



Bild: © ktsdesign/shutterstock.com

» **Neue Werkzeuge** erledigen die bestehende Arbeit schneller und besser. **Neue Methoden** gestalten die Arbeit um, so dass insgesamt weniger zu tun ist. «

Neue Wege und Methoden für Business Intelligence

Agilität durch Automatisierung des Data Warehouse

Data Warehouse Automation bedeutet, dass die Entwicklung und Pflege der Datenaufbereitung aus wenigen Informationen heraus generiert wird und somit radikal beschleunigt. Diese Automation ist nicht ein Tool-Thema, sondern vor allem auch ein methodisches Thema. Neue Methoden im Bereich Business Intelligence, kurz BI, erlauben neben der Beschleunigung durch die Generierung auch eine höhere Agilität in der Bereitstellung von Daten und neue Wege in der Kommunikation mit dem Kunden.

Veränderungen im Bereich BI sind seit Jahren vor allem ein Thema von neuen Werkzeugen: ETL-Werkzeuge für eine schnellere Entwicklung der Datenaufbereitung, neue, schnellere Datenbanken oder Datenbanken für polystrukturierte Daten (NoSQL), um nur ein paar zu nennen. Dennoch hat sich im Kern nicht viel geändert. Die Zahl der Anforderungen übersteigt die Umsetzungsmöglichkeiten. Zudem entsteht beständig neuer Bedarf an Auswertungen und somit steigen die Anforderungen weiter.

Zeit für ein Umdenken. Es wird Zeit, das Problem grundsätzlich anzugehen. Eine BI-Lösung teilt sich immer in Datenbereitstellung und Datenauswertung. Die meiste Entwicklungszeit wird auf die Bereitstellung der Daten verwