



Daten- management

Ein Thema für alle!

Egal in welcher Branche - Daten sind und werden zu einem immer wichtigeren Rohstoff. Der Siegeszug datengetriebener Geschäftsmodelle macht ihre zentrale Bedeutung zunehmend sichtbar. Rapide wachsende Datenmengen, die Forderung nach Realtime-Verfügbarkeit sowie die Notwendigkeit zur bewussten organisatorischen Verankerung stellen große Herausforderungen an das strategische Datenmanagement. Neue Technologien eröffnen Wege und Möglichkeiten, die es zu nutzen gilt.

In Daten und deren strategischem Management liegt wertvolles Potenzial, darin sind sich alle einig. Wer davon nachhaltig profitieren will, muss verinnerlichen, dass es sich bei Daten um ein zentrales, äußerst wertvolles Gut mit abteilungsübergreifender Relevanz handelt, und dies schnellstens in seine Unternehmenskultur aufnehmen. Denn eine auf Daten basierende Information, die an der einen Stelle vielleicht wenig Bedeutung hat, kann an einer anderen ein äußerst wirksamer Hebel sein.

Vor diesem Hintergrund darf Datenmanagement auch nicht mehr allein als IT-Thema behandelt werden, sondern sollte auf der Agenda aller Abteilungen stehen. Es beinhaltet gleichermaßen klassische und moderne Aufgaben und reicht von der Sammlung und Automatisierung der Datenarbeit über die Optimierung von Geschäftsabläufen bis zur Aufbereitung für die unternehmensweite Informationsversorgung und für neue Geschäftsmodelle.

Es geht um Master Data Management und Data Quality oder Themen wie Data Discovery, Data Lineage, Datenvisualisierung und -distribution. Hinzu kommen moderne Konzepte und Technologien wie Cloud, Mobility, Big Data oder Predictive und Prescriptive Analytics, die weiterführende Dimensionen öffnen. An den Prozessen beteiligt sind Management, IT und Fachabteilungen, aber auch Anwender, Kunden und Multiplikatoren – im Grunde also „alle und jeder“.

Strategisches Datenmanagement: eine weitreichende Disziplin, die das gesamte Unternehmen betrifft

Die Bildung organisatorischer Einheiten im Unternehmen, z.B. einer Data Factory oder eines Business Intelligence Competence Center (BICC), ist die logische Konsequenz, und eine Data-Governance-Richtlinie, die Datenhoheit, Zuständigkeiten und Prozesse definiert, ist unabdingbar. Denn erst eine nachhaltige Organisation von Daten im Unternehmen ermöglicht strategisches Arbeiten. In diesem Sinne thematisiert das Konzept der Data Governance mit seinem auf wertvolle und qualitativ hochwertige Daten zielenden ganzheitlichen Ansatz gleichermaßen Prozesse, Technologien und Menschen, die als Data Executive, Data Owner, Data Steward oder Data Specialist klar definierte Verantwortlichkeiten und Rollen übernehmen.

Daten dürfen nicht mehr nur alphanumerisch betrachtet werden

Unter "Daten" versteht man zunächst vor allem die klassischen Stammdaten wie Kunden, Artikel und Gesellschaften sowie Bewegungsdaten wie Umsatz und Angebote aus dem Vertrieb, Responsezahlen aus Marketing-Kampagnen oder Finanzdaten aus der Buchhaltung. Zu diesen primär strukturierten, alphanumerischen Daten, die in sogenannten relationalen, transaktionalen oder auch multidimensionalen Systemen gehalten werden, kommen zunehmend unstrukturierte Daten wie Texte, E-Mails, Nachrichten in sozialen Netzwerken, Sensordaten, Audiodateien oder Videos. Nur etwa drei Prozent dieser Daten lassen sich über ein Schlagwort suchen, die anderen sind komplett unstrukturiert. Die Menge an Daten im Internet verdoppelt sich dabei derzeit alle zwei Jahre.

