

Einflüsse auf den Implementierungserfolg von NoSQL-Systemen: Erkenntnisse einer quantitativ-empirischen Untersuchung

Patrick Cato · Simon Brumm · Philipp Gölzer · Walter Demmelhuber

Eingegangen: 31. März 2016 / Angenommen: 24. Mai 2016 / Online publiziert: 23. Juni 2016
© Springer Fachmedien Wiesbaden 2016

Zusammenfassung NoSQL-Systeme sind eine neue Generation von Datenbanksystemen, mit denen sich eine Reihe von neuen und innovativen Anwendungsfällen realisieren lässt. Erfahrungsberichte von Praktikern zeigen jedoch, dass die Implementierung von NoSQL-Systemen in der Praxis oftmals komplexitätsbedingt scheitert. Ziel der Studie des Instituts für Wirtschaftsinformatik Nürnberg war es daher, die zentralen Einflussgrößen auf den Implementierungserfolg zu erfassen und empirisch zu validieren. Acht Faktoren haben einen signifikanten Einfluss auf den Implementierungserfolg von NoSQL-Systemen: ausreichende Ressourcenausstattung des Projekts, datengetriebene Entscheidungskultur im Unternehmen, Datenqualität der Quellsysteme, Managementunterstützung, Data Science Fähigkeiten, Laboransatz zur Exploration des Anwendungsfalls, adäquate Technologiewahl sowie Berücksichtigung von datenschutzrechtlichen Aspekten in der Designphase (Privacy by Design).

Schlüsselwörter Big Data · NoSQL · Erfolgsfaktoren · Implementierungserfolg

Abstract NoSQL systems are a new generation of data base systems that enable the implementation of new and innovative use cases. However, experience reports from practitioners show that the implementation of NoSQL systems is a complex undertaking and many NoSQL projects fail. This paper summarizes the key results of a study that investigated the central factors that have an impact on the implementation success of NoSQL systems. The study has identified eight factors that have a significant impact: Sufficient resources, data-driven culture, data quality of the

P. Cato (✉) · S. Brumm · P. Gölzer · W. Demmelhuber
Lehrstuhl für IT-Management, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lange
Gasse 20, 90403 Nürnberg, Deutschland
E-Mail: patrick.cato@fau.de

sourcing systems, management support, data science skills, proof-of-concept phase, adequate technology selection and privacy by design.

Keywords Big data · NoSQL · Success factors · Implementation success

1 Einleitung

Die zunehmende Digitalisierung verschiedener Industrie- und Lebensbereiche hat in den letzten Jahren zu einer signifikanten Zunahme des generierten Datenvolums geführt (Gantz und Reinsel 2012). Insbesondere die großen Internetunternehmen stellten bereits früh fest, dass klassische, relationale Datenbanksysteme die anfallenden Daten nicht mehr effizient verarbeiten können (Schmarzo 2013). Daher wurden von einigen Internetunternehmen (z. B. Amazon, Ebay, Google) alternative Datenbanksysteme entwickelt, die sich vom relationalen Datenmodell lösen und unter anderem durch Parallelverarbeitung und abfrageorientierte Datenmodelle große und komplexe Datenmengen (Big Data) effizient verarbeiten können. Die Entwicklungserkenntnisse und Erfahrungen mit den neuen NoSQL Datenbanksystemen wurden größtenteils in Fachpublikationen (Chang et al. 2006; DeCandia et al. 2007; Laksham und Prashant 2010) offengelegt. Insbesondere die Erklärung des Big Data Frameworks Hadoop als Apache Top-Level-Projekte im Jahr 2008 führte zu einem regelrechten Hype um das Thema Big Data/NoSQL. So sind ab 2008 unzählige Publikationen erschienen, die sich mit neuen Anwendungsfällen von NoSQL-Systemen und Hadoop im Besonderen auseinandersetzen. Feststellungen wie „data is the new oil“ (World Economic Forum 2011, S. 5; Forbes 2012, o. S.) bzw. „data is the new gold“ (Europäische Kommission 2011, o. S.) sind hier charakteristisch für die Diskussion der Potenziale von NoSQL-Systemen. Insbesondere Erfahrungsberichte von Praktikern (Infochimps 2013; Hewlett-Packard Enterprise 2014) zeigen jedoch, dass NoSQL bzw. Big Data Projekte in Unternehmen häufig erfolglos abgebrochen werden.

