



Qualität geht vor Quantität

Mit modernen Datenintegrationslösungen lassen sich nahezu alle relevanten Unternehmensdaten in einem Data Warehouse zusammenfassen. Voraussetzung dafür ist jedoch eine Datenintegration auf Basis gesunder Daten. Hier helfen Lösungen wie beispielsweise „Informatica Data Quality“ von Informatica, mit denen sich die Datenqualität bereits innerhalb der Prozesse verbessern lässt.

(erschieden in Sapport 07-08/2007, S. 25-28 Autor: Thomas Schumacher)

Ein ganz normaler Tag in Deutschland: 78 Gewerbebetriebe werden neu angemeldet und 75 wieder abgemeldet. Sechs Unternehmen sind aus dem Ausland zugezogen, sieben ins Ausland abgewandert. Neun Unternehmen sind übernommen, fünf sind insolvent geworden. 90 Personen haben geheiratet, 48 sind geschieden worden, acht haben Privatinsolvenz angemeldet und 4500 ihre Adresse geändert. Das alles geschieht nach Angaben des statistischen Bundesamtes (Quellen: destatis.de/ regionalstatistik.de) täglich binnen 45 Minuten. Glaubt man den Analysten von PriceWaterhouseCoopers, so können allein aufgrund von Umzug, Scheidung, Heirat oder Tod rund zwei Prozent der Einträge in einem CRM-System in nur einem Monat obsolet werden. Hinzu kommen fehlerhafte Dateneinträge und Sammelp Prozesse, die dazu führen, dass unrichtige Informationen und Dopplungen Datenbanken sukzessive verwässern. Infolgedessen geben allein Marketingabteilungen jährlich Millionen aus, um teure Broschüren doppelt oder an fehlerhafte Adressen zuzustel-

len. Schlimmer noch: Schätzungen zufolge treffen drei Viertel der Anbieter von Finanzdienstleistungen geschäftliche Entscheidungen auf Basis nicht optimaler Daten.

Qualität wurde bisher sträflich vernachlässigt

Grund dafür ist, dass in den vergangenen Jahren das Thema der Qualität von Daten sträflich vernachlässigt worden ist. Die IT-Abteilungen der Firmen haben enorme Summen in Systeme investiert, mit denen sich zahlreiche interne Abläufe automatisieren lassen. In der heutigen Unternehmenspraxis führen die Daten daher eine Art Eigenleben: sie stoßen Prozesse an, werden mit anderen Daten verknüpft und die daraus resultierenden Ergebnisse dienen letztlich dem Management als Grundlage von Entscheidungen. Dazu werden Daten aus CRM-, ERP oder Business Intelligence-Systemen in großen Data Warehouse-Installationen gespeichert und von dort aus den Analyseprogrammen zur Verfügung gestellt. Mit modernen Datenintegrationslösungen lassen sich nahezu alle relevanten Unternehmensdaten in ei-

nem Data Warehouse zusammenfassen - ob strukturierte Daten aus Datenbanken und ERP-Anwendungen oder unstrukturierte, etwa aus E-Mails, Präsentationen, Word-Dokumenten oder PDFs. Die fortschreitende Ausweitung der Datenbasis verspricht naturgemäß exaktere Informationen über Vorgänge im Unternehmen, bringt aber auch Probleme mit sich. Neue Compliance-Regeln, inter-nationaler Wettbewerb sowie Firmenzusammenschlüsse und -übernahmen zwingen Unternehmen dazu, sich genauer mit den Daten und ihrer Qualität auseinanderzusetzen. Wer will schon eine scheinbar profitable Firma übernehmen, nur um nach einem Quartal festzustellen, dass der neue Partner in Wirklichkeit pleite war? Die Frage ist bloß: wie verlässlich, komplett oder korrekt sind diese Informationen? Den Verantwortlichen ist nach und nach klar geworden, dass in ihren Datenbanken bestenfalls eine Momentaufnahme festgehalten wird, die durch eigene Bemühungen ständig auf dem aktuellen Stand gehalten werden muss.

Datenintegration auf Basis gesunder Daten

Innovative Softwareunternehmen haben diese Lücke erkannt und Lösungen entwickelt, mit denen sich die Datenqualität bereits innerhalb der Prozesse verbessern lässt. Dazu zählt das amerikanische Softwarehaus Informatica, das mit der Übernahme der irischen Similarity Systems Anfang des Jahres seiner Datenintegrationsplattform PowerCenter führende Data Quality-Technologie hinzufügen konnte. Heute sorgen innerhalb von PowerCenter die Module IDE (Informatica Data Explorer, vormals Similarity Systems Evoke Axio) und IDQ (Informatica Data Quality, vormals Similarity Athanor) für eine Steigerung der Informationsqualität. Während IDE für schnelle Analysen und die Profilierung von Daten aus heterogenen Quellsystemen eingesetzt werden kann, lassen sich mit IDQ Regeln für die Datenvalidierung entwickeln und Datenbereinigungen und -konsolidierungen durchführen. Im Zusammenspiel mit der bewährten Datenintegrationslösung entsteht so eine Datendrehzscheibe, die auch die Qualitätsprobleme adressiert. Dabei gibt es eine Vielzahl von Definitionsansätzen, was Datenqualität letztlich ausmacht, im Kern geht es dabei darum, dass die Daten den Anforderungen der Benutzer entsprechen müssen. Informatica hat für die Definition und Messung der Datenqualität sechs Dimensionen entwickelt:

1. Vollständigkeit: Welche Daten fehlen oder sind unbrauchbar?
2. Konformität: Welche Daten werden in einem Nicht-Standard-Format gespeichert?
3. Konsistenz: Welche Datenwerte geben Aufschluss über Konflikte?
4. Genauigkeit: Welche Daten sind unkorrekt oder veraltet?
5. Dopplungen: Welche Einträge bestehen doppelt?
6. Integrität: Welchen Daten fehlen wichtige Beziehungsverknüpfungen?

Beispielsweise kann ein Marketingmanager eine Kampagne auch dann noch ausführen, wenn 20 Prozent der Daten fehlerhaft oder unvollständig sind. Die Kampagne wird zu teuer, ineffizient und Verschwendung von Ressourcen sein, aber es ist unwahrscheinlich, dass sie gänzlich scheitert. Bei einer Bank, die ihre Risiken für die Einhaltung der Basel II-Regeln kalkuliert, oder einer Versicherung, die Policen für Öltanker unterschreibt, sieht das anders aus.

Prozesse und Technologie gehen Hand in Hand

Bei Projekten zur Verbesserung der Datenqualität ist es von herausragender Bedeutung, nicht nur die gesammelten Daten zu betrachten, sondern bereits die Prozesse genau zu untersuchen, die zu schlechter Datenqualität führen. Häufig sind die Unzulänglichkeiten in den Systemen begründet in schlecht dokumentierten Änderungen, mangelhafter Ausbil-

dung der Anwender oder fehlerhaften Benutzerhandbüchern. Hinzu kommt, dass Daten sehr oft über den ursprünglich angedachten Einsatzzweck hinaus genutzt werden. Besonders der letzte Punkt ist heute in Geschäftsumgebungen ein großes Problem, dann dabei fließen Daten in wichtige Entscheidungsprozesse, die dafür überhaupt nicht vorgesehen waren. Gleichzeitig können die fragmentierten und verteilten IT-Systeme Doppelpfeile, mangelnde Konformität zwischen Systemen und andere Diskrepanzen zur Folge haben. Neben den Daten müssen daher vor allem die internen Abläufe angepasst werden. Derzeit haben die meisten Unternehmen das Problemfeld Datenqualität mit taktischen Lösungen angegangen, um die Qualität innerhalb einer Abteilung oder eines Prozesses zu verbessern. So lassen sich zwar konkrete Fehler in Teilen der Organisation kurzfristig ausmerzen. Langfristig scheitern solche Initiativen aber daran, in großem Maßstab eine Verbesserung der Datenqualität herbeizuführen. Der Denkfehler dabei ist, dass Datenqualität nicht einmalig herbeigeführt und dann vergessen werden kann. Langsam setzt sich die Erkenntnis durch, dass Daten im Grunde genommen verderbliche Güter sind und auch als solche behandelt werden müssen. Unternehmen müssen darauf reagieren, indem sie bereits genutzte Daten bereinigen, Daten schlechter Qualität davon abhalten, in die Systeme zu gelangen und sich in Erinnerung rufen, dass der Aktuali-

tätsgrad einer Information ab dem Zeitpunkt der Erhebung kontinuierlich nachlässt, so dass Projekte in diesem Bereich eine fortwährende Aufgabe sind.

Eine nicht nur technische Herausforderung

Datenqualität ist eher ein strategisches Thema als ein technisches und sollte aus diesem Grund auch organisatorisch bei der Unternehmensleitung und den Datenkonsumenten aufgehängt sein, die dann in Zusammenarbeit mit der IT nach gemeinsamen Lösungsansätzen suchen. Zur Lösung des Problems muss ein unternehmensweiter Ansatz herangezogen werden, der eine Kombination aus Technologie und Organisation, Struktur und Prozesswechsel enthält. Dabei sollte dieser Ansatz aus Sicht des Geschäfts und nicht aus dem Blickwinkel der IT verfolgt werden. Wichtig ist auch, dass dieses Thema in den Unternehmen so schnell wie möglich und mit einem umfassenden Lösungsansatz angegangen wird. Die Tatsache, dass es sich um eine Herausforderung mit offenem Ende handelt, sollte nicht als Alibi benutzt werden. Mit Werkzeugen wie Informatica Data Quality und dem Informatica Data Explorer lassen sich nachhaltige Verbesserungen erzielen. Unternehmen profitieren von geringeren Streuverlusten und fundierteren und damit vermutlich auch besseren Entscheidungen. Die Investitionen rechnen sich schnell, denn eine Steigerung der Datenqualität bewirkt bes-

sere Unternehmensentscheidungen und optimalen Mitteleinsatz. Ergebnisse von denen letztlich alle Bereiche eines Unternehmens profitieren. Eigentlich kann es sich kein Unternehmen leisten, auf dieses Potenzial zu verzichten.