



b.telligent

München • Düsseldorf • Frankfurt • Hamburg • Stuttgart • Zürich

Datenqualität erhöhen

mit automatisierten Prozessen

Verbesserung der Qualität durch kontinuierliche Überwachung

Hohe Datenqualität ist einer der am meisten unterschätzten Faktoren für den Erfolg von Informationssystemen. b.telligent beugt Datenqualitätsproblemen erfolgreich durch Präventivmaßnahmen auf Prozessebene vor: Mit Hilfe von Datenqualitätsmonitoren werden Probleme frühzeitig und automatisiert im Data Warehouse erkannt und korrigiert. So erfährt die Datenqualität eine kontinuierliche Überwachung und Verbesserung. Abweichungen oder Fehler werden systemübergreifend rasch erfasst, gezielt lokalisiert und damit kostengünstiger behoben.

Das DWH spielt in der IT-Infrastruktur eines Unternehmens eine entscheidende Rolle. Ein Großteil der Fachbereiche verwendet es für Auswertungen und Unternehmenssteuerung. So wird ein entsprechend hoher Wert auf Qualität der Daten, Effizienz und Automatisierung gelegt.

Mangelnde Datenqualität (DQ) im DWH führt im besten Fall zu suboptimalen Entscheidungen. Im schlechtesten Fall werden Fehlentscheidungen getroffen, die dem Erreichen von Bereichs- und Unternehmenszielen entgegenwirken. Über das Verfehlen von quantitativen Zielen hinaus können Fehler in Managementreports zu unternehmensinternen Unstimmigkeiten führen. Fehler in Quartalsberichten können sich bis hin zu externem Vertrauensverlust entwickeln.

Allgemein betrachtet mündet mangelhafte Datenqualität direkt in Ertragsverluste, Verluste durch Fehlentscheidungen oder auch sonstige Risikokosten. Doch das ist nur die Spitze des Eisbergs: Indirekt entstandene Kosten sind zum Beispiel Kosten durch entgangene Geschäfte, erhöhte Personalkosten oder Kundenabwanderung, wie Abbildung 1 verdeutlicht.

Besonders in komplexen Systemen wie einem DWH setzen sich Datenqualitätsmängel schnell fort. Deren Korrektur in

nachgelagerten Teilsystemen ist daher umso zeitaufwändiger und teurer. Eine zielgerichtete Erkennung von Datenqualitätsproblemen schon beim Laden des DWH ist also unerlässlich.

Aus diesem Grund hat b.telligent die Datenqualitätsmonitoringplattform (DQMP) entwickelt.

DQMP besteht aus verschiedenen Prüfverfahren (DQ-Module), die in Form von Prüfroutinen (DQ-Monitore) auf konkret zu prüfende Objekte angesetzt werden können. Kritische Daten werden durch dieses regel- und erfahrungsbasierte Werkzeug analysiert und geprüft. Wie beim automatisierten Unit Testing üblich werden kleine Prüfbausteine regelmäßig ausgeführt und ausgewertet.



Abbildung 1: Kosten von DQ-Problemen

Ein DQ-Monitor beschreibt eine Qualitätsprüfung und besteht aus der Angabe eines Moduls und einigen weiteren Parametern, wie zum Beispiel:

- Prüfverfahren/Modul
- Daten über zu prüfende Objekte und Entwickler (Namen, Schemata etc.)
- konkrete Prüfabfrage(n)
- Angaben zur Ausführungszeit und Fehlertoleranz

Ein DQ-Monitor vergleicht beispielsweise täglich das Volu-

men durchgeführter Transaktionen zwischen dem Quellsystem und dem auswertbaren Datenbestand in den Data Marts. Die gefundenen Abweichungen können als Problemindikatoren behandelt werden. Statistische Ausreißer oder eine direkte Überschreitung von Grenzwerten deuten auf eine fehlerhafte oder unvollständige Extraktion beziehungsweise außergewöhnliche Datenlieferung durch das Quellsystem hin.

