich ist auch bei den agilen Ansätzen das Sachziel durch die User Story wohldefiniert. Die agile Durchführung soll die explizite Ausprägung der Lösungen besser an den Anforderungen der Bedarfsträger ausrichten, jedoch unter Einhaltung eines festgelegten Budgets und Zeitrahmens. Hierbei ist der Qualitätsanspruch bei einem agilen Vorgehen, wie etwa Scrum, über das DoD (Definition-of-Done: festhalten der Fertigstellungskriterien (Qualitätsanforderungen) des Teams zur Erstellung des Produktes bei Scrum, um die Wünsche des Product Owners an Funktionalität, Qualität, Skalierbarkeit etc. zu erfüllen) ebenfalls fest definiert.

Jedoch zeigt sich ein Mehrwert vieler Big-Data-Ansätze, ähnlich den Data-Mining-Ansätzen, in der Ermittlung bzw. Schärfung eines Sachziels. Dieses explorative Vorgehen entspricht einem Lean-Startup-Ansatz [Ries 2014]. Hierbei werden über das im explorativen Vorgehen inhärente validierte Lernen die Entscheidungen bzw. Modelle anhand von Daten kontinuierlich überprüft und geschärft.

Im Folgenden betrachten wir den möglichen Lebenszyklus eines Big-Data-Vorhabens und zeigen auf, dass die Aufgaben und Herausforderungen von Big Data sich im Verlauf des Lebenszyklus verändern. Somit liegt es nahe, auch ein adäquates Vorgehen pro Phase zu wählen und dadurch im Big-Data-Lebenszyklus einen Mix an Vorgehen und Methoden zu verwenden.

3.3 Projektarten

Um die unterschiedlichen Ausprägungen der Projektart auf den Lebenszyklus zu verdeutlichen, haben wir einige Big-Data-Projekte hinsichtlich ihrer betriebswirtschaftlichen Zielsetzung und der benötigten Implementierungsart betrachtet. In Tabelle 3-1 sind die Schwerpunkte dieser Cluster in Bezug auf den Betrachtungsgegenstand *Lebenszyklus* herausgestellt.

Cluster	Schwerpunkte und Herausforderung
Analyse	<p>Ergebnis: Aussage über eine betriebswirtschaftlich relevante Erkenntnis, oft mit einem strategischen Inhalt.</p> <p>Schwerpunkt: Das Vorhaben besteht in der Regel nur aus einer explorativen Phase, eine operative Implementierung und der Betrieb sind nur auf die Laborumgebung beschränkt.</p> <p>Herausforderung: Der Nutzen der Erkenntnis liegt erst in nachfolgenden Veränderungen der operativen Tätigkeiten. Aufwand/Ertrag von Big Data ist (wenn überhaupt) nur mit hohem zeitlichem Verzug nachweisbar.</p>
Intelligente/ neue Produkte	<p>Ergebnis: Erkenntnis wird Basis eines neuen Geschäftsmodells, das auf einem (Informations-) Produkt für Dritte basiert.</p> <p>Schwerpunkt: Operative Implementierung der Modelle und Technologie durch eine Produktentwicklung. Hierbei muss die Laborumgebung sukzessive in eine Produktivumgebung überführt werden.</p> <p>Herausforderung: Einbindung der IT zur Entwicklung eines neuen noch unsicheren, aber IT-basierten Geschäftsmodells.</p>
Analytische Quelle	<p>Ergebnis: Erkenntnis dient als weitere Quelle für die bestehende BI-Welt.</p> <p>Schwerpunkt: Produktivsetzung der analytischen Verfahren mittels Big Data und anschließende Integrationsleistung bei Einbindung der Ergebnisdaten in die Datenbewirtschaftung von BI.</p> <p>Herausforderung: Verstärkte Einbindung der IT (BI-Entwickler) bei Entwicklung, Wartung und Betrieb.</p>

→

Cluster	Schwerpunkte und Herausforderung
Analytisches System	<p>Ergebnis: Erkenntnis wird zu einem eigenen dispositiven (Informations-)System ausgebaut.</p> <p>Schwerpunkt: Aufbau eines Produktivsystems aus der Laborumgebung</p> <p>Herausforderung: Verstärkte Einbindung der IT bei der Entwicklung und Betrieb.</p>
Effizientere Prozesse	<p>Ergebnis: Erkenntnis wird als Algorithmus in den operativen Systemen verankert.</p> <p>Schwerpunkt: Verankerung der Modelle und Algorithmen von Big Data in den operativen Systemen; wobei dies sowohl klassische Applikationen wie auch Event-orientierte bzw. Data-Stream-orientierte Systeme sein können.</p> <p>Herausforderung: Softwareentwicklungsprojekt der IT mit Big-Data-Techniken.</p>

Tab. 3-1 Schwerpunkte und Herausforderungen der Projektarten

Die unterschiedlichen Schwerpunkte der technischen Umsetzung bei Big-Data-Projekten bedingen somit ein unterschiedliches Vorgehen bei der Steuerung des Big-Data-Projekts. Die dargestellte Spannbreite reicht vom Betrieb der Infrastrukturen für eine Laborumgebung oder ein Data Lab über die Systemintegration zur BI-Welt oder der Applikationslandschaft bis hin zur Entwicklung eigener Informationsprodukte.

3.4 Der Big-Data-Lebenszyklus

Das Application Lifecycle Management (ALM) ist »das Management des Assets *Applikation*, über den gesamten Lebenszyklus von der Idee bis zum End-of-Life, mit dem Zweck, die Anwendungssysteme (hier Big-Data-Lösung) zeitgerecht, verlässlich und anforderungsbezogen zu liefern und gleichzeitig den Wertbeitrag der Applikation kontrolliert und an den Bedürfnissen des Geschäfts ausgerichtet zu gestalten« [Scheuch 2013].

Bei der Betrachtung des Lebenszyklus wird ein fundamentaler Unterschied eines Big-Data-Vorhabens zu den klassischen Vorgehensweisen transparent: der explorative Charakter der Big-Data-Vorhaben in den

frühen Phasen des Lebenszyklus [Scheuch 2015]. Das eigentliche Sachziel ist in dieser Phase der Gewinn an Erkenntnis selbst, und das Sachziel für die weiteren Projektphasen richtet sich an dieser gewonnenen Erkenntnis aus. Ferner gestaltet sich das Sachziel und somit das Produkt im Laufe des Projekts immer detaillierter aus. Diese grundsätzliche stetige Verfeinerung und Anpassung entspricht einem iterativ, inkrementellen Vorgehen, wie etwa beim Rational Unified Process (RUP) oder dem Spiralmodell von Boehm. Kritikpunkt an diesen Ansätzen ist der *Wasserfall*-Ansatz innerhalb der einzelnen Iterationen. Beide Vorgehensweisen legen ihren Schwerpunkt auf die kontrollierte Weiterentwicklung bei Großprojekten. Big-Data-Projekte zeichnen sich eher durch eine hohe Dynamik und kleine bis mittlere Projektgrößen aus. Gleichwohl lässt sich die grundlegende Idee für ein Vorgehen bei Big Data übertragen.

Eine weitere Herausforderung liegt in der Veränderung der technologischen Schwerpunkte bei Big-Data-Projekten im Laufe des Lebenszyklus. Um einen nachhaltigen Nutzen der gewonnenen Erkenntnis einer frühen Phase zu einer Innovation bzw. einem Nutzen für das Unternehmen auszubauen, muss in den folgenden Phasen ein operativer Mehrwert für das Unternehmen entstehen. Dies kann in der Regel nicht ohne eine Einbeziehung der Unternehmens-IT erfolgen.

Es lassen sich vier grundlegende Phasen identifizieren, die wiederum aus Iterationen bestehen, jedoch einen klaren Übergang zur nächsten Phase besitzen. Im Folgenden werden die typischen Phasen eines Big-Data-Vorhabens, wie auch in der Abbildung 3-1 dargestellt, erläutert und die Nutzbarkeit von unterschiedlichen Methoden beschrieben. Ein entscheidender Faktor bei der Auswahl der Methodik ist die Eindeutigkeit und Genauigkeit der Definition des Sachziels. Dies kann in der zeitlichen Abfolge auch zu einem sinnvollen Wechsel bei der Auswahl des geeigneten Vorgehensmodells innerhalb der Phasen führen. Bei der nachfolgenden Erläuterung der Phasen (es wurde bewusst auf die abschließende End-of-Life-Phase verzichtet, um den Schwerpunkt auf die Methodik zu erhalten) soll das einzelne Big-Data-Projekt im Mittelpunkt stehen und nicht eine umfassende Big-Data-Initiative im Unternehmen.

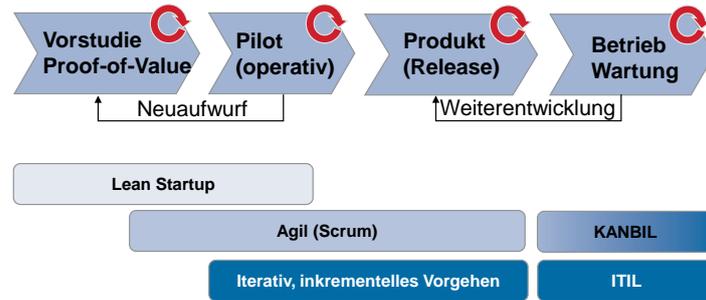


Abb. 3-1 Lebenszyklus von Big Data und Vorgehensweisen

3.4.1 Phase: Vorstudie/Proof-of-Value

Meist beginnen Big-Data-Projekte mit einem Assessment, einer Studie bzw. Analyse mit einem abschließenden Proof-of-Value, um eine vage Idee zu erproben und in der Folge ein Sachziel zu formulieren. Ziel ist es, neben der technischen Machbarkeit vor allem eine unternehmerische Idee und deren wirtschaftlichen Mehrwert zu evaluieren. Sinnvoll erscheint hier ein exploratives Vorgehen bzw. ein klassisches Lean-Startup-Vorgehen, das auch bei den Data/Text-Mining-Ansätzen Verwendung gefunden hat. Als Ergebnis liefert die Projektgruppe neue Erkenntnisse, die zu der Formulierung eines Sachziels (inklusive der Beschreibung eines möglichen wirtschaftlichen Nutzens) führt. Im Big-Data-Kontext ist dies meist eine wirtschaftlich relevante Aussage bzw. ein mathematisch/statistisches Modell. In der Folge werden die Ergebnisse bewertet und häufig durch einen Proof-of-Value der erwartete Mehrwert dargestellt, und im positiven Falle machen die Ergebnisse eine weitere Verfolgung des Sachziels sinnvoll. Das Motto *fail fast, fail early* kann auch dazu führen, dass ein Vorhaben mit einem negativen Ergebnis abgebrochen wird oder ein komplett anderes Sachziel sich als sinnvoll erweist. Das **Ergebnis** dieser Phase ist meist eine Erkenntnis durch die Datenanalyse oder ein Modell zur Simulation von Zusammenhängen.