Für das Datenmanagement bedeutet dies, dass die Erfassung, Organisation und Ordnung verschiedener Datenarten in die strategischen Überlegungen mit einbezogen werden müssen. Dabei geht es nicht mehr nur um unternehmensinterne Daten, sondern es muss auch ein performanter und zuverlässiger Zugang zu öffentlichen Quellen geschaffen werden, beispielsweise über standardisierte Web-Service-Schnittstellen und Datenmarktplätze. Ebenso gilt es darauf zu achten, dass wirklich alle relevanten Daten berücksichtigt werden, denn Wissenseinheiten jedweder Art werden erst durch ihre Erfassung überhaupt systemrelevant. Was dagegen nicht erfasst wurde, ist in der Welt der Daten faktisch nicht passiert. Nur verfügbare Daten können modelliert, aggregiert und analysiefähig gemacht werden und ihren Informationswert nachhaltig entfalten – ein Aspekt, den Unternehmen oft vernachlässigen. Damit werden wertvolle Potenziale verschenkt oder sogar aufgrund unvollständiger Informationen fatale Schlüsse gezogen. →

von Dr. Ilias Michalarias und Markus Sümmchen

Eine valide Datenbasis ist die kompromisslose Grundlage für unternehmerische Entscheidungen

Um das Thema Datenmanagement in den Griff zu bekommen, haben sich viele Unternehmen in den letzten Jahren mit dem Aufbau eines Data Warehouse beschäftigt und damit eine zentral verwaltete, aus den Daten der unterschiedlichsten Systeme und Medien zusammengeführte einheitliche und verfügbare Datenbasis geschaffen. Für die Datenbewirtschaftung dieser Plattform stehen am Markt ausgereifte ETL (Extract-Transform-Load)-Werkzeuge bereit, die einen hohen Automatisierungsgrad aufweisen, Data Cleansing-Routinen für die Datenbereinigung beim Import enthalten und Schnittstellen zu verschiedensten Vorkomplexen wie SAP, Microsoft Dynamics AX und Navision oder vielen anderen Standardsystemen mitbringen. Hierunter fallen sowohl Informatica als Marktführer für Datenintegration, als auch Microsoft mit den SQL Server Integration Services als Datenbankhersteller sowie spezialisierte Tools wie beispielsweise der IDL.IMPORTER.

Eine der Hauptdisziplinen zur Gewährleistung einer konsistenten und zuverlässigen Datenbasis im Data Warehouse ist die Stammdatenverwaltung, sprich das Master Data Management (MDM). Ziel des MDM ist, die Stammdaten zu konsolidieren, zu harmonisieren und Informationen über deren Beziehungen untereinander zu dokumentieren. Das bedeutet, dass beispielsweise ein Kunde, der sowohl im CRM- als auch im ERP-System geführt wird, eine eindeutige Kundennummer im Data Warehouse hat. Dadurch ist er in beiden Systemen eindeutig identifizierbar, und auch alle weiterführenden Informationen können klar zugeordnet werden.

Data Quality wirkt nach Innen und nach Außen

Nur auf der Grundlage einer eindeutigen und validen Datenbasis lassen sich unternehmerische Maßnahmen vernünftig aufsetzen. Das Thema Datenqualität ist daher kompromisslos zu verfolgen – und zwar über den gesamten Daten-Lebenszyklus hinweg, von der Erhebung über die Aufbereitung bis zur Bereitstellung. In diesem Sinne sind sämtliche Prozesse und Mechanismen im Sinne von Data Quality Management und Data Cleansing umzusetzen. Sie helfen beispielsweise dabei, Postleitzahlen automatisiert zu überprüfen und zu korrigieren, Dubletten abzugleichen und zu eliminieren oder fehlende Daten aus anderen Daten

abzuleiten (z.B. die Anrede aus dem Vornamen). Zudem haben viele Unternehmensbereiche spezielle Qualitätsanforderungen, die berücksichtigt werden müssen. Dazu gehören u.a. anonymisierbare Daten im Personalbereich, die Aggregations- und Analysefähigkeit für das Controlling oder die Testierbarkeit von Finanzdaten. In Summe gilt es, Qualitätsstandards zu definieren und diese sowohl bei wiederkehrenden Prozessen als auch bei nicht alltäglichen Aufgaben (z.B. die Eingliederung einer neuen Konzerngesellschaft und die damit verbundenen Erweiterungen in den Datenflüssen) konsequent anzuwenden.