Ziel der Studie des Instituts für Wirtschaftsinformatik war es daher, die Einflussfaktoren des Implementierungserfolgs von NoSQL-Systemen zu ergründen und Wissenschaftlern und Praktikern einen gesamtheitlichen Blick auf relevante „Erfolgsfaktoren“ im Kontext von NoSQL-Systemimplementierungen zu geben. Im vorliegenden Beitrag werden die Kernergebnisse der empirischen Studie dargestellt. Die gesamten Ergebnisse der Studie sind in Cato (2016) ausführlich dokumentiert.

2 Hintergrundinformationen zur Studie

Die Studie, aus der ausgewählte Ergebnisse präsentiert werden, wurde als Online Befragung im Zeitraum von 22.07.15 bis 30.09.15 durchgeführt. Zielgruppe der Studie waren Projektbeteiligte von NoSQL Projekten in deutschen Unternehmen. Ein NoSQL-Projekt wird im Rahmen dieses Beitrags als Implementierung eines NoSQL-Systems innerhalb eines Unternehmens verstanden. Der Implementierungsbegriff wird in Anlehnung an Müller (1990, S. 208) als „umfassender Gestaltungsansatz“

definiert. Das heißt, der Implementierungsbegriff bezieht sich auf den gesamten Entwicklungs- und Einführungsprozess und schließt somit die Abstimmung von technologischen und organisatorischen Komponenten ein.

Die Auswertung der Studie basiert auf 101 verwertbaren Fragebögen. An der Befragung nahmen unterschiedliche Projektsteakholder teil. Die Mehrheit der Befragten hatte eine beratende bzw. technische Funktion in dem NoSQL Projekt. So waren rund 31,7 % der Befragten als Berater in dem Projekt eingebunden. Rund 23 % bzw. 21 % der Studienteilnehmer hatte die Rolle eines Entwicklers bzw. Data Scientists.

An der Befragung nahmen überwiegend große Unternehmen teil. So hatten rund 43 % der Befragten eine Unternehmensgröße von über 5000 Mitarbeitern. 24 % der Befragten gaben an, dass ihr Unternehmen zwischen 501 und 5000 Mitarbeiter beschäftigt.

Die Stichprobe repräsentiert verschiedene Branchen. Die stärkste Gruppe ist mit 18 % der Groß- und Einzelhandel. Dieser Branche folgen mit 13 %, 12 % sowie 11 % die Branchen Banken/Versicherungen, Telekommunikation sowie Informati-onstechnik.

Die Probanden wurden ebenfalls gefragt, in welchen Fachabteilungen bzw. Fachbereichen die einzelnen NoSQL Projekte durchgeführt wurden. Die relative Mehrheit (37 %) der Projekte ist dem Bereich Marketing/Vertrieb zuzuordnen. 18 % der Probanden gaben an, dass das implementierte NoSQL-System innerhalb der IT-Abteilung eingesetzt wird. Ca. 8 % der Projekte wurden von den Probanden dem Bereich Produktion zugeordnet.

3 Ausgewählte Studienergebnisse

Im Folgenden werden die Kernergebnisse der Studie präsentiert. Zunächst wird in Abschn. 3.1 die Stichprobe weiter charakterisiert. In Abschn. 3.2 werden die Ergebnisse der kausalanalytischen Untersuchung dargestellt.

3.1 Deskriptive Ergebnisse

Mit den NoSQL-Systemen aus dem Sample werden unterschiedliche Zielsetzungen (siehe Abb. 1) verfolgt. So stimmen rund 21,8 % bzw. 33,7 % der Befragten vollkommen zu, dass mit ihrem NoSQL-System primär operative Prozesse bzw. Kosten optimiert werden. Interessant ist, dass lediglich 13,9 % der Befragten vollkommen zustimmen, ein neues Geschäftsmodell mit den NoSQL-Systemen zu realisieren. Dies zeigt, dass sich die Unternehmen aus dem Sample hauptsächlich auf Optimierungsprojekte konzentrieren.

Die befragten Unternehmen setzen unterschiedliche NoSQL-Systeme zur Erreichung der Projektziele ein (siehe Abb. 2). In 31,7 % der Projekte wurde das Framework Hadoop eingesetzt. Dem folgen mit 19,8 % bzw. 18,8 % MongoDB bzw. Amazon Dynamo.

