

Edition TDWI

Datenqualität erfolgreich steuern

Praxislösungen für Business-Intelligence-Projekte

von

Detlef Apel, Wolfgang Behme, Rüdiger Eberlein, Christian Merighi

3., überarbeitete und erweiterte Auflage

dpunkt.verlag GmbH Heidelberg 2015

Verlag C.H. Beck im Internet:

www.beck.de

ISBN 978 3 86490 042 6

Vorwort zur 3. Auflage

Nach Schätzungen (vgl. [Crosby 1979, S. 15] und [Juran 1988, S. 1]) verursacht schlechte Datenqualität in Unternehmen Verluste in Höhe von bis zu 25 Prozent des operativen Gewinns. Aufgrund der zunehmenden Integration von IT in die Geschäftsprozesse der Unternehmen, der Anforderungen hinsichtlich Compliance sowie der Einbeziehung unternehmensexterner Daten (z.B. Big Data) nimmt die Bedeutung von Datenqualität nochmals erheblich zu. Die Hoffnung vieler Unternehmen auf Lösung der Datenqualitätsproblematik durch die Einführung von Standardsoftware für Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Supply Chain Management (SCM) u.a. hat sich nicht erfüllt und macht endlich Platz für wirksame Maßnahmen.

Im Mittelpunkt dieses Buches steht die Vermittlung langjähriger Erfahrungen aus BI-Projekten mit Datenqualitätsmanagement-Aktivitäten bei Unternehmen unterschiedlicher Branchen. Neben der anwender- und praxisorientierten Darstellung der verschiedenen Bereiche von Datenqualitätsmanagement (DQM) werden die Best Practices und Lessons Learned dargestellt, sodass der Leser eigene Projekte in diesem Umfeld erfolgreich durchführen kann.

Generell werden Daten über eine Benutzerschnittstelle erfasst oder durch Geschäftslogik von IT-Systemen erzeugt. Meistens fließen die Daten weiter in andere IT-Systeme und werden dabei transformiert. Ein Datenfluss kann viele Stationen haben. Das Data Warehouse ist häufig nur die »Endstation« solcher Datenflüsse. Werden fehlerhafte Daten nicht erkannt und behandelt, führen sie im Verlauf des Datenflusses zu Folgefehlern, die sich leicht zu größeren Problemen aufschaukeln können. Es liegt also auf der Hand, ein Datenqualitätsmanagement möglichst frühzeitig im Datenfluss anzusetzen. Nachhaltiges Datenqualitätsmanagement ist daher idealerweise eine unternehmensweite Aktivität, die ggf. von Vorhaben für Business Intelligence oder auch Customer Relationship Management angestoßen werden muss.

In den meisten Unternehmen kommen fehlerhafte Daten erst im Data Warehouse ans Licht. Das liegt daran, dass dort alle Daten in Gänze und verdichtet betrachtet werden, während beim Datenzugriff durch operative Systeme nur einige Felder in dem einen oder anderen Datensatz zutage treten. Schlechte Datenqualität lässt sich im Data Warehouse nicht verbergen. Allerdings ist es oft

genau diese schlechte Datenqualität, die die Akzeptanz der BI-Anwendung durch den Endanwender in den Fachbereichen verhindert und häufig direkt zum Misserfolg des mit dem Data Warehouse verbundenen Vorhabens führt. Wer will schon wichtige geschäftliche Entscheidungen auf fehlerhafte Daten stützen? Da lässt es sich noch besser aus dem Bauch heraus entscheiden.

Dieses Buch hat nicht den Anspruch eines unternehmensweiten Datenqualitätsmanagements, sondern fokussiert auf den Bereich Business Intelligence, wo der Schmerz mit fehlerhaften Daten am größten ist. Unter Business Intelligence (BI) wird ein integrierter, unternehmensspezifischer, IT-basierter Gesamtansatz zur Unterstützung betrieblicher Entscheidungen verstanden. »Business Intelligence ist der Prozess, der Daten in Informationen und weiter in Wissen umwandelt« (Definition von Howard Dresner (Gartner) 1989). Unternehmensentscheidungen und Prognosen stützen sich auf dieses Wissen und führen zu geschäftlichem Mehrwert. Business Intelligence kommt sowohl zur Unterstützung strategischer Entscheidungen als auch im operativen Bereich zum Einsatz.

Business Intelligence umfasst ein breites Spektrum an Anwendungen und Technologien und ist der Oberbegriff für Data Warehousing, Data Mining, Online Analytical Processing und Analytische Anwendungen. Im weiteren Sinne umfasst Business Intelligence auch die Erschließung unstrukturierter Daten mittels Content- und Dokumentenmanagement. Letztgenannte Bereiche sind jedoch nicht Gegenstand dieses Buches. Betrachtet wird lediglich die Business Intelligence im engeren Sinn, also auf strukturierte Daten bezogen.