3.4.2 Phase: Pilot

Bei einer positiven Bewertung des nun formulierten Sachziels aus der explorativen Phase folgt meist eine Pilotphase. Mit einer noch recht vagen Idee versucht man den produktiven Nutzen zu validieren. Als Methodik

bietet sich als Vorgehen ein *Lean-Startup*-Ansatz oder eine agile Methodik an, da beide Ansätze ein bewegliches Sachziel zulassen. Damit der Pilot operativ, wenn auch eingeschränkt, nutzbar ist, wird die IT mit der Betriebsorganisation eingebunden. Hierbei durchläuft der Pilot die klassischen Phasen einer Produktentwicklung, jedoch mit einem eingeschränkten Funktionsumfang und dem Ansatz, ein *Minimal Viable Product* (MVP) [Ries 2014] zu erstellen.

Als **Ergebnis** steht eine operativ nutzbare Pilotinstallation zur Verfügung und kann seitens der Organisation auf ihre Anwendbarkeit und Vorteilhaftigkeit hin überprüft werden. Mit dem Abschluss dieser Phase sollte die Organisation in der Lage sein, eine Entscheidung über den wirtschaftlichen Nutzen zu treffen und im positiven Fall die weiterführende Produktentwicklung zu beauftragen.

3.4.3 Phase: Produktentwicklung

In dieser Phase wird man das Sachziel, etwa als User Story oder als Lastenheft, formulieren können und zu einem typischen iterativen, inkrementellen oder agilen Modell bei der Projektdurchführung überwechseln. Die Entscheidung für eine eher agile oder iterative Vorgehensweise hängt von der Komplexität der Integrationsleistung zu operativen Systemen, der Unternehmenskultur selbst und der Eindeutigkeit des Sachziels als Basis eines Gewerks ab. Als Ergebnis der Phase steht nun ein Produkt in einer ersten operativ nutzbaren Qualität zur Verfügung. Erweist sich die Lösung als *stabil* und werthaltig, kann das Rollout in der ganzen notwendigen Breite in der Organisation erfolgen.

Der Adaption dieser spezifischen Big-Data-Lösung und der Schaffung einer Innovation durch Big Data steht damit nichts mehr im Wege.

3.4.4 Phase: Betrieb und Wartung

In der Folge werden weitere Anpassungen und neue Ideen durch die Applikationswartung permanent umgesetzt und schnellstmöglich produktiv gestellt. Die Big-Data-Lösung befindet sich nun in der frühen Phase der Applikationswartung mit vielen kleinen Veränderungen. Diese sind nicht unbedingt als neue Releasestände, die als eigenständige Projekte angesehen werden sollten, zu verstehen, sondern als *Requests for Changes* (RFC) im Sinne der ITIL. Aus diesem Grunde ist eine Nutzung der agilen Methode des KANBIL, wie auch in Abbildung 3-1 aufgeführt, bei der Wartung ein guter Ansatz, um die Vielzahl an kleineren RFCs effizient

abzuarbeiten. KANBIL ist als eine Synthese aus dem prozessorientierten Framework ITIL und den agilen Praktiken von Kanban zu verstehen.

Wie bei allen Produkten folgt eine ruhigere Phase, die aus Sicht der unternehmerischen Effektivität die *Blütezeit* der Lösung darstellt. Die Verantwortlichen der IT betreiben dieses spezifische Big-Data-System in aller Ruhe, der Fachbereich steuert die Lösung im Rahmen eines Produktmanagements und führt gezielt Modernisierungs- und Refactoring-Maßnahmen durch.

Die eigentliche Aufgabe der Organisation ist es nun, ein spezifisches Application Lifecycle Management (ALM) für ihr Big-Data-Vorhaben zu entwickeln, das seitens der Methodik und des Vorgehens den Lebenszyklus und die Zielstellung, wie in Tabelle 3-1 ausgeführt, beachtet und nicht dogmatisch an einem Vorgehen über alle Phasen festhält. Insbesondere sind die notwendigen agilen Ansätze zu nutzen, um sukzessive die *richtige* Big-Data-Lösung zu erstellen und über ein agiles Anwendungsmanagement weiterzuentwickeln [Hüttermann 2012]. Abbildung 3-2 zeigt die Veränderung der Mitwirkung unterschiedlicher Organisationseinheiten im Lebenszyklus auf. Diese Kriterien helfen, das eigene Big-Data-Vorhaben einzuschätzen und einen adäquaten Application Lifecycle zu planen.

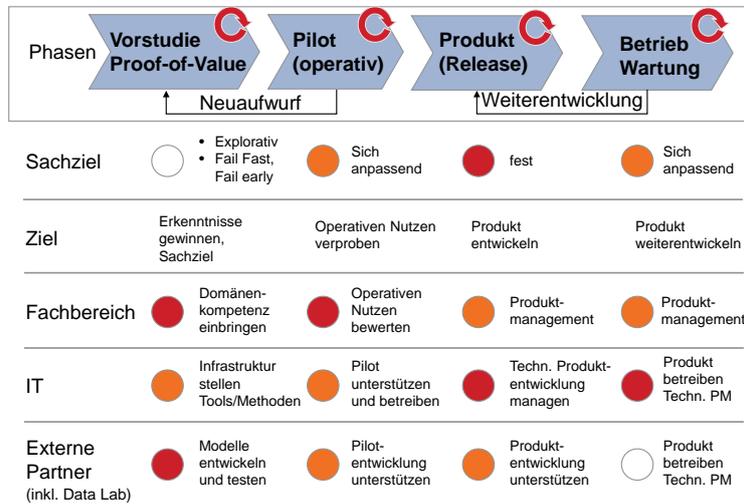


Abb. 3-2 Veränderung der Mitwirkung im Lebenszyklus

3.5 Data Lab, Governance und Unternehmens-IT

Aktuell fokussieren viele der Big-Data-Ansätze alleine auf ein Ergebnis, das sich als *Erkenntnisgewinn* charakterisieren lässt. Die Implementierung von Data Labs ist ein gängiger Ansatz, bei dem Analysten zusammengezogen werden, damit diese Gruppe unabhängig von der Linienorganisation agieren kann. Das (Big) Data Lab wird als *Forschungsstelle* angesehen und hat somit auch mit ähnlichen Akzeptanzhürden zu kämpfen. Dieser Ansatz ist vergleichbar mit einem Outsourcing der Forschungsleistung im Unternehmen. Den erkennbaren Vorteilen hinsichtlich gemeinsamer Nutzung einer Big-Data-Infrastruktur und eines Austauschs der spezialisierten Analysten steht auch ein grundlegender Nachteil gegenüber. Mit einem Data-Lab-Ansatz wird es schwierig, den Mehrwert für die Organisation durch Big Data nachzuweisen, da zum einen die Sachziele dieser Gruppe durch die Linienorganisation aufgegriffen werden müssen und zum anderen der wirtschaftliche Nutzen des Data Lab nur indirekt über den Nutzen für die Linienorganisation ermittelt werden kann; und dies oft mit einem erheblichen Zeitversatz.

Die unterschiedlichen Ansätze von Big Data im Unternehmen benötigen eine starke Einbeziehung der Fachbereiche bei Entwicklung und Change Management, um aus Sicht des Fachbereichs die *richtige* Lösung (oder das *richtige* Produkt) zu erstellen. Die inhaltliche Führung und Ausrichtung der Big-Data-Forscher, auch in einem Data Lab zentral organisiert, obliegt den Fachbereichen, damit eine grundlegende Steuerung der Ausrichtung des explorativen Arbeitens erfolgt. Man verliert eventuell überraschende Ergebnisse eines unabhängigen Forschers, aber gewinnt Stringenz bei den Forschungsinhalten. Ferner steuert das Produkt-/Programmmanagement die gesamten Tätigkeiten über alle Phasen des geschichteten Lebenszyklus – unabhängig von der Form der technischen oder organisatorischen Implementierung.

Von Bedeutung ist ferner der Abgleich (Governance) mit einem zentralen Competency Center (CC) für Big-Data-Vorhaben im Unternehmen. Zum einen soll das spezifische Big-Data-Vorhaben Rücksicht auf die Big-Data-Strategie nehmen. Zum anderen erscheint es sinnvoll, dass die Big-Data-Vorhaben von den Best Practices, auch bei der Auswahl und dem Betrieb der Big-Data-Infrastrukturen, im Unternehmen profitieren sollten. Die Rolle des CC zur inhaltlichen Steuerung sehen wir in diesem Zusammenhang eher kritisch, da hierbei der explorative und kreative

Charakter durch Planungsprozesse behindert wird und das CC inhaltlich nicht die notwendige fachliche Kompetenz aufbauen wird.

Ferner wird durch das Modell des Big-Data-Lebenszyklus die entscheidende Rolle der Unternehmens-IT deutlich. In den frühen Phasen beschränkt sich die IT auf die Bereitstellung der Infrastruktur und einigen Hilfestellungen bei der Datenextraktion und Betrieb. In den späteren Phasen erhöht sich die notwendige Integrationstiefe zu der Applikationslandschaft und die IT muss bei der eigentlichen Produktentwicklung eine gewichtigere Rolle einnehmen. Die Abbildung 3-3 beschreibt im Lebenszyklus mögliche Übergabepunkte der technischen Projektverantwortung an die Unternehmens-IT, um das Big-Data-Projekt mit dem nunmehr stark IT-lastigen Schwerpunkt operativ fortzuführen.