Ferner wurden die Probanden gebeten, anzugeben welche Datentypen von den NoSQL-Systemen verarbeitet werden (siehe Abb. 3). Interessant ist, dass mehrheit-

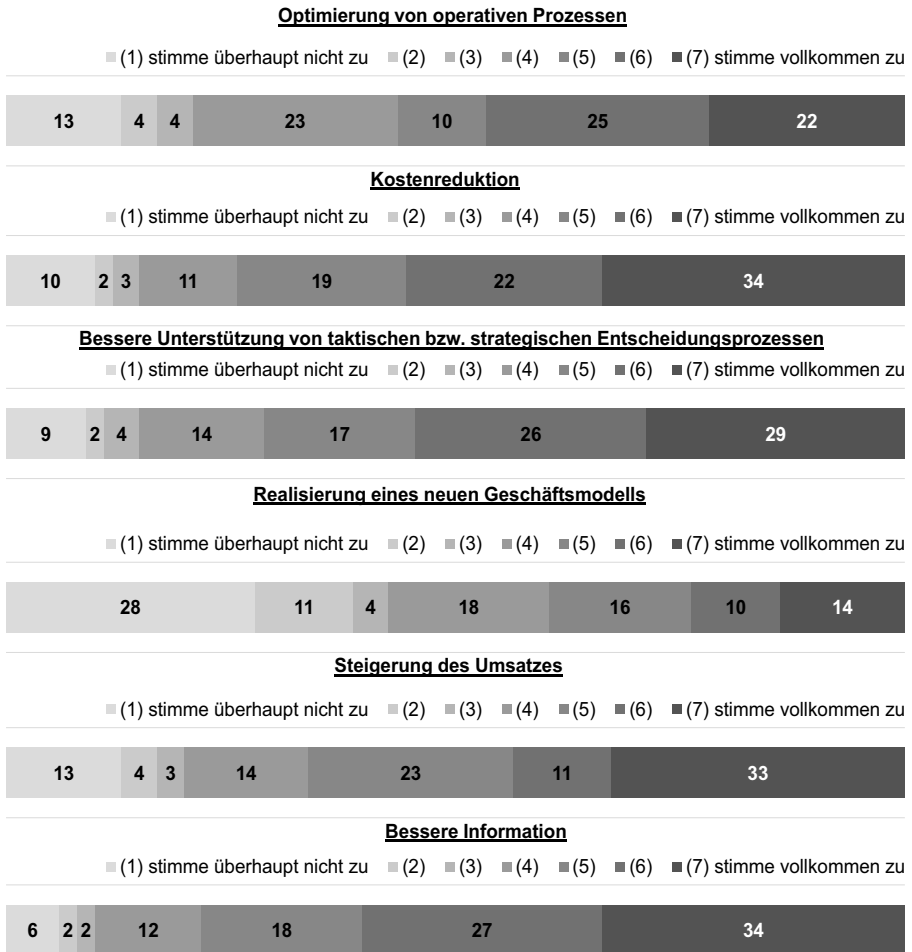
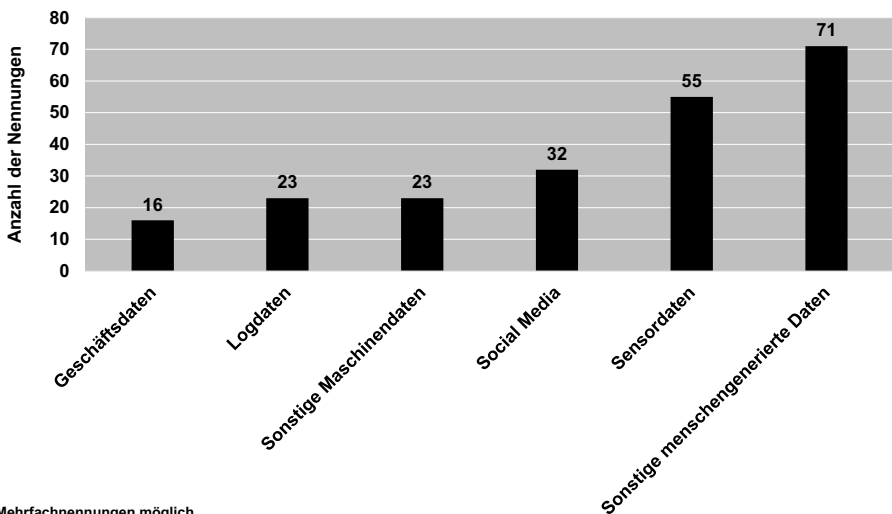
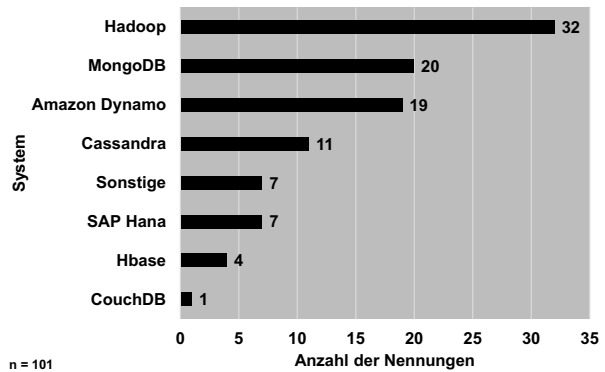