Die einzelnen DQ-Monitore werden entweder durch das DQM-Tool selbst ausgeführt (zeitbasiert) oder von einem beliebigen anderen Scheduler per Webservice getriggert (eventbasiert).

Bei Überschreitung festgelegter Schwellwerte kann ein vordefinierter Empfängerkreis automatisch per E-Mail oder über ein Ticketingsystem informiert werden, so dass kritische Situationen sofort behandelt werden können. Die Monitordefinition und -wartung sowie die Aufbereitung und Darstellung der Laufergebnisse werden über ein webbasiertes grafisches Frontend realisiert.

Die Module (siehe Abbildung 2) definieren generische Prüfverfahren, also die Art und Weise, in der eine Prüfung abläuft. Dies kann beispielsweise der direkte Vergleich von Aggregationen oder die Whitelist-Prüfung sein. Hinter einem Modul verbirgt sich ein Java-Programm, das immer eine bestimmte Datenkonstellation analysieren kann.

Einbettung in den Entwicklungsprozess

Die DQMP ist nahtlos in den Entwicklungsprozess (siehe Abbildung 3) der DWH-Anwendungsentwicklung eingegliedert. In den meisten Fällen ist bereits in der Analysephase die Erstellung von SQL-Anfragen erforderlich, da die zu verwendenden fachlichen Daten untersucht und vollständig verstanden werden müssen. Dazu werden unter anderem Techniken des Data Profiling verwendet. Eine Auswahl dieser Anfragen kann erfahrungsgemäß direkt als Monitoreingabeparameter verwendet werden. Der Zusatzaufwand für die Erstellung eines DQ-Monitors ist aus diesem Grund sehr gering. Mit der Wahl des Moduls und der Angabe einiger weniger Parameter (siehe oben) ist ein Monitor damit startbereit und kann aktiviert werden. Es ist daher sinnvoll, dass bereits während der Implementierungs- und Testphasen möglichst viele Monitore aktiv sind. So werden Testläufe auf einfache Art und Weise automatisch überwacht und ausgewertet.