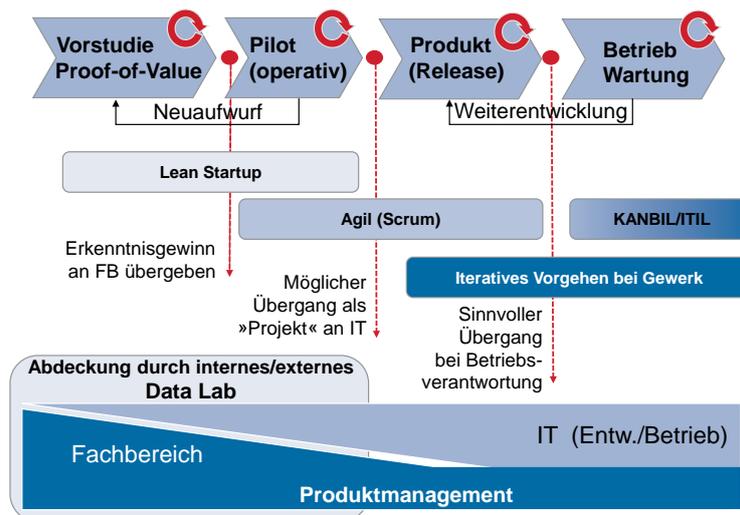


Abb. 3-3 Mögliche Übergabepunkte an die Unternehmens-IT

3.6 Empfehlung

Das Big-Data-Vorgehen entspricht im Wesentlichen dem zyklischen Vorgehen beim Spiralmodell von Boehm, wobei in jedem Zyklus bzw. jeder Iteration das Sachziel oder das Ergebnis wieder neu evaluiert und geschärft wird. Jedoch bieten sich andere Methoden der Projektsteuerung in den Iterationen an, da Big-Data-Projekte sich von einem explorativen Vorgehen über eine iterative Produktentwicklung hin zu einer Betriebs- und Wartungsphase verändern. Jede Phase (mit ihren Iterationen) hat Optionen für die Auswahl einer geeigneten Methodik für die Projektsteuerung. Einzig bei der frühen Phase eines explorativen Vorgehens erscheint es uns empfehlenswert, einen Lean-Startup-Ansatz zu wählen. Die Folgephasen sollten ein angemessenes Vorgehen und vertraute Methodik verwenden, wobei dies von der Integrationstiefe der Lösung zu den operativen Systemen, der vorherrschenden Unternehmenskultur und der Genauigkeit des Sachziels inklusive der Qualitätsanforderungen abhängt. Data Labs sehen wir als einen effizienten Ansatz, Big-Data-Kompetenz zu bündeln, jedoch kann dies nur der Startpunkt für die Produktentwicklung und operative Implementierung inklusive des IT-Betriebs sein. Hierbei benötigt man die Kompetenz der Unternehmens-IT, um das (Software-) Produkt zu entwickeln und nachhaltig zu betreiben. Trotz der Bedeutung der Unternehmens-IT bei Entwicklung und Betrieb *gehört* das Produkt dem Fachbereich und wird durch ein Produktmanagement als Führungssystem gesteuert. Das Produktmanagement ermöglicht eine Erfolgsmessung von dem Big-Data-Vorhaben über den gesamten Lebenszyklus. Die einseitige Ausrichtung der klassischen Methoden und Vorgehensweisen auf Effizienz und Kosten wird dem meist explorativen Charakter der Big-Data-Ansätze nicht gerecht. Effektivität und Veränderung müssen als wesentliche Faktoren im Lebenszyklus Beachtung finden und durch die Verwendungen agiler Methoden und Vorgehen unterstützt werden.

3.7 Literatur

[Ferstl/Sinz 2008] Ferstl, O.-K.; Sinz E.: Grundlagen der Wirtschaftsinformatik. 6. Auflage, Oldenbourg Wissenschaftsverlag, 2008.

[Dittmar 2016] Dittmar, C.: Die nächste Evolutionsstufe von AIS: Big Data. Erweiterung klassischer BI-Architekturen mit neuen Big Data Technologien. In: Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Business-Intelligence-Technologien und -Anwendungen. 5., vollständig überarbeitete Auflage. Unter Mitarbeit von Peter Chamoni, Springer-Verlag, 2016, S. 55-65.

[Hüttermann 2012] Hüttermann, M.: Agile ALM. Manning, 2012.

[PMI 2006] PMI: The Standard for Program Management. Project Management Institute, Inc., Pennsylvania, USA, 2006.

[Ries 2014] Ries, E.: The lean startup: how today's entrepreneurs use continuous innovation to create radically successful businesses. Crown Publishing, 2011, 2014.

[Scheuch 2013] Scheuch, R.: Warum versagen typische Alm Ansätze. IM +io Magazin, 2013.

[Scheuch 2015] Scheuch, R.: Mehr Effizienz für Big Data – Neue Ansätze für das Applikationsmanagement. IT-Management, September 2015.

[Trahasch/Zimmer/Krawatzek/Gansor 2014] Trahasch, S.; Zimmer, M.; Krawatzek, R.; Gansor, T. (Hrsg.): Memorandum für Agile Business Intelligence. dpunkt.verlag, 2014.

4 Big-Data-Architektur und -Technologie

Ralf Finger und Carsten Dittmar

Die Abkehr von unrealistischen klassischen dispositiven Architekturparadigmen wird durch neue Big-Data-Architekturen und Technologien befruchtet.

Klassische dispositive Architekturen und Big-Data-Architekturelemente und Technologien werden aktuell häufig noch als zwei getrennte Welten aufgefasst, langfristig werden beide Welten zu einem analytischen Ökosystem zusammenwachsen.

Der Markt für Big-Data-Technologien ist sehr vielfältig und dynamisch. Klare Abgrenzungskriterien zur Unterscheidung von Teilssegmenten von Big-Data-Technologien fehlen. Zudem sind in vielen Segmenten deutliche Überschneidungen zu klassischen Technologien festzustellen.

Big-Data-Architekturelemente und Technologien bieten teilweise neue Möglichkeiten für die Analyse großer und schlecht strukturierter Datenmengen. Damit können sie klassische BI-Ansätze unterstützen und neue Nutzungsszenarien ermöglichen.

4.1 Klassische, zentralistische Architekturen auf dem Prüfstand

Die *richtige* Architektur für die dispositive Datenverarbeitung war lange klar definiert. Ein idealtypisch singuläres (Enterprise) Data Warehouse sammelt in einem Hub&Spoke-Ansatz aus den unterschiedlichen operativen Quellsystemen die relevanten Daten auf und harmonisiert, integriert und persistiert diese in einem mehrschichtigen Datenintegrations- und Datenveredelungsprozess. Aus diesem vielzitierten *Single Point of Truth* werden anschließend Datenextrakte in der Regel multidimensional in voneinander fachlich abgrenzbaren Data Marts gehalten, die dann wiederum den Presentation Layer mit seinen spezifischen Reporting- und Analysewerkzeugen (Business Intelligence im engen Sinne) versorgen. Diese Mehrschichtenarchitektur hat in bestimmten Anforderungskontexten weiterhin seine Gültigkeit. So zum Beispiel im Einzelunternehmen, in dem übergreifend konsistente Auswertungen über abgestimmte Auswertungssichten zum Beispiel nach Produkt, Kunde oder Region bereitgestellt werden sollen.

Die Einhaltung dieser idealtypischen klassischen Architektur wird jedoch immer schwieriger, wenn Fachbereiche eine höhere Änderungsdynamik fordern, als diese von einer zentralen Data-Warehouse-Infrastruktur geleistet werden kann. Auch in Konzernstrukturen, in denen sich mitunter sehr unterschiedliche Geschäftsmodelle unter einem Dach finden, ist ein zentrales Data Warehouse oft nicht die richtige Lösung. Denn, wenn sich die Steuerungsnotwendigkeiten unterschiedlicher Geschäftsmodelle stark unterscheiden, behindert ein technisch integriertes Data Warehouse mehr, als dass es Synergien schafft. So finden sich in der Praxis selten vollständig konsolidierte und zentralisierte Architekturansätze für analytische Anwendungen. Vielmehr sind Informationssilos und heterogene Architekturen in gewachsenen dispositiven Landschaften eher die Regel als die Ausnahme. Dadurch kann im Allgemeinen auch die fachliche und/oder technische Transparenz über die Datenveredlungs- und Datenkonsolidierungsprozesse nicht aufrechterhalten werden. Ein starkes Anwachsen der Schattensysteme in Form von sogenannten *Spread Marts* oder autark und gegebenenfalls anarchisch agierenden *Self-Service*-Lösungen unter Umgehung von architekturellen Vorgaben verstärkt diesen Trend noch.

4.2 Der Trend zu flexiblen analytischen Ökosystemen

Big Data als Analyse und Echtzeitverarbeitung großer unstrukturierter und kontinuierlich fließender Datenmengen aus einer Vielzahl unterschiedlicher Datenquellen zur unmittelbaren Entscheidungsunterstützung fordert im Gegensatz zu den klassischen Ansätzen der dispositiven Datenverarbeitung keinen *Single Point of Truth*. Es sollen vielmehr nur die jeweiligen Datenquellen verbunden werden, die für den jeweiligen Big-Data-Anwendungsfall benötigt werden, um Datenvolumen-, Datentypen- und Datenquellenvielfalt sowie Anforderungen an die Verarbeitungsgeschwindigkeit zu erfüllen. Damit setzt Big Data den Trend zur Dezentralisierung von dispositiven Architekturen fort und führt zur Bildung von Multi Platform Environments oder analytischen Ökosystemen. Traditionelle Data Warehouses mit konsistentem Datenbestand bilden zwar in solchen analytischen Ökosystemen nach wie vor die ideale Plattform für Reporting, Dashboarding, Performance Management und OLAP. Doch neue *Stand-alone*-Datenplattformen entstehen in Ergänzung zur bestehenden Architektur und dienen der Verarbeitung und der Analyse von Big-Data-Anwendungsfällen.

