Mai 2017 / by MIOsoft

Verlässlich – oder nicht? Datenqualitätserfassung in heterogenen IT-Systemlandschaften

Aufgrund der unterschiedlichen Anforderungen im Bereich Datenverarbeitung, mit denen Unternehmen sich im digitalen Zeitalter konfrontiert sehen, verwenden die meisten Betriebe eine Vielzahl von Softwaresystemen. Da der rasante Fortschritt in der Informationstechnologie zudem laufend für Neuerungen sorgt, sind die IT-Systemlandschaften in etablierten Unternehmen heute stark heterogen. Hieraus ergeben sich bei strukturellen Veränderungen – etwa im Rahmen von Fusionen, Zukäufen oder Auslagerungen – häufig gravierende Probleme: Datensätze werden unabsichtlich dupliziert oder falsch zugeordnet. Dadurch sinkt die Datenqualität – unter Umständen so stark, dass die Daten nicht mehr als Informationsgrundlage für strategische Entscheidungen taugen. Folge: Die Handlungsfähigkeit des jeweiligen Unternehmens ist kaum oder gar nicht mehr gegeben.

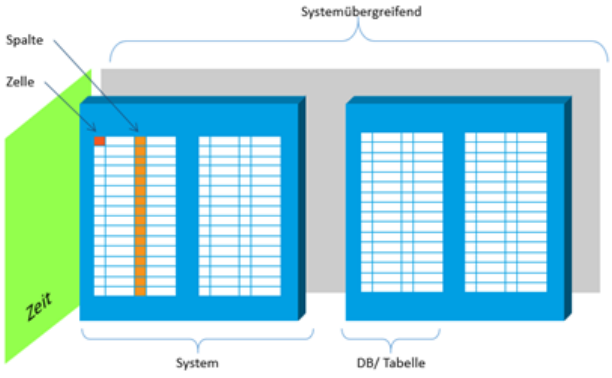
Dabei ahnt man in den IT-Abteilungen durchaus, wo der Hase im Pfeffer liegt. Mangels geeigneter Messmethoden kann man die Datenqualität aber nicht erfassen, und schon gar nicht ist man in der Lage, Datenfehler großräumig auszumerzen.

Der Weg zu einer guten Datenqualität führt über eine Datenqualitätsplattform

Um Datenqualitätsprobleme effizient und souverän lösen zu können, muss man zunächst *Data Quality Indicators* (DQIs) definieren – diese erlauben es, die Verlässlichkeit von Daten darzustellen. Außerdem benötigt man eine Datenqualitätsplattform, auf der man fest definierte Regeln aufstellen und umsetzen kann und auf der sich ferner Messungen über Systemgrenzen hinweg und unter Berücksichtigung von Prozessabhängigkeiten vornehmen lassen. Dass die Ergebnisse skalierbar sein müssen, versteht sich. Und auch Zeit ist ein Faktor: Alles muss extrem schnell funktionieren. Je nach Unternehmensstruktur und Geschäftsausrichtung können Datenfehler schließlich Umsatzverluste in Millionenhöhe verursachen – da zählt jede Stunde.

Messung Schritt für Schritt

Die Grundlage jeder Datenqualitätsmessung sind technische Basisprüfungen innerhalb der einzelnen Systeme. Bei diesen Prüfungen werden die Standardregeln zur Eindeutigkeit, zur Vollständigkeit, zur Gültigkeit und zur korrekten Syntax auf die jeweiligen Systeminhalte angewendet. Der nächste Schritt ist eine systemübergreifende Konsistenzprüfung – hierbei wird die Konsistenz zwischen redundanten Daten in verschiedenen Systemen überprüft sowie die referenzielle Integrität (gibt es für gleiche Kategorien auch gleiche Referenzen?). Dazu muss eine persistente transitive Hülle über das Datenkontingent „gestülpt" werden, die es erlaubt, die Einhaltung fachlicher und prozessabhängiger Datenqualitätsregeln zu überprüfen.



Die Menge macht's

Nur durch eine Kombination beider Schritte ist es möglich, die Datenqualität zu messen und sie in der Folge auch zu verbessern. Konkret heißt das: Nur dann, wenn technische und fachliche Regeln ganzheitlich, also systemübergreifend, auf die zur Verfügung stehenden Daten angewendet werden, können DQIs und *Key Performance Indicators* (KPIs) miteinander verknüpft werden, sodass sich erkennen lässt, wie es um die Verlässlichkeit der Daten bestellt ist. Datenqualitätsmessungen für Millionen oder gar Milliarden von Datensätzen benötigen allerdings viel Zeit – oder eben sehr große Hardware-Ressourcen. Hoch skalierbare Plattformen, wie etwa die Plattform MIOvantage des Software-Anbieters MIOsoft, gestatten indes selbst auf enorm große Datenmengen einen performanten Zugriff. Dabei können die Regeln jederzeit angepasst werden bzw. es können je nach Bedarf auch neue Regeln aufgestellt und angewendet werden. Des Weiteren lässt sich das Datenvolumen nahezu beliebig vergrößern, und es können auch neue Datenquellen hinzugefügt werden. In einem von MIOsoft betreuten Projekte wurde eine Datenqualitätsmessung über sieben Kernsysteme mit insgesamt rund 11 Milliarden Records durchgeführt – diese Zahlen sprechen für sich.