Das Thema Data Quality ist aber nicht nur aufgrund interner Anforderungen wichtig, sondern wird hoch brisant, wenn es um multinationale Konzerne, börsennotierte Unternehmen oder Finanz- und Kreditinstitutionen und deren Pflichtveröffentlichungen geht (aktuelles Stichwort: FINREP). Hier ist an eine Vielzahl von Melde- und Ausweispflichten zu denken, deren Stimmgenauigkeit weitreichende Konsequenzen hat. Data-Quality-Prozesse fungieren zunächst als Automatisierungshilfe und als Schutzmechanismen gegen eigene unbeabsichtigte Fehler. Sie können aber auch Fehler vermeiden helfen, die bewusst von außen eingespielt werden – man denke hier an das Thema Fraud Detection.

Von Business Intelligence bis Big Data:
Finden, wovon ich noch gar nichts wusste und Unternehmen damit proaktiv steuern

Will man aus den Daten Informationen generieren, ist es mit der Ablage und Organisation operativer Rohdaten nicht getan, denn diese sind per se nur selten aussagekräftig. Eine Disziplin, die sich mit der Transformation von Daten in Information sowie deren Visualisierung und Bereitstellung beschäftigt, ist die Business Intelligence. Angefangen von der Datenhaltung in multidimensionaler und analytischer Form in OLAP-, MOLAP- oder InMemory-Datenbanken über die Ad-hoc-Analyse, das Data Mining, die Nutzung statistischer Algorithmen im Sinne von Predictive und Prescriptive Analytics oder dem Aufsetzen moderner Data-Discovery-Strategien für die datengetriebene Zusammenarbeit aller Unternehmensbereiche bis hin zur Bereitstellung der Ergebnisse in Form von Berichten, Reports und Dashboards oder zur Unterstützung der Unternehmensplanung: Bei Business Intelligence geht es immer darum, Daten zu aggregieren, um werthaltige Informationen zu generieren. Ihr Ziel ist es, durch das Analysieren und Durchforschen von Datenmaterial Auffälligkeiten zu identifizieren, Zusammenhänge aufzudecken, Möglichkeiten durchzuspielen, Maßnahmen abzuleiten und schließlich ein Unternehmen auf Basis dieser Informationen und Erkenntnisse proaktiv zu steuern. Die Methoden und Technologien haben sich in den letzten Jahren kontinuierlich weiterentwickelt. Die Tools und Plattformen am Markt, mit denen die entsprechenden Lösungen – idealerweise aufsetzend auf Data Warehouses – aufgebaut werden, sind zahlreich. Zu nennen sind hier die großen Marktplayer wie Microsoft, IBM, Oracle, SAP ebenso wie IDL als einer der führenden Spezialanbieter.

Eine der neuen Methoden bei der Datenanalyse ist das Thema Data Lineage. Hier geht es darum, den Datenfluss sowohl bei aktuellen als auch bei archivierten Daten bis hin zu komplexesten Daten-Transformationen jederzeit zu tracken, um damit beispielsweise nachvollziehen zu können, wie ein angegebener Umsatz zustande gekommen ist oder wie sich der Ergebnisbeitrag pro Landesgesellschaft bemisst. Eine Historisierung im Sinne von temporaler Datenhaltung sorgt dafür, dass stets lückenlos rekonstruiert werden kann, welche Werte zu welchem Zeitpunkt gültig waren. Das wiederum ist wesentliche Voraussetzung für Vergleiche entlang der Zeitachse, Abweichungsanalysen, Simulationen und Forecasts. Dass Speicherplatz (Storage) heute so günstig ist wie nie und die Archivierung auch umfassender Datenmengen damit keine großen Kosten mehr verursacht, kommt dieser Strategie sehr entgegen.

Self Service BI birgt Chancen und Risiken

Ebenso muss hier das Stichwort Big Data genannt werden. Es steht für neue Technologien und Vorgehensmodelle, die das Sammeln und Auswerten immenser strukturierter sowie unstrukturierter Datenvolumina ermöglichen, um diese in Realtime oder Near-time auszuwerten, in entscheidungsrelevante Informationen umzuwandeln und neue Erkenntnisse zu gewinnen – ganz im Sinne von: Ich suche nach irgendetwas, das mir bislang noch nicht bekannt war.