Die integrierte Analyselandschaft der Zukunft wird vielfältiger und komplexer und stellt die Flexibilität in den Vordergrund. Die Integration erfolgt hier eher logisch, durch einheitliche Metadaten, Data Governance und Stammdaten. Der physische Integrationsanspruch tritt wieder zurück. Anwenderunternehmen müssen hier auch ihre Grundsatzentscheidungen bezüglich Make-or-Buy überprüfen. Durch die zunehmende Technologievielfalt ist es eine Strategie, die vorkonfektionierten Lösungen etablierter Anbieter zu adaptieren. Die klassischerweise aus dem Bereich Open Source entstammenden Technologien wirken hier auf den ersten Blick günstiger, es muss aber mehr Basis-Know-how aufgebaut werden, um die Interoperabilität von Architekturkomponenten sicherzustellen. Daher wird möglicherweise zukünftig ein Nebeneinander von selbst erstellten und integrierten vorkonfektionierten Lösungen und extern zum Beispiel in einer Cloud bereitgestellten Architekturkomponenten festzustellen sein. Klassische Grenzen zwischen operativen und dispositiven Systemen, zwischen disjunkten Werkzeugen für unterschiedliche Analysebedarfe, zwischen Backend-orientierter Datenintegration und Frontend-orientierter Datenanalyse, aber auch zwischen Produktiv- und temporären Entwicklungs- und Evaluationsumgebungen konvergieren zunehmend in diesen analytischen Ökosystemen.

Einen Überblick über ein entsprechendes analytisches Ökosystem liefert die Abbildung 4-1. Neben den Komponenten, die aus einer klassischen BI-Architektur heraus bekannt sind, sind dort auch Elemente einer Lambda-Architektur für die Realtime-Verarbeitung zum Beispiel von Sensordaten mit einem Speed Layer, Batch Layer und Service Layer enthalten. Dadurch können Realtime-Daten historisiert abgelegt, aber auch an ein Realtime-Monitoring durchgereicht werden. In einigen Anwendungsfällen kann es dabei sinnvoll sein, den Batch Layer mit den Backend-orientierten Komponenten der klassischen BI-Architektur zusammenzuführen. Zudem enthält die Architektur auch eine Laborumgebung (Analytic Lab) für Datenanalysen auf Basis flexibler Schemata (*Any Schema* oder *noSQL*).

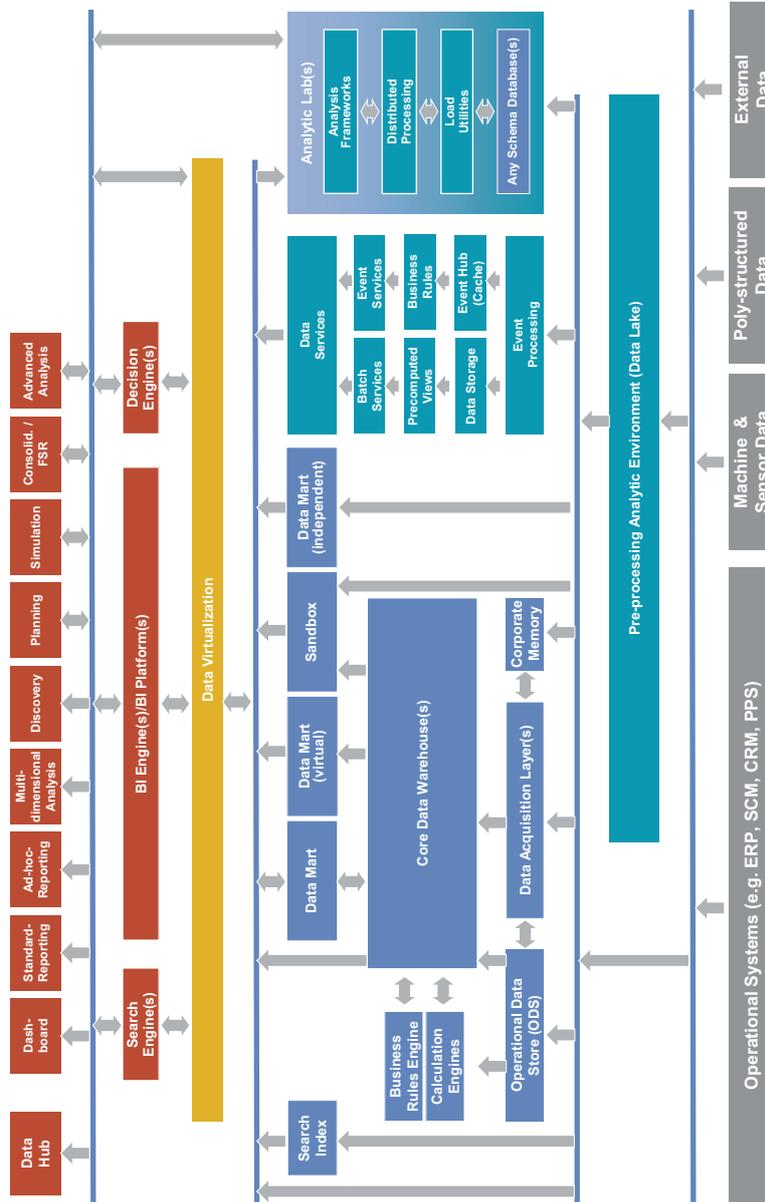


Abb. 4-1 Analytisches Ökosystem im Überblick

4.3 Big-Data-Technologien

Der Begriff der *Big-Data-Technologie* wird heute sehr unscharf verwendet. Zudem zeichnet sich der Markt für Technologien für den Einsatz im Umfeld Big Data durch eine extreme Dynamik aus, sodass neben den etablierten Technologien, die aus dem klassischen BI-Umfeld bekannt sind, regelmäßig neue Lösungen und Anbieter auftauchen. So treten Open-Source-Lösungen sowie kommerzielle Anbieter mit innovativen Technologien in direkte Konkurrenz zu etablierten *Full-Stack*-BI-Anbietern, die mit erprobten und über Jahre weiterentwickelten umfangreichen Softwarelösungen glänzen. Die Anzahl der Alternativen und Lösungen zur Umsetzung einer oben beschriebenen Architektur wächst exponentiell. Die Dynamik im Markt ergibt sich zudem auch aus der Tatsache, dass sich die angebotenen Lösungen oder Technologien ergänzen und diese dann gegebenenfalls zu einer Lösung durch entsprechende Kooperationen bzw. Übernahmen verschmelzen. Zum Teil wechseln auch Technologien ihren Fokus und das damit verbundene Marktsegment. Die Übersicht in Abbildung 4–2 gibt einen ersten Überblick über die verschiedenen Segmente, erhebt dabei aber keinen Anspruch auf Vollständigkeit.

Die Big-Data-Technologien werden gemäß der Übersicht in folgende Segmente aufgeteilt:

- **Infrastructure**
Basistechnologien zum Data Management von insbesondere Big Data
- **Analytics**
Werkzeuge für Anwender (z.B. Analysten, Data Scientists) zur Datenanalyse mit unterschiedlichen Analysemethoden
- **Applications**
Bereits vorkonfektionierte (pre-packaged) zum Teil branchen- bzw. funktionsbereichsbezogene Big-Data-Anwendungen
- **Cross-Infrastructure/Analytics**
Lösungen von den klassischen *Full-Stack*-BI-Anbietern
- **Open Source**
Open-Source-Initiativen und Frameworks, die wiederum die anderen Segmente herausfordern bzw. die Basis für kommerzielle Distributionen darstellen
- **Data Sources & API**
Zugriff auf externe Data Brocker mit anwendungsbezogenen relevanten Spezialdaten

Die Zuordnung einzelner Technologien zu den jeweiligen Subsegmenten im Detail erfolgte anhand der Kernfunktionalitäten bzw. Kerneigenschaften und liefert im Detail viele Überschneidungen. Das unscharfe Clustering zeigt, dass es etablierte Standardarchitekturen bzw. Technologiekombinationen aus sich jeweils substituierenden Technologiekomponenten (noch) nicht gibt, sondern identische Anwendungsfelder auch mit komplett unterschiedlichen Technologiewerkzeugen gelöst werden können.

Viele der Big-Data-Technologien lassen sich anhand von charakteristischen Eigenschaften beschreiben. Die neuen Technologien unterstützen so häufig die Speicherung und Auswertung von teil- und unstrukturierten Daten, ermöglichen die schnelle Datenverarbeitung durch effiziente Nutzung von Arbeitsspeicher bzw. arbeitsspeicherähnlichen Ressourcen und erlauben die Verarbeitung von Echtzeitdatenströmen.

So lassen sich mithilfe der Technologien häufig neben den strukturierten Daten auch teil- und unstrukturierte Daten (z.B. XML, JSON, Logdaten von Sensoren oder Webinhalte, Multimediainhalte) speichern und auswerten. Zum Teil muss dafür die Struktur der Daten im Vorfeld nicht explizit modelliert werden, sodass Daten zunächst gespeichert werden können, ohne die genaue Zielrichtung der Datenablage bereits zu kennen. Dies wird in Abgrenzung zu etablierten relationalen Datenbanken auch als *Schema-Later-Ansatz* bezeichnet, da vor Beladung und Auswertung bei diesen Datenbanken in der Regel keine oder nur geringe Datenmodellvorgaben existieren müssen. Besondere Popularität und Verbreitung hat hier die Open-Source-Lösung Hadoop erreicht. Wesentliche Charakteristika von Hadoop sind ein verteiltes Dateisystem HDFS (Hadoop Distributed File System) und ein Computing Framework, das einen Rahmen bietet, um eine Vielzahl von darauf aufsetzenden Lösungen, beispielsweise Abfragesprachen (z.B. Hive), Suchsysteme (z.B. Solr) oder Machine-Learning-Umgebungen (z.B. Mahout) in einem gemeinsamen, hochskalierbaren Cluster ablaufen zu lassen sowie Daten in großer Menge auf kostengünstiger Standardhardware zu speichern.

HDFS ist ein in Java geschriebenes Dateisystem und stellt über einen einstellbaren Redundanzfaktor sicher, dass alle Dateien automatisch auf mehreren (üblich sind drei) Knoten abgespeichert werden. Somit ist bei einem Ausfall eines Knotens kein Datenverlust zu befürchten. HDFS unterstützt per se keine Veränderung von Dateien, sondern nur Append- und Delete-Operationen. Das Computing Framework wird im klassischen Hadoop-Ökosystem durch MapReduce repräsentiert. Die Execution Engine MapReduce ist daraufhin optimiert, Jobs auf einzelne Knoten zu verteilen und nach Abarbeitung wieder zusammenzuführen. Dabei wird sichergestellt, dass auch bei fast beliebigen Ausfällen von Knoten der Job sicher zu einem korrekten Ende geführt wird. Mit YARN (Yet Another Resource Negotiator) ist mittlerweile der Nachfolger zu MapReduce entstanden. YARN sichert Prozessen dediziert Ressourcen (CPU, Memory) zu. Durch die Ausgliederung des Ressourcenmanagements aus MapReduce können nun auch andere Execution Engines mit MapReduce im selben Cluster laufen und trotzdem kann der gesamte Ressourcenbedarf an zentraler Stelle verwaltet und zugeteilt werden. Das wiederum ermöglicht es, Batch- und Realtime-Anwendungen in einem Cluster zu mischen.