Abb. 1 Zielsetzung der NoSQL-Projekte (Cato 2016)

lich unstrukturierte Daten verarbeitet werden. Lediglich 16 der 101 NoSQL-Systeme verarbeiten klassische, strukturierte Daten.

3.2 Kausalanalytische Ergebnisse

Im Vorfeld der Studie wurden basierend auf einer qualitativen Inhaltsanalyse von Industriepublikationen sowie Experteninterviews (Cato et al. 2015) potenzielle Einflussfaktoren des Implementierungserfolgs exploriert. Die Probanden wurden zu jedem identifizierten Einflussfaktor gebeten, die Relevanz der einzelnen Faktoren für ihr Projekt auf einer siebenstufigen Likert-Skala zu bewerten. Ferner bewerteten die Studienteilnehmer den Implementierungserfolg des NoSQL-Systems. Basierend auf theoretischen Überlegungen wurde ein Hypothesenmodell formuliert und in ein Strukturgleichungsmodell überführt (siehe Abb. 4). Im Anschluss wurden die Wir-

Abb. 2 Verwendete NoSQL-Systeme (Cato 2016)**Abb. 3** Verarbeitete Datentypen der NoSQL-Systeme (Cato 2016)

kungszusammenhänge mittels einer varianzbasierten Strukturgleichungsanalyse untersucht. Hierfür wurde die Software SmartPLS 3 eingesetzt.

Im Folgenden werden ausschließlich die Wirkungszusammenhänge dargestellt, die statistisch signifikant sind (siehe Abb. 4). Hier wird ein Signifikanzniveau von $\alpha = 0,05$ zu Grunde gelegt. In Anlehnung an Chin (1998) wird erst ab einem Pfadkoeffizienten von 0,19 ein bedeutender Wirkungszusammenhang vermutet. In Abb. 4 sind zudem die Stärke der Pfadkoeffizienten (PK) sowie das Bestimmtheitsmaß (R^2) der abhängigen Variablen dokumentiert.

3.2.1 Einfluss auf den projektbezogenen Implementierungserfolg

Der projektbezogene Implementierungserfolg wird definiert als der wahrgenommene Erfolgsgrad hinsichtlich der Einhaltung von Kosten-, Zeit-, und Qualitätszielen. Auf diese Erfolgsdimension hat die Ausstattung des NoSQL-Projekts mit ausreichend

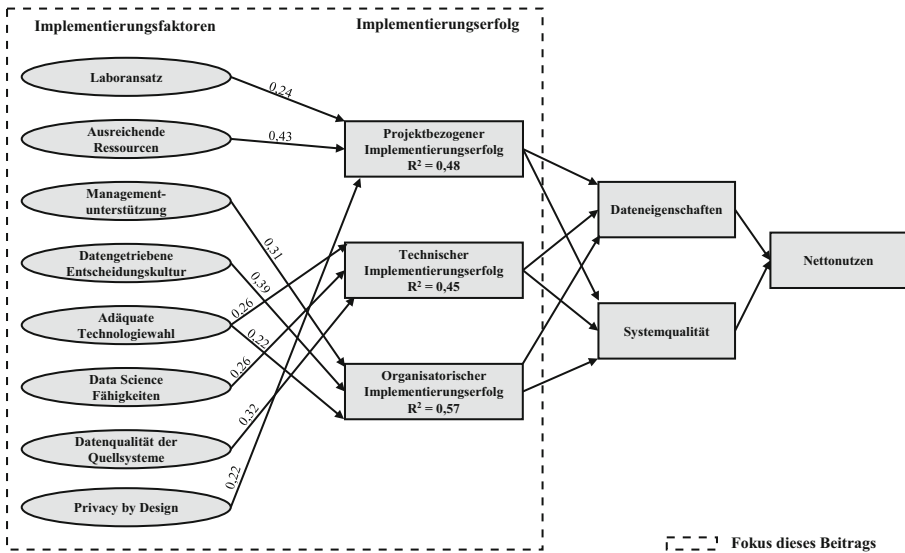


Abb. 4 Ergebnisse der Strukturgleichungsanalyse (Grafische Darstellung in Anlehnung an Wixom und Watson 2001)