Im Idealfall sind alle Daten, deren Analysemöglichkeiten und die daraus abgeleiteten Informationen allen nur erdenklichen Nutzerkreisen jederzeit und überall in der jeweils gewünschten Form und entsprechend ihrer jeweiligen Rechte und Bedürfnisse zugänglich, und zwar auf dem aktuell gewünschten Endgerät – stationär wie mobil. Dieses Ziel ist durch ausschließliche Zusammenarbeit mit der IT-Abteilung nicht zu erreichen, denn diese verfügt meist nicht über die notwendigen Kapazitäten und Ressourcen, um die beständig wachsenden Datenmengen und fachlichen Anforderungen zu bewältigen. Vor diesem Hintergrund sind Self Service BI und ihre Tools, wie Microsoft Power BI oder SAP Lumira, mit denen die Anwender autark Analysen und Berichte aufbauen und nutzen können, in den letzten Jahren zu einem eigenen Business-Case geworden und haben Einzug in vielen Fachabteilungen gehalten.

Doch bei aller Freude über diesen Ansatz der Eigenständigkeit muss auch klar sein, dass dieser eine Herausforderung für das Thema Datenqualität darstellt und im Hinblick auf das zentrale und strategische Datenmanagement kontraproduktiv wirken kann. Denn die typischen Insel-Lösungen (meist auf Basis von Excel), die durch den Einsatz durchgängiger BI-Applikationen zunehmend verdrängt wurden, werden damit nun aufs Neue aufgebaut – und zwar mit sämtlichen bekannten Nebenwirkungen. →

Cloud und SaaS bringen die Antwort auf wachsende Datenvolumina und Realtime-Anforderungen

Aber wie soll man sonst mit den ständig wachsenden Datenvolumina sowie den neuen Ansprüchen sowohl in qualitativer als auch in zeitlicher Hinsicht umgehen? Und wie trägt man der Skalierbarkeit Rechnung und schafft die notwendigen Infrastrukturen? Über solche Fragestellungen gelangt man sehr schnell zum Thema Cloud-Computing und den dahinterstehenden tragfähigen Strategien, die es erlauben, grundlegende IT-Ressourcen wie Rechenleistung und Datenbanken, Speicher- und Netzwerkkapazitäten nebst entsprechender Applikationen (Software as a Service / SaaS) bedarfsgerecht über das Internet zu beziehen.

Damit liegt auch das „Datenmanagement as a Service“ nahe, zumal sich über service-orientierte Ansätze nicht nur Dienste wie Master Data Management und Data Quality, sondern beispielsweise auch ETL-Prozesse, Data Mining oder Self-Service-Analytics auf Basis von In-Memory-Technologien effizient und sicher abbilden lassen. Und sind die operativen Daten erst einmal in der Cloud, also im Internet, verfügbar, können sie auch auf mobilen Endgeräten wie Tablet und Smartphone bereitgestellt werden, was wiederum Basis für neue Geschäftsmodelle ist.

Datenmanagement ist zentrales Thema der digitalen Zukunft

Der strategische Umgang mit den Daten ist in der digitalen Ära also längst zur überlebenswichtigen Aufgabe geworden, und das systematische Datenmanagement nimmt einen festen Platz im Unternehmensmanagement ein. Das Datenmanagement wird zukünftig mehr und mehr Einfluss auf den unternehmerischen Erfolg haben, denn die Gefahr, dass die digitalen Unternehmen die nicht digitalen verdrängen, steht schon heute im Raum. ●

Dr. Ilias Michalarias

ist seit neun Jahren in den Bereichen Business Intelligence und Data Warehousing tätig. 2007 wurde ihm der Dokortitel durch den Fachbereich Informatik von der Freien Universität Berlin für seine Forschungsarbeit zum Thema „Mobiles Reporting“ verliehen. Seit Mitte 2012 ist er Business Analyst bei der white duck GmbH, einem Spezialisten für cloud-basierte Software-Entwicklung und -Beratung. Davor zeichnete er als Produkt Manager des BI-Anbieters Cubeware für das Produktportfolio verantwortlich.
www.whiteduck.de



Markus Sümichen

ist geschäftsführender Gesellschafter der white duck GmbH, einem Spezialisten für Software-Entwicklung von Anwendungen und Beratung rund um Cloud-Computing. Er ist seit über 15 Jahren im Software-Entwicklungsumfeld tätig und war bis Mitte 2012 als Mitglied der Geschäftsleitung bei dem BI-Anbieter Cubeware für das Produktportfolio verantwortlich. Die Schwerpunkte von Markus Sümichen liegen in der Konzeption und dem Design von modernen Cloud-Anwendungen sowie der Beratung bei der Überführung bestehender Anwendungen in die Cloud.
www.whiteduck.de