Ergänzend zu Hadoop werden heute vier Gruppen schemaloser Datenbanken diskutiert. Schwerpunkt dieser Datenbanken ist stets die Speicherung und Auswertung von Daten mit SQL-Mechanismen, aber auch darüber hinausgehenden Prinzipien. Aus diesem Grund spricht man bei diesen Datenbanken auch von NoSQL (im Sinne von *not only SQL*).

Dabei sind zunächst die Key-Value-Datenbanken zu nennen, deren Datenablage aus identifizierenden Schlüsseln besteht, die auf Array-Strukturen verweisen. Die interne Speicherung hierfür basiert auf Hash-Tabellen und ist daher sehr performant, Der Zugriff auf die Werte im Array erfolgt allerdings ausschließlich über den Key, sodass im Bereich der Datenanalyse Begrenzungen zu beachten sind. Ein Beispiel für Key-Value-Datenbanken ist Apache Accumulo.

Hiervon zu unterscheiden sind spaltenorientierte Datenbanken (Columnar Databases) wie zum Beispiel Apache Cassandra. Hier werden Datensequenzen nicht wie bei herkömmlichen Datenbanken zeilenweise, sondern spaltenweise abgelegt. Dadurch lässt sich je Datenzeile leicht eine flexible Anzahl Spalten anlegen. Oft werden auch sich wiederholende Werte nicht erneut gespeichert, sodass sich eine Kompression des Datenvolumens im Vergleich zu zeilenorientierter Speicherung ergibt. Diese Form der Datenspeicherung ist besonders für analytische Lesezugriffe

über Spaltenausprägungen geeignet, besitzt jedoch Einschränkungen bei Update- und Append-Operationen in Datensätzen.

Graph-Datenbanken bieten besondere Mechanismen zur Speicherung stark verknüpfter Daten, wie sie zum Beispiel bei der Auswertung sozialer Netzwerke üblich sind. Dafür unterscheiden diese Datenbanken die Objekte *Knoten* und *Link* und bieten damit eine selbstbeschreibende Form zur Speicherung und Analyse netzwerkartiger Strukturen. Graph-Datenbanken sind weniger verbreitet, was jedoch auf die ausgeprägte Zweckorientierung zurückzuführen sein dürfte.

Schließlich sind die sehr populären Document Stores zu nennen. Diese Datenbanken (z.B. MongoDB) unterstützen die Ablage insbesondere von JSON-Dokumenten (Java Script Object Notation). Hierbei handelt es sich um eine sehr effiziente textartige Dateistruktur, die sehr flexibel, hierarchisch aufgebaut werden kann. Da die Semantik in den JSON-Dokumenten selbst abgelegt ist, können diese in Document Stores gespeichert werden, ohne einem übergreifenden Schema zu genügen. Der typische Einsatzbereich für JSON-Datenbanken sind Webanwendungen. Aber auch Sensordaten werden heute nicht selten über Gateway-Rechner im JSON-Format an Message Broker publiziert.

Auch durch die Verarbeitung von Daten im Arbeitsspeicher bzw. arbeitsspeicherähnlichen Komponenten wird ein besonderer Geschwindigkeitsvorteil erreicht, da die Datenablage und Datenverarbeitung innerhalb der gleichen Komponente durchgeführt wird und so Datentransporte und zeitaufwendige (mechanische) Plattenzugriffe vermieden werden. Um größere Datenmengen In-Memory vorzuhalten, kommen dabei in der Regel innovative Caching-Verfahren, hohe Kompressionsraten sowie die Verarbeitung von komprimierten Daten ohne Dekompression und auch unterschiedliche Speicherformen von Daten (spaltenbasiert, hybrid, indexorientiert etc.) zum Einsatz. Die Verwendung von derartigen In-Memory-Ansätzen wird von einer Vielzahl von Big-Data-Technologien eingesetzt. Aus der Open-Source-Community ist auf Apache Spark zu verweisen. Da Spark nicht alle Verarbeitungsschritte persistiert, ist Spark deutlich schneller als MapReduce.

Für die Echtzeitverarbeitung von Big Data sind Replikation, Bus-Architekturen, Microbatches oder Continuous-Realtime-Streaming-Werkzeuge auf Basis von Message Queues in die Architektur zu integrieren. Auch für diesen Bereich existieren zahlreiche Plattformen der etablierten Softwarehersteller sowie eine Vielzahl von Technologie-Newcomer. Apa-

che Storm ist als Vertreter der Open-Source-Community zu nennen. Storm ist eine Stream Processing Engine und bietet ein verteiltes Framework zur Verarbeitung von Daten in der Bewegung. Storm zeichnet sich insbesondere dadurch aus, dass es Einzeldatensätze verarbeitet und diese nicht zu Microbatches zusammengesetzt werden. Zusätzlich laufen Storm-Tasks, anders als MapReduce-Prozesse, solange, bis sie mittels eines Eingriffs beendet werden. Sie kennen daher kein dem Prozess inhärentes Ende.

4.4 Big-Data-Plattformen

Neben dem Begriff der analytischen Ökosysteme wird im Markt gerade von den etablierten, großen Anbietern (vgl. *Cross Infrastructure/Analytics* in Abbildung 4–2) das Leistungsspektrum einer integrierten Big-Data-Plattform betont. Dieses wird bewusst im Gegensatz zu der aktuellen Anbieter- und Werkzeugvielfalt im Bereich Big Data positioniert, indem die Vorteile eines integrativen Infrastrukturansatzes herausgestellt werden. Diese sind aufgrund der Heterogenität der technologischen Optionen nicht ganz von der Hand zu weisen und akzentuieren zum Beispiel eine vereinfachte Installation, Konfiguration, Administration durch vorpaketierte Routinen, verbesserten Support und Dokumentation sowie eine abgestimmte Integration zwischen den einzelnen Komponenten zum Beispiel für den Datenaustausch zwischen unterschiedlichen Big-Data-Store-Technologien.

Besonderes Merkmal dieser Positionierung ist auch, dass nicht nur die Big-Data-typischen neuen Anwendungsfelder (unstrukturierte Daten, realtime) dem Begriff Big Data zugerechnet werden, sondern auch die klassischen BI-Stacks des jeweiligen Anbieters. Dadurch entsteht das Verständnis einer Big-Data-Plattform im weiteren Sinne, das Core Data Warehouse, Data Marts und traditionelle BI-Frontends umfasst. Diese Abgrenzung mag für Experten, die den Markt für BI und Data Warehousing langjährig begleiten, gewöhnungsbedürftig sein. Sie erlaubt den *Cross-Infrastructure-Analytics*-Anbietern jedoch, die Vorteile herauszustellen, die sich insbesondere an den Übergängen zwischen traditionellen BI-Architekturkomponenten und Big-Data-Architekturkomponenten realisieren lassen.

Dazu gehören zum Beispiel integrierte paketierte Lösungen für die Anwendung von Predictive-Analytics-Modellen auf Basis historisierter Daten im Core Data Warehouse und unstrukturierter Daten in Big Data

Stores auf Realtime-Datenströme im Sinne einer Lambda-Architektur (vgl. Abschnitt 4.2). Auch die automatisierte Datenüberleitung zwischen Big Data Stores (z.B. JSON) in relationale Datenbanken mittels integrierter Schema-Discovery-Mechanismen per Batch ist hier zu nennen, ebenso wie der relationale Zugriff auf Hadoop direkt aus der Datenbank per Views basierend auf Hive.

Insgesamt ist auch aus technologischer Sicht darauf aufmerksam zu machen, dass Big-Data-Basistechnologien mehr und mehr in die Plattformen der großen Hersteller vollständig integriert werden. Gleichsam bringen sich diese Hersteller wiederum aktiv in die Open-Source-Community als Zulieferer ein. So basieren beispielsweise neuere Datenintegrationswerkzeuge und Calculation Engines auf Apache Spark. Daher werden Anwenderunternehmen zukünftig viele als Big-Data-Technologien bekannte Werkzeuge aufgrund einer strategischen Plattformscheidung faktisch einsetzen, ohne diese im Einzelnen betrachtet zu haben. Des Weiteren beinhaltet das Leistungsversprechen dieser Big-Data-Plattformen im weiteren Sinne auch die verbesserte Abstimmung der Datenspeicher aus fachlicher Sicht über einheitliche Metadaten und Governance. Dies wird gerne mit einem weiteren V als *Veracity* im Sinne von *Wahrhaftigkeit* charakterisiert.

Das Spektrum an Technologien und resultierenden Architekturen ist vielfältig, zeigt aber auch die inhaltliche Nähe von Big Data und klassischer BI, sodass ein interessanter Pfad der Integration zu erwarten ist, da die Mehrzahl der Anwendungsszenarien in einer mittelfristigen Perspektive additiv zu bestehenden sind. Gleichsam ist aber die Technologie immer nur ein Mittel zum Zweck und bedarf einer geeigneten Anwendung. Big-Data-Technologien werden demzufolge nicht Anwendungsfälle von selbst finden, sondern der Anwendungsfall stellt den Ausgangspunkt dar.

4.5 Literatur

[Turck 2014] Turck, M.: Big data landscape v 3.0, <http://de.slideshare.net/mjft01/big-data-landscape-matt-turck-may-2014> (letzter Abruf: 10.05.2016).

5 Auswirkungen von Big Data auf die Organisation

Lars Tams

Big Data muss organisatorisch in die zentrale Unternehmens-IT. – Nur die Fachbereiche können erfolgreiche Big Data initiieren. – Big Data ist eine eigene Organisationseinheit.

Diese Kontroverse zeigt das aktuelle Dilemma, es gibt keine eindeutige organisatorische Einbindung von Big-Data-Projekten.