personellen und finanziellen Mitteln den größten Einfluss. Hier wurde ein moderat-starker Einfluss (PK = 0,43) gemessen. Einen ebenfalls bedeutenden Einfluss hat das methodische Vorgehen im Projekt. So hat die Wahl eines Laboransatzes ebenfalls einen bedeutenden Einfluss (PK = 0,24) auf den projektbezogenen Implementierungserfolg. Beim Laboransatz handelt es sich um eine spezielle Vorgehensweise zur Exploration des Anwendungsfalls während der Projektvorbereitungsphase. Hierbei sind folgende Merkmale charakteristisch für den Laboransatz:

- Der Anwendungsfall wird iterativ entwickelt.
- Es wird ein Prototyp entwickelt bzw. ein Machbarkeitsnachweis (Proof-of-Concept) erbracht.
- Den Projektbeteiligten wird ausreichend Freiraum zum Experimentieren gegeben.

Es wurde ebenfalls untersucht, welchen Einfluss die Berücksichtigung von datenschutzrechtlichen Aspekten in der Konzeptionsphase des Systems (Privacy by Design) hat. Für diesen Faktor konnte ebenfalls ein bedeutender Wirkungszusammenhang (PK = 0,22) nachgewiesen werden.

Die Variable projektbezogener Implementierungserfolg hat ein Bestimmtheitsmaß von 0,48. Somit erklären die unabhängigen Variablen rund 50 % der Varianz des Zielkonstrukts. Dieser Wert kann nach Chin (1998) als gut bewertet werden.

3.2.2 Einfluss auf den technischen Implementierungserfolg

Der technische Implementierungserfolg wird definiert als der wahrgenommene Erfolgsgrad hinsichtlich der Lösung von technischen Problemen. Auf dieses Zielkonstrukt haben drei Faktoren einen signifikanten Einfluss. Den größten Einfluss (PK =

0,32) auf den technischen Implementierungserfolg hat die Datenqualität der Quellsysteme. Auch im Kontext von Big Data hat die Beschaffenheit der Quelldaten einen bedeutenden Einfluss auf die technische Implementierung. So führen beispielsweise korrupte oder unvollständige Quelldaten zu einer Reihe von technischen Problemen. Ferner wurde ein signifikanter Einfluss der Data Science Fähigkeiten des Projektteams auf das Zielkonstrukt festgestellt. So treten signifikant weniger technische Probleme in Projektteams auf, die über fundierte Data Science Fähigkeiten verfügen. In der Studie wurden Data Science Fähigkeiten unter anderem durch folgende Merkmale erfasst:

- Geschäftswissen des Projektteams.
- Kompetenzen in Statistik sowie in der Auswertungsalgorithmik des Projektteams.
- Selbsteinschätzung der Data Science Fähigkeiten durch den Studienteilnehmer.

Des Weiteren wurde ein bedeutender Einfluss ($PK = 0,26$) der Technologiewahl auf das Zielkonstrukt festgestellt. So hat insbesondere die Auswahl eines für den Anwendungsfall geeigneten Datenbanksystems bzw. Datenmodells einen erheblichen Einfluss auf den technischen Implementierungserfolg.

Die Variable technischer Implementierungserfolg hat ein Bestimmtheitsmaß von 0,44. Die Variable projektbezogener Implementierungserfolg hat ein Bestimmtheitsmaß von 0,48. Somit erklären die unabhängigen Variablen rund 45 % der Varianz des Zielkonstrukts. Dieser Wert kann nach Chin (1998) als gut bewertet werden.