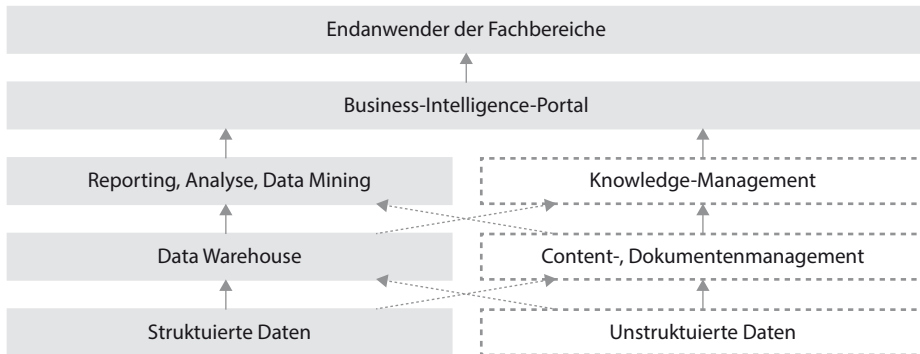


Abb. 1 Grobe Architektur für Business Intelligence

Das Data Warehouse ist eine konsolidierte Datenhaltung zur Unterstützung von Reporting und Analyse. »Ein Data Warehouse ist eine themenorientierte, integrierte, chronologisierte und persistente Sammlung von Daten, um das Management bei seinen Entscheidungsprozessen zu unterstützen« (vgl. [Inmon 1996]).

Das Buch gliedert sich in drei Teile. Im ersten Teil wird beschrieben, was Datenqualitätsmanagement ausmacht. Der zweite Teil befasst sich mit der Umsetzung und stellt insbesondere technische Hilfsmittel dar. Im dritten Teil wird

erklärt, wie man Verfahren, Methoden, Organisation und Werkzeuge des Datenqualitätsmanagements in der Praxis einsetzt.

Zu Beginn des ersten Teils (Kapitel 1) werden die wesentlichen Begriffe im Zusammenhang mit Datenqualitätsmanagement definiert. In Kapitel 2 wird erklärt, woran sich schlechte Datenqualität festmacht und wo die Ursachen dafür liegen. In Kapitel 3 wird dargelegt, warum es sich lohnt, ein Datenqualitätsmanagement aufzusetzen. In Kapitel 4 werden die organisatorischen Belange in Bezug auf die Datenqualität ausführlich geschildert. Die Architektur für BI-Anwendungen wird unter dem Blickwinkel der Datenqualität in Kapitel 5 betrachtet. In Kapitel 6 wird Big Data Analytics mit den Herausforderungen an das Datenqualitätsmanagement diskutiert. Hierbei wird auch auf den Bereich der unstrukturierten Daten eingegangen. Das Kapitel 7 beschreibt, wie sich Datenqualität messen lässt.

Im zweiten Teil des Buches werden wichtige Prinzipien der technischen Umsetzung des Datenqualitätsmanagements beschrieben. Dabei werden die Werkzeuge zur Unterstützung des Datenqualitätsmanagements betrachtet, angefangen beim Metadatenmanagement über Data Profiling, die Validierung, Bereinigung und Anreicherung von Daten bis hin zur fortlaufenden Überwachung der Datenqualität. Anschließend wird auf die Integration der Werkzeuge in die Anwendungslandschaft der jeweiligen IT-Umgebung eingegangen. Am Ende dieses Buchteils werden Kriterien zur Produktauswahl aufgeführt.

Der dritte und letzte Teil des Buches bildet Datenqualitätsmanagement auf das Vorgehen in BI-Projekten ab. Dabei werden die einzelnen Phasen eines BI-Projekts von der Vorstudie über Spezifikation, Design und Umsetzung bis zum Betrieb im Unternehmen betrachtet. Für jede Projektphase werden die jeweils einzusetzenden Elemente des Datenqualitätsmanagements benannt, die im zweiten Teil des Buches beschrieben wurden. Somit bietet der dritte Buchteil für Projektverantwortliche eine unverzichtbare Hilfestellung zur erfolgreichen Durchführung von Projekten.

Was hat sich in der 3. Auflage geändert? Im 1. Teil des Buches wurde das Thema Big Data neu aufgenommen, da es für die Welt der Business Intelligence eine neue Evolutionsstufe darstellt und somit Auswirkungen auf das Datenqualitätsmanagement hat. Weiterhin wurden in allen Kapiteln Aktualisierungen vorgenommen.

Detlef Apel, Wolfgang Behme, Rüdiger Eberlein, Christian Merighi
Troisdorf, Hannover, München, Wien, im Dezember 2014