5.1 Big Data – schöne neue Datenwelt: Wer ist Treiber, wer Sponsor und wer Owner?

Wer verantwortet Big Data im Unternehmen? Wer ist Treiber? Wer profitiert? Wer ist Budget Owner? Wen brauchen wir? All dies sind Fragestellungen mit Blick auf die Unternehmensorganisation, die sich bei der Einführung von Big-Data-Aktivitäten ergeben. In diesem Kapitel werden anhand verschiedener Szenarios ein Überblick sowie die Vor- und Nachteile unterschiedlicher Organisationsformen dargestellt. Darüber hinaus erfordern Big-Data-Aktivitäten auch neue Rollen und Fähigkeiten, die im weiteren Verlauf beschrieben werden.

Daten sind das Öl des neuen Jahrtausends, Information gilt als vierter Produktionsfaktor, all dies wird durch Big-Data-Technologie und die fortschreitende Digitalisierung weiter beschleunigt und entwickelt sich zu einem festen wie auch kritischen Erfolgsfaktor im Unternehmen. Die Besonderheit an Big Data ist, dass es neben direkten Potenzialen, wie zum Beispiel der Optimierung von Produktionsprozessen, Sekundärpotenziale bis hin zu komplett neuen Geschäftsmodellen ermöglicht. Diese sind allerdings im Moment der Datensammlung oftmals noch nicht bekannt und müssen erst entwickelt werden. Mit Blick auf die Verantwortlichkeit für Big-Data-Initiativen im Unternehmen bedeutet dies, dass hohe fachliche Kompetenz, Innovationsfähigkeit für neue Geschäftsmodelle sowie technische Kompetenz sichergestellt werden müssen. Diese Anforderung wurde schon 2011 in der Detecon/Bitkom-Studie *IT-Organisation 2015*

prognostiziert, wonach sich die IT-Organisation zu einer Steuerungs- und Demand-Organisation mit zunehmendem Business-Know-how entwickelt haben wird. »Die IT entwickelt sich vom reinen *Operator* zum *Business Enabler*, bei dem die Service-Erbringung gegenüber der Service-Steuerung in den Hintergrund tritt« [Detecon/Bitcom 2011, S. 3].

5.2 Big-Data-Strategie

Ein wesentlicher Einflussfaktor für die organisatorische Eingliederung ist die Big-Data-Strategie eines Unternehmens. Ist diese Teil der generellen IT-Strategie, besteht die Gefahr, dass hier eher operativ funktional (Software, Hardware, Infrastruktur) und weniger fachlich, inhaltlich ausgerichtet an Big-Data-Fragestellungen herangegangen wird [Gansor/Totok 2015, S. 44-45]. Dies hat insbesondere Auswirkung auf die Akzeptanz von Big-Data-Initiativen im Management.

Die Big-Data-Strategie sollte analog zur BI-Strategie idealtypisch an der Unternehmensstrategie ausgerichtet sein, deren Ziel es ist, den langfristigen Erfolg des Unternehmens zu sichern [Hungenberg 2014, S. 4]. Hiervon lassen sich dann die Vision sowie die fachlichen, technologischen und organisatorischen Ziele ableiten. Generell gilt, dass die unternehmensstrategische Bedeutung, der Reifegrad von Business/IT-Alignment und auch der Grad der digitalen Transformation ausschlaggebend für die organisatorische Einbettung von Big Data sind.

5.2.1 Verantwortung für Big Data im Unternehmen

2013 entwickelten Tom Casey, Kumar Krishnamuthy und Boris Abezgauz von Booz & Company ein Modell, das die Abhängigkeiten der Business- und IT-Ownership in drei Organisationsszenarios darstellte [Casey/Krishnamuthy/Abezgauz 2013]. Hierbei wurden die in Abbildung 5-1 gezeigten drei wesentlichen Ausprägungen herausgestellt:

1. Organisation mit Verantwortung bei der IT
2. Organisation mit Verantwortung beim Fachbereich
3. Matrixorganisation

Jede Organisationsform hatte der Studie zufolge Vor- und Nachteile. Wenig überraschend war die Empfehlung einer Matrixorganisation als optimale organisatorische Einbindung. Wobei die Autoren allerdings die

Antwort nach der Art und Weise der Einbindung in die Gesamtunternehmung einer solchen Organisationform schuldig blieben. Eine weitere Erkenntnis der Untersuchungen war, dass die meisten Unternehmen damals (noch) keine klare Big-Data-Strategie und definierte Verantwortung für Analytics hatten.

Grad der fachlichen Verantwortung	hoch	Szenario 2 Ein funktional aufgestellter Fachbereich ist für die Entwicklung der Datenverarbeitungskapazitäten verantwortlich. Die IT übernimmt deren Aufträge.	Szenario 3 Eine Matrixorganisation, vertreten durch einen Vorgesetzten, der sowohl die fachliche Seite als auch die IT-Seite repräsentiert. Hier liegt die Verantwortung zur Entwicklung der Datenverarbeitungskapazitäten in Kohärenz mit der Unternehmensstrategie.
	niedrig	Das Unternehmen hat weder eine klare Verantwortung bezüglich der Daten oder der Analysekompetenzen noch für deren Strategie.	Szenario 1 Die Unternehmens-IT ist für den Aufbau und die Verwaltung der Datenverarbeitungskapazitäten verantwortlich.
		niedrig	Grad der IT-Verantwortung
			hoch

Abb. 5-1 Strategy & Business [Casey/Krishnamuthy/Abezgauz 2013]

Ausgehend von dieser These möchten wir versuchen, die organisatorische Einbettung anhand der Dimensionen *Reifegrad von Business/IT-Alignment* und *strategische Bedeutung von Big Data für das Unternehmen* herzuleiten. Abhängig davon, ob Big-Data-Projekte nur einzelne (Fach-) Bereiche mit einem klaren Business Case betreffen (z.B. die Optimierung von Kundenansprachen im Marketing), ob es eine hohe Business-Affinität innerhalb der IT-Organisation und eine überschaubare Anzahl von Anwendungsszenarien gibt oder ob Unternehmen Big Data als Innovationsstrategie für ihr Geschäftsumfeld verstehen und auch ohne klaren Business Case Big-Data-Projekte starten, um künftige Geschäftsmodelle zu sichern und auszubauen, ergeben sich unterschiedliche Organisationsmöglichkeiten.

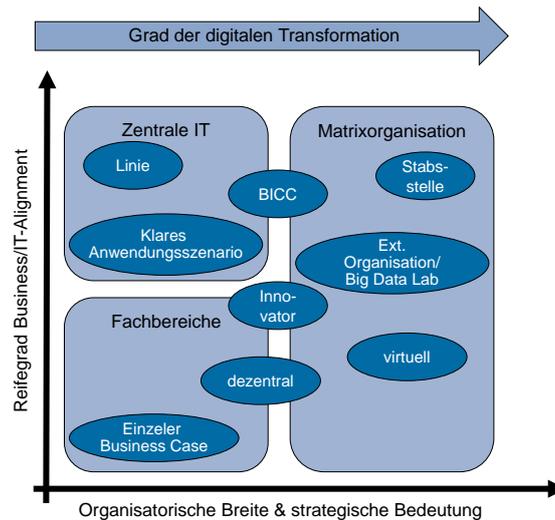


Abb. 5-2 Digitale Transformation

5.2.2 Big Data als Teil der zentralen IT-Organisation

Auf den ersten Blick erscheint es logisch, die Verantwortung für Big Data in der zentralen IT und wenn vorhanden hier innerhalb der BI-Organisation anzusiedeln. Diese stellt Infrastruktur und Analyseressourcen bereit, die mit der bestehenden IT-Architektur konform gehen, und ermöglicht den Fachbereichen neue Datenquellen und Analyseverfahren. Die Verantwortung für Daten und Betrieb bleibt damit in einer Hand, eine stabile, sichere und in die Gesamtorganisation eingebettete Big-Data-Infrastruktur wird gewährleistet.

In diesem Fall wäre also die Technologie der Treiber, die IT der Owner und der Fachbereich und die internen Sponsoren müssten entsprechend befähigt werden. Im Falle einer eher reaktiven IT-Organisation besteht hier das Risiko, dass die Geschäftspotenziale nicht über die eher technologisch geprägte Kommunikation der IT in den Fachabteilungen erkannt werden.

Diese Organisationsform eignet sich deshalb eher für Unternehmen mit einer recht hohen fachlichen Affinität innerhalb der IT, für Big-Data-Projekte mit relativ eindeutigem Business Case, die zum Beispiel innerhalb der BI-Organisation betrieben werden, und für Unternehmen mit einheitlichen Data-Governance-Anforderungen. Gerade für Unternehmen mit einer definierten BI-Strategie und einem Business Intelligence Competency Center (BICC) als organisatorischem Überbau erweitert Big Data den Fokus von BI und gewährleistet die Vernetzung innerhalb der Organisation [Gansor/Totok 2015, S. 16].

Es besteht aber das Risiko, dass sich die Big-Data-Aktivitäten gerade durch die zentrale Organisation zu weit vom eigentlichen Geschäftsbetrieb in den verschiedenen Fachbereichen entfernen und unterschiedliche, fachbereichsspezifische Potenziale und neue Geschäftsfelder nicht entdeckt werden. Darüber hinaus besteht die Gefahr von Akzeptanzproblemen im Fachbereich durch die Ownership in der IT, die sich am Ende bis zum Management hin ausbreiten können.

5.2.3 Big Data in Verantwortung der Fachbereiche

Eine weitere Organisationsmöglichkeit wäre der Fachbereich als Owner von Big-Data-Projekten. Es gäbe dann keine übergreifende, zentrale Big-Data-Organisation, sondern einzelne Fachbereiche würden Big-Data-Projekte nach ihren individuellen und teilweise sehr unterschiedlichen Anforderungen treiben, zum Beispiel Marketing oder QS in der Produktion. Die Fachbereiche kennen ihre Anforderungen im Detail und sind bestens dafür geeignet, neue Innovationen und Geschäftsfelder durch Big Data zu entwickeln. Die Fachbereiche sind in diesem Fall Budgetverantwortliche und beauftragen die Projekte intern oder auch extern. Die IT hat in diesem Szenario eine reine Delivery-Funktion.

Die Vorteile liegen in einem klar definierten Primärnutzen, zum Beispiel durch eine Social-Media-Analyse im Marketing. Der Fachbereich kann die Anforderungen besser beschreiben und durch den hierdurch definierten Business Case leichter einen Sponsor im Management finden, um die Projekte zu finanzieren. Einzelne Big-Data-Projekte können somit deutlich schneller umgesetzt werden.

Dieses Szenario erscheint gut geeignet für Unternehmen, die nur in einzelnen Fachbereichen klare Big-Data-Projekte haben – ohne eine strategische Gesamtbedeutung für die ganze Organisation – und wenn sich die IT eher als Delivery-Organisation sieht. Voraussetzung ist eine gewisse Innovations- und Analytik-Affinität im Fachbereich.

Als Alternativszenario ist diese Organisationsform auch für Big-Data-Pilotprojekte denkbar, die aus einzelnen fachlichen Anforderungen heraus getrieben werden und im Vorfeld einem ausgewählten Kreis an Anwendern zur Verfügung gestellt werden sollen, damit sich diese einen ersten Überblick verschaffen können. Das Schadensrisiko dürfte hierbei niedrig sein, da es sich um einen ausgewählten Anwenderkreis handelt. Im Rahmen eines Lean-Startup-Ansatzes lässt sich ein Pilotprojekt schneller erproben, dies entspricht dem gängigen Verfahren einer Pilotphase bei der Einführung neuer Applikationen. Vor einem breiteren Rollout im Unternehmen kann man erste Erfahrungen nutzen. Nach erfolgreicher Pilotphase ist die Verantwortlichkeit im Unternehmen neu zu organisieren.

Risiken entstehen bei diesem organisatorischen Ansatz durch die meist fehlende technische und analytische Kompetenz im Fachbereich und die Akzeptanz durch die IT als reine Delivery-Organisation. Durch die dezentrale Organisation in den einzelnen Fachbereichen existiert keine konsistente Infrastruktur, was zu erhöhten Administrations- und Wartungsaufwänden führt. Außerdem findet kein fachbereichsübergreifender Austausch statt. Hier besteht zusätzlich das Risiko, dass Fachbereiche untereinander konkurrieren und ihre Daten und Analysen sichern. Es gibt keinen *Single Point of Truth* und keine einheitliche Governance.

5.2.4 Big Data als Querschnittsorganisation

Die beste Einbindung aller Interessengruppen innerhalb eines Unternehmens wird durch eine Querschnittsorganisation erreicht, in der sowohl die Kompetenz der Fachbereiche als auch die der IT vertreten sind. Hier empfiehlt sich der Aufbau eines Big Data bzw. Analytics Competence Center mit entsprechender Durchsetzungsverantwortung. In Anlehnung an [Dittmar/Oßendoth 2012, S. 60 ff.], die die Aufgaben der BI-Organisation in die vier Dimensionen Management und Governance, Kundenmanagement, Daten- und Informationsmanagement sowie Applikations- und Infrastrukturmanagement einteilen, kann dies auf die Aufgaben der Big-Data-Organisation übertragen werden.

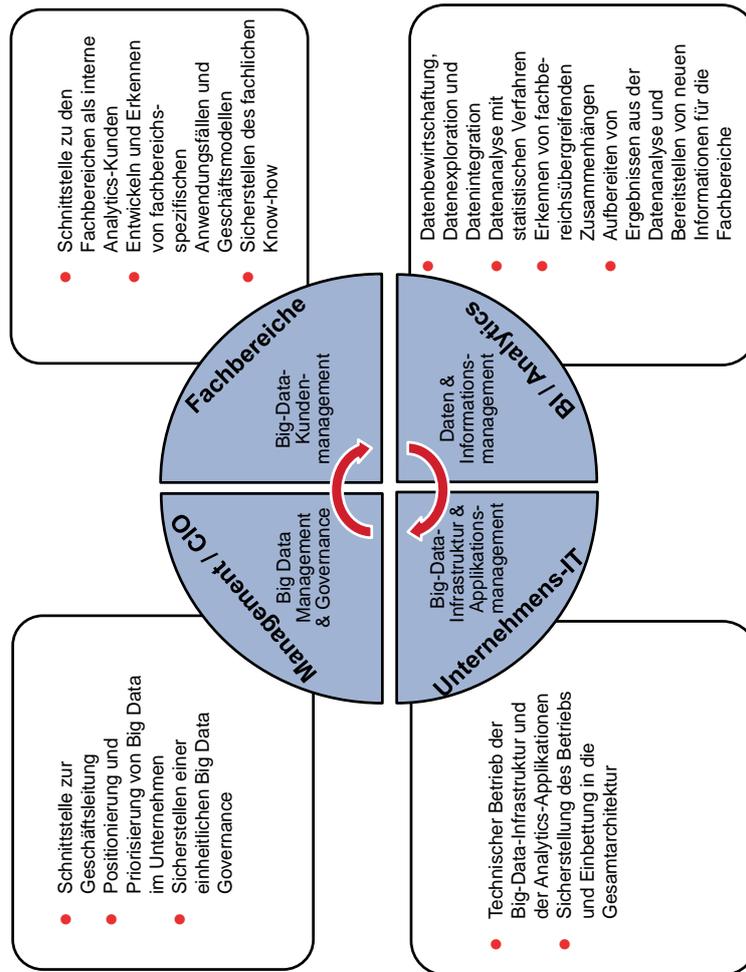


Abb. 5-3 Big-Data-Organisation

Diese Organisationsform ermöglicht den bestmöglichen Austausch unter den beteiligten Parteien und erzeugt gleichzeitig eine gute Sichtbarkeit für das Thema in der gesamten Organisation. Voraussetzung ist ein entsprechender Support aus der Geschäftsleitung, die Big Data für sich als strategisches Unternehmensziel versteht und bereit ist, Ressourcen in eine übergeordnete Big-Data-Organisation zu investieren, bevor der Business-Nutzen für alle Bereiche klar definiert ist. Dieser Ansatz stellt sicher, dass die Potenziale in Bezug auf Ausbau und Entwicklung von Geschäftsmodellen durch Big Data bestmöglich genutzt werden. Natürlich gibt es nicht die *eine* Querschnittsorganisation, sondern verschiedene Möglichkeiten für ihre Einbindung. Im Folgenden gehen wir kurz auf die Ausprägung als Stabsstelle, eigene und virtuelle Organisationseinheit ein.

5.3 Big Data als Stabsstelle

In einer primär hierarchisch aufgebauten Organisation kann die Big-Data-Verantwortung gut als Stabsstelle an der Geschäftsleitung aufgehängt werden. Diese Form der Einbindung sichert eine Kommunikation über Abteilungsgrenzen, entlastet die Führungsverantwortlichen in den Abteilungen und gewährleistet eine ganzheitliche Sicht auf Big Data. Es ergeben sich allerdings die üblichen Risiken von Stabsstellen, so haben diese faktisch Entscheidungsgewalt ohne Entscheidungsbefugnis und somit auch ohne die Verantwortung hierfür tragen zu müssen. Eine weitere Herausforderung ist die Definition der Budgethoheit für die Big-Data-Projekte (Fachbereich, IT). Die Big-Data-Organisation wird in diesem Modell als ein reines Costcenter dargestellt [Scheuch/Gansor/Ziller 2012, S. 213].

5.4 Big Data als eigene Organisationseinheit

Das BICC wird als eigenständige Einheit in die Gesamtorganisation integriert und idealerweise bei einem Chief Data Officer oder beim CIO aufgehängt. Wichtig ist, dass die Big-Data-Einheit über eigenes Budget und über eigenes Personal verfügt, um hier in der Umsetzung autark agieren zu können. Das Team setzt sich interdisziplinär mit IT (Einbindung der bisherigen BI-Organisation), Business und speziellem Big-Data- und Analytics-Know-how zusammen. Die Fachbereiche können ihre Big-Data-Aufgabenstellungen bei der Big-Data-Organisation einbringen

und diese agiert als interner Dienstleister für die Fachbereiche. Im Rahmen einzelner Projekte können Mitarbeiter der beauftragenden Fachabteilungen operativ fachlich eingebunden werden. In den meisten Fällen wird diese Organisationsform eine reine Costcenter-Struktur haben, da die Wertschöpfung innerhalb der Fachbereiche liegt. In einer Idealwelt kann die Wertschöpfung der Big Data Business Cases beziffert werden und die Big-Data-Organisation entwickelt sich zu einem Big-Data- bzw. Analytics-Profitcenter. In der Ausprägung als eigene Organisationseinheit könnte noch unterschieden werden in eine Big-Data/Analytics-Delivery- und eine Data-Lab-Struktur.

In ersterer setzt die Big-Data-Einheit Business Cases aus den Fachbereichen um, entwickelt diese weiter und liefert gewonnene Erkenntnisse an die Fachbereiche. Durch die zentrale Organisation ist ein abteilungsübergreifender Blick auf das Unternehmen und die verschiedenen Datenquellen sichergestellt.

In der Ausprägung als Data Lab steht primär die Entwicklung von neuen Business Cases im Vordergrund. Im Data Lab entwickeln die Mitarbeiter schnell und unbürokratisch Ideen, setzen sie in Laborumgebung als Softwareprototypen um und erproben diese. Dazu steht ihnen modernste Hard- und Software zur Verfügung. Mitarbeiter der Konzern-IT kooperieren dabei eng mit externen Partnern. Gehen einzelne Big-Data-Prototypen in den Regelbetrieb, werden diese organisatorisch außerhalb des Labs betrieben. Das Big Data Lab ist hier in erster Linie Innovator. Das Big Data Lab kann organisatorisch aber auch Teil der externen Big-Data-Organisation sein.