3.2.3 Einfluss auf den organisatorischen Implementierungserfolg

Der organisatorische Implementierungserfolg wird definiert als wahrgenommener Grad der Integration von NoSQL-Systemen in betriebliche Prozesse sowie als wahrgenommener Grad der Akzeptanz von NoSQL-Systemen in der Organisation. Auf dieses Zielkonstrukt haben drei Faktoren einen signifikanten Einfluss. Den größten Einfluss auf den organisatorischen Implementierungserfolg hat die Unternehmenskultur. So wurde ein starker Wirkungszusammenhang ($PK = 0,39$) zwischen der datengetriebenen Entscheidungs- und Unternehmenskultur und dem organisatorischen Implementierungserfolg festgestellt. Eine datengetriebene Entscheidungs- und Unternehmenskultur zeigt sich in der Regel durch folgende Merkmale:

- Eine Vielzahl von operativen und strategischen Entscheidungen basiert auf Datenanalysen.
- Die Menschen im Unternehmen stehen Datenanalysen offen gegenüber.
- Die Menschen im Unternehmen haben eine positive Einstellung gegenüber Datenanalysen.

Ebenfalls einen moderat-starken Einfluss ($PK = 0,31$) hat das Management. So wiesen die NoSQL-Projekte mit hoher Managementunterstützung signifikant bessere Werte für den Implementierungserfolg auf als Projekte mit geringerer Managementunterstützung. In diesem Zusammenhang wurde die Stärke der Managementunterstützung durch folgende Kriterien erfasst:

- Das Management lässt sich regelmäßig über den Projektfortschritt unterrichten.
- Das Management wirkt unterstützend, das heißt es unterstützt das Projekt fachlich und sachlich.

In dem Sample wurde ebenfalls ein bedeutender Wirkungszusammenhang ($PK = 0,22$) der Technologiewahl auf das Zielkonstrukt festgestellt. Dies kann dadurch erklärt werden, dass viele NoSQL-Projekte in den Unternehmen oftmals Pilotcharakter haben. Bei falscher Technologiewahl stellen sich die gewünschten Erfolge nicht ein und in der Konsequenz entwickelt sich in den Unternehmen eine gewisse Skepsis gegenüber NoSQL-Systemen.

Die Variable projektbezogener Implementierungserfolg hat ein Bestimmtheitsmaß von 0,57. Die Variable projektbezogener Implementierungserfolg hat ein Bestimmtheitsmaß von 0,48. Somit erklären die unabhängigen Variablen rund 60 % der Varianz des Zielkonstrukts. Dieser Wert kann nach Chin (1998) als gut bewertet werden.

4 Fazit und Schlussfolgerungen für die industrielle Praxis

Die empirische Studie zeigt, dass das Thema Big Data/NoSQL kein reines IT-Thema ist, sondern vielmehr organisatorische und technische Faktoren gleichermaßen auf den Implementierungserfolg wirken. Eine erfolgreiche Implementierung eines NoSQL-Systems erfordert daher sowohl neue technische als auch organisatorische Herangehensweisen. Aus den Ergebnissen lässt sich eine Reihe von Hinweisen für die industrielle Praxis ableiten. Diese werden im Folgenden dargestellt.

4.1 Gewährung von Freiräumen und Managementunterstützung

Eine zentrale Herausforderung jedes NoSQL-Projekts ist die Identifikation des Anwendungsfalls. Da bis dato nur sehr wenige Standardlösungen verfügbar sind, muss der Anwendungsfall in der Regel vom Projektteam exploriert werden. Hier hat sich in der Praxis ein Laboransatz bzw. ein exploratives Vorgehen bewährt. Das heißt, dass der Anwendungsfall sukzessive durch ein experimentelles Vorgehen entwickelt wird. Hierzu suchen die Projektbeteiligten (z. B. IT-Vertreter, Fachbereichsvertreter, Berater) im ersten Schritt gemeinsam Ideen für mögliche Anwendungsfälle. Dem schließt sich eine Bewertung der technischen Machbarkeit an. In der Regel wird hierzu der Zugriff auf diverse Datenquellen sowie die Qualität der Quelldaten bewertet. Ebenfalls wird ein Prototyp entwickelt, um die technische Realisierbarkeit abzuschätzen. Erst nach einem erfolgreichen Machbarkeitsnachweis sollte ein konkreter Anwendungsfall implementiert werden.

Die Untersuchungsergebnisse der empirischen Studie lassen darauf schließen, dass für eine erfolgreiche Implementierung eines Big Data Systems eine hinreichende Unterstützung des Managements notwendig ist. Dem Thema Big Data sollte daher seitens des Managements ausreichend Priorität eingeräumt und die notwendigen Ressourcen bereitgestellt werden (Cato 2016).