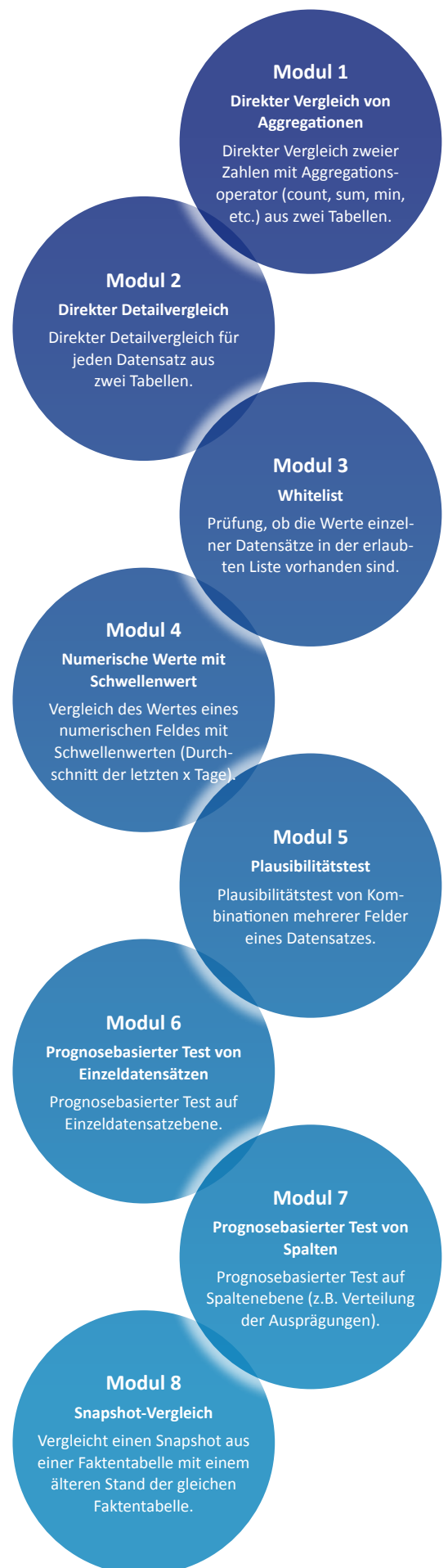


Abbildung 2: Module

DQ-Monitoring

Wird ein DQ-Problem durch einen Monitor erkannt, kann durch DQMP z. B. ein neues Fehlerticket eröffnet oder eine Personengruppe per E-Mail informiert werden. Der Betrieb der DQMP erfordert durch die automatische Überwachung praktisch gar keine Zeit. Lediglich das Einrichten der einzelnen Monitordefinitionen bzw. die Definition von Prüf-SQLs benötigt ein paar Klicks. Im Vergleich zum höheren Aufwand der Behebung von zu spät erkannten Datenqualitätsproblemen ist die Betreuung mit Hilfe der DQMP sehr effizient.

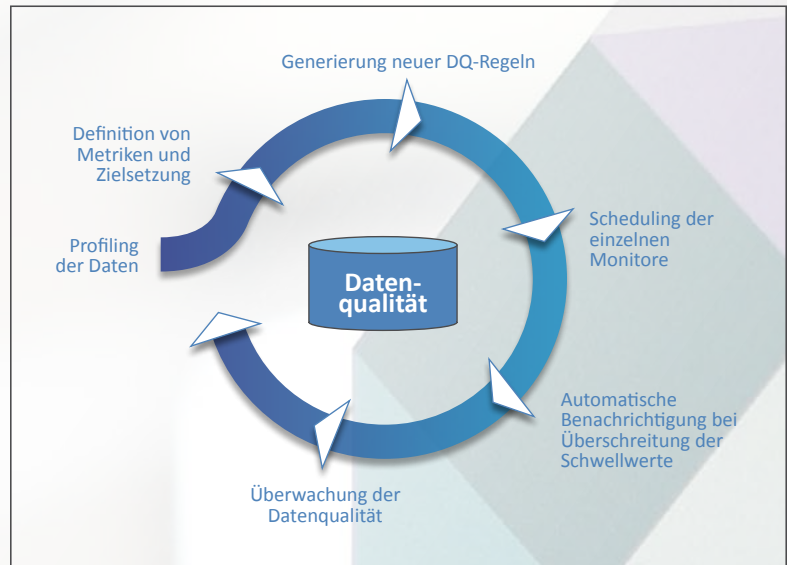


Abbildung 3: Entwicklungsprozess



Über die Autorin

Helene Fuchs ist Leiterin des Competence Centers Data Warehouse bei b.telligent und berät Kunden aus verschiedenen Branchen, wie z.B. der Telekommunikation, dem

Handel oder dem Bank- und Finanzwesen zu den Themen Data Warehousing und Business Intelligence seit nunmehr fünfzehn Jahren.

Über b.telligent

Die Unternehmensberatung b.telligent hat sich auf die Einführung und Weiterentwicklung von Business Intelligence, Customer Relationship Management, DWH und Big Data in Unternehmen in Massenmärkten spezialisiert.

Der Fokus liegt dabei auf der kontinuierlichen Optimierung

von Geschäftsprozessen, Kunden- und Lieferantenbeziehungen durch den Erkenntnisgewinn aus der Verdichtung, Analyse und Prognose von systemübergreifenden Geschäftsdaten. So lassen sich Margen erhöhen, Kosten senken und Risiken besser kontrollieren. www.btelligent.com

b.telligent

Walter-Gropius-Straße 17 • 80807 München
info@btelligent.com • Tel: +49 (89) 122 281 110

Helene Fuchs

Leiter Competence Center DWH
helene.fuchs@btelligent.com

Oliver Gräfe

Team Lead
oliver.graefe@btelligent.com

Follow us