Kriterium	Vorteile	Nachteile
Personal	Optimale Ausstattung mit Personal und den erforderlichen Kompetenzen	Mögliche Spannung zwischen dem Big-Data-Team und den operativen Einheiten
Durchsetzungsvermögen	Nähe zur Geschäftsleitung sorgt für optimale Durchsetzungsmöglichkeiten	Entfernung von den Fachbereichen birgt die Gefahr von Akzeptanzproblemen und <i>Elfenbeinturmsicht</i>
Verantwortung	Klare Verantwortlichkeit	Machtkonzentration kann zu Machtkämpfen mit den Fachbereichen führen
Kommunikation	Klare Organisation hilft bei der Kommunikation	Die Mannschaft in den Fachbereichen wird unter Umständen nicht erreicht
Erfolgsmessung	Zentrale Organisation hilft bei der Erfolgsmessung	Kosten und Budget liegen im Big-Data-Team, die Wertschöpfung entsteht im Fachbereich
Komplexität	Transparente und eindeutige Strukturen	Macht/Einfluss des Big-Data-Teams
Governance	Zentralistische Organisation unterstützt Umsetzung der Strategie und Ziele	Nicht nah genug an der Realität im Fachbereich?
Kosten	Klare Kostenstrukturen	Personal-, Infrastruktur- und Betriebskosten ohne eigene Wertschöpfung

Tab. 5-1 Vor- und Nachteile von Big Data als eigene Organisationseinheit

5.5 Big Data als virtuelles Team

Wird das BICC virtuell organisiert, verbleiben die beteiligten Personen in ihren Fachabteilungen und werden für Big-Data-Projekte ganz oder teilweise freigestellt. Die Zusammensetzung erfolgt als Querschnitt über die verschiedenen für das Projekt benötigten Organisationseinheiten. (Fachbereich, IT etc.). Das Team kann so je nach Projektanforderungen skaliert werden. Bestimmte Big-Data-spezifische Rollen (z.B. Data Scientist) sind aufgrund der speziellen fachlichen Anforderungen idealerweise mit Vollzeitmitarbeitern in der Big-Data-Organisation zu besetzen. Weitere werden nach Bedarf mit Personen besetzt, die für die Projektzeit freigestellt werden. Hierdurch wird erreicht, dass die Projektmitarbeiter aus den unterschiedlichsten Bereichen zusammenarbeiten. Im Zeitverlauf lässt sich das Team flexibel und anforderungsgerecht variieren. Risiken bestehen darin, dass es in der Weisungsbefugnis und zwischen Projekt- und Originärrolle zu Konflikten kommen kann [Scheuch/Gansor/Ziller 2012, S. 215 ff.].

Es gibt keinen Königsweg für die organisatorische Einbindung von Big Data. Wichtig ist, dass die Unternehmen die individuelle und strategische Bedeutung von Big Data für die Gesamtunternehmung bewerten und in die Organisation einfließen lassen. In der Praxis wird dies oft auch ein Entwicklungszyklus sein, in dem nach einer Anfangsinitiative die strategische Bedeutung zunimmt und sich damit auch die organisatorische Einbindung verändert. Je nach Reifegrad sollten die bestehenden BI-Strukturen die Grundlage des Analytics-Teams bilden. Die nächste Evolutionsstufe des BICC wird das Big Data/Analytics Competence Center!

5.6 Literatur

[Casey/Krishnamuthy/Abezgauz 2013] Casey, T.; Krishnamuthy, K.; Abezgauz, B.: Who should own Big data? Strategy & Business vom 12.08.2013, <http://www.strategy-business.com/article/00211?gko=44b8e> (letzter Abruf: 17.02.2016).

[Detecon/Bitcom 2011] Detecon; Bitcom: Studie IT-Organisation 2015. <https://www.detecon.com/sites/default/files/DEB-IT-Organisation-2015-Fit-fuer-die-Zukunft-2011.pdf> (letzter Abruf: 17.02.2016).

[Dittmar/Oßendoth 2012] Dittmar, C.; Oßendoth, V.: Die Organisatorische Dimension von BI. In: Chamoni, P.; Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Springer/Gabler, 2012, S. 60 ff.

[Hungenberg 2014] Hungenberg, H.: Strategisches Management in Unternehmen. 8. Auflage, Springer/Gabler, 2014.

[Madsudeira 2014] Madsudeira, K.: Key skills needed by big data engineers. Dice.com, 21.08.2014, <http://insights.dice.com/2014/08/21/key-skills-needed-big-data-engineers/> (letzter Abruf: 17.02.2016).

[Scheuch/Gansor/Ziller 2012] Scheuch, R.; Gansor, T.; Ziller, C.: Master Data Management. 1. Auflage, dpunkt.verlag, 2012.

6 Neue Rollen und Skillsets im Big-Data-Kontext

Lars Tams

Den (Alleskönner) Data Scientist gibt es nicht und für erfolgreiche Big-Data-Projekte wird er auch nicht gebraucht.

Big Data stellt neben organisatorischen Fragestellungen auch neue Kompetenzanforderungen in den Bereichen Business-, Daten-, IT- und Analysefähigkeiten an die Projektmitarbeiter. Es gibt in aktuellen Veröffentlichungen zahlreiche Rollenbeschreibungen in verschiedenen Ausprägungen. Besonders der Data Scientist wird hier als Big-Data-Wunderwaffe herausgestellt [Madsudeira 2014].

6.1 Aufgaben und Rollen des Data Scientist

Die Aufgabe des Data Scientist ist das Erkennen und Umsetzen von komplexen analytischen Zusammenhängen. Er übernimmt die Suche nach sinnvollen Zusammenhängen im Big-Data-Universum, hierfür erstellt er komplexe analytische Modelle und interpretiert die Analyseergebnisse [Gansor/Totok 2015, S. 177]. Seine Rolle erfordert eine breite Palette von interdisziplinären Kompetenzen, von komplexen statistischen Analyseverfahren über Kenntnisse von Werkzeugen des Big-Data-Systems wie zum Beispiel Hadoop, Kenntnisse im Umgang mit relationalen Datenbanken, Machine Learning, Auswerten von unstrukturierten und textuellen Daten, Datenvisualisierung bis zu Fachkenntnissen der Betriebswirtschaftslehre, um die Analyseergebnisse korrekt zu interpretieren. Das Problem ist, dass dieses Anforderungsprofil in der Realität überhaupt nicht existent ist. Das ist auch nicht weiter kritisch, viele Kompetenzen sind bereits im Unternehmen vorhanden. So sind zum Beispiel Datenmanagement-Skills innerhalb der BI-Organisation bereits vorhanden, müssen aber natürlich um Big-Data-Technologien erweitert werden. Leitet man die benötigten Rollen und Kompetenzen an der Big-Data-Wertschöpfung her, kann das in Tabelle 6-1 angegebene Rollenmodell definiert werden [Thomsen/Bleiholder 2015].

Big-Data-Wert-schöpfungsstufe	Big-Data-Rolle	Aufgaben	Ressourcen
Datenakquisition	Enterprise-Architekt ETL-Entwickler	Sammeln und Bereitstellen von Daten aus heterogenen Datenquellen (Sensor, Web, ERP etc.)	Vorhanden in der BI-Organisation
Datenmodellierung	DB-Administrator Data Engineer Data-Architekt	Aufbereiten von Daten aus heterogenen Quellen für Analyseverfahren	Vorhanden in der BI-Organisation
Datenanalyse	Statistiker SW-Entwickler	Durchführen von komplexen analytischen Verfahren, Anwenden von Big-Data-Technologien (z.B. MapR, Hadoop)	Statistik-Know-how wird in den meisten Fällen neu geschaffen werden müssen. Softwareentwickler sind in Kern-IT vorhanden.
Dateninterpretation	Domain-Experte	Interpretation und Überprüfen der Analyseergebnisse und Business-Case-Sicht	Diese Rolle muss aus dem Fachbereich besetzt werden, ist vorhanden, bedarf aber einer Affinität für Big Data und Analyseverfahren.

→

Big-Data-Wertschöpfungsstufe	Big-Data-Rolle	Aufgaben	Ressourcen
Reporting/ Exploration	BI-Entwickler Data Steward	Daten- visualisierung Aufbereiten der <i>neuen</i> Informationen für die Fachbereiche	Vorhanden in der BI-Organisation

Abb. 6-1 Big-Data-Rollen

Der Data Scientist ist eine gute Umschreibung für das Gesamtbild der Rollen, die für eine Big-Data-Organisation benötigt werden. In der Umsetzung teilt sich die Vielzahl der Aufgaben in verschiedene Rollen auf. Die bestehende BI-Organisation kann viele hiervon bereits abdecken, insofern erscheint es nur logisch, dass diese eine tragende Rolle in der Führung und dem Aufbau eines Big Data Competence Center und der Umsetzung von Big-Data-Projekten einnimmt. Diese unterschiedlichen Rollen aus IT, Fachbereichen und Betrieb sollten idealerweise die Grundlage für ein umfassendes Big Data/Analytics Competence Center bilden, das die unterschiedlichen Anforderungen und Sichtweisen auf Big Data widerspiegelt.

6.2 Literatur

[Gansor/Totok 2015] Gansor, T.; Totok, A.: Von der Strategie zum BICC. 2. Auflage, dpunkt.verlag, 2015.

[Thomsen/Bleiholder 2015] Thomsen, C.; Bleiholder, J.: 10 Big-Data-Mythen geradegerückt. IT-Management 29.10.2015, <http://www.it-daily.net/analysen/11637-geradegerueckt-10-big-data-mythen> (letzter Abruf: 10.03.2016).

unabhängig.
praxisorientiert.
persönlich.



Die Plattform für Ihre BI-Weiterbildung

- Community
- Konferenzen
- Seminare
- BI-SPEKTRUM
- Fachbücher
- Online Wissensportal

www.tdwi.eu

Jetzt Mitglied werden!
Informationen auf www.tdwi.eu





Carsten Dittmar, Carsten Felden, Ralf Finger,
Rolf Scheuch, Lars Tams

Big Data

Was ist oder was sind Big Data? Wie kann man Big Data nutzen? Was bedeutet es, wenn man Big Data nutzen möchte? Wie verändern sich Architekturen durch Big Data? Wie gestalten sich Big-Data-Projekte? Wer verantwortet Big Data im Unternehmen? Wer ist Treiber? Wer profitiert? Wer ist Budget Owner? Wen brauchen wir? Viele Fragestellungen begleiten das Thema Big Data. In der breiten öffentlichen Diskussion zu Big Data löst die damit verbundene Idee, aus der Vielzahl und Vielfalt der verfügbaren Daten schnell werthaltige Informationsschätze zu heben, sehr viele positive, aber durchaus auch negative Reaktionen aus. Das erhöht den Bedarf an inhaltlicher Orientierung.

Im Namen des TDWI Germany e.V. haben die Autoren aus ihrem Know-how die Themen zusammengestellt, welche die Breite des Big-Data-Themenkomplexes reflektieren, und damit verbundene Handlungsempfehlungen erarbeitet. Damit spricht diese Broschüre insbesondere Leser aus der Praxis an, um ihnen beim Management eigener Big-Data-Projekte Unterstützung zu geben.

Art.-Nr.: 077.95739

www.dpunkt.de