4.2 Adäquate Technologiewahl

Es ist wichtig, zu verstehen, dass für den konkreten Anwendungsfall ein passendes Datenbanksystem bzw. Datenmodell gewählt werden sollte. In der Praxis ist es nicht unüblich, dass sich bereits zu Projektbeginn die Projektbeteiligten auf eine Technologie – oftmals Hadoop – festlegen. Insbesondere die intensive Berichterstattung über Hadoop hat hier sicherlich zu einer gewissen Überbewertung bzw. übertriebenen Euphorie beigetragen. So lassen sich einige Anwendungsfälle (z. B. Echtzeitsysteme) mit einem batchorientierten Hadoop System nicht umsetzen (Cato 2016).

4.3 Entwicklung einer datengetriebenen Entscheidungs- und Unternehmenskultur

Die Auswertung zeigt, dass sich die Entscheidungs- und Unternehmenskultur signifikant auf den organisatorischen Implementierungserfolg auswirkt. Das heißt, dass sich in solch einer Unternehmenskultur NoSQL Systeme schneller und erfolgreicher in betriebliche Prozesse integrieren lassen. Hieraus lässt sich schließen, dass für eine erfolgreiche Integration von Big Data Systemen in den betrieblichen Leistungserstellungsprozess eine datengetriebene Entscheidungs- und Unternehmenskultur hilfreich ist. Insbesondere etablierte Unternehmen mit einer konservativen Unternehmenskultur stehen hier vor der Herausforderung, einen Kulturwandel einzuleiten, um nachhaltig die Potenziale von NoSQL Systemen zu erschließen. Eine der Schlüsselherausforderungen besteht darin, Entscheidungsprozesse anzupassen und im Unternehmen ein Bewusstsein zu schaffen, dass Datenprodukte bzw. Data-services einen wesentlichen Wertschöpfungsbeitrag leisten können (Cato 2016).

Literatur

- Cato P (2016) Einflüsse auf den Implementierungserfolg von Big Data Systemen – Ergebnisse einer inhalts- und kausalanalytischen Untersuchung. Verlag Dr. Kovac, Hamburg
- Cato P, Gölzer P, Demmelhuber W (2015) An investigation into the implementation factors affecting the success of big data systems. In: 11th Int Conf Innov Inf Technol. S 134–139
- Chang F, Dean J, Ghemawat S et al (2006) Bigtable: A distributed storage system for structured data. In: OSDI'06: Seventh Symposium on Operating System Design and Implementation, Seattle, WA, November 2006, S 205–218
- Chin W (1998) The partial least squares approach for structural equation modeling. In: Marcoulides G (Hrsg) Mod. Methods Bus. Res. Lawrence Erlbaum Associates, Mahwah, S 295–336
- DeCandia G, Hastorun D, Jampani M et al (2007) Dynamo: amazon's highly available key-value store. *Acm Sigops Oper Syst Rev* 41(6):205–220
- Europäische Kommission (2011) Data is the new gold. http://europa.eu/rapid/press-release_SPEECH-11-872_en.pdf
- Forbes (2012) Is data the new oil? <http://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/>
- Gantz J, Reinsel D (2012) The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the Far East. Idc Whitepaper. <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>. Zugegriffen: 21 Juni 2016
- Hewlett-Packard Enterprise (2014) HP bringt Big-Data-Lösung für das Rechenzentrum. <http://www8.hp.com/de/de/hp-news/press-release.html?id=1426189#.VpEhyfnhCUk>

- Infochimps (2013) CIOs & big data: what your it team wants you to know. Infochimps Whitepaper. <http://www.infochimps.com/resources/report-cios-big-data-what-your-it-team-wants-you-to-know-6/>. Zugegriffen: 21 Juni 2016
- Laksham A, Prashant M (2010) Cassandra – A decentralized structured storage system. *Acm Sigops Oper Syst Rev* 44:35–40
- Mülder W (1990) Implementierung. In: *Lexikon der Wirtschaftsinformatik*
- Schmarzo B (2013) *Big data: understanding how data powers big business*. John Wiley & Sons, Indianapolis
- Wixom BH, Watson H (2001) An empirical investigation of the factors affecting data warehousing success. *Mis Q* 25:17–41
- World Economic Forum (2011) Personal data: the emergence of a new asset class. <http://www.weforum.org/reports/personal-data-emergence-new-asset-class> Zugegriffen: 21 Juni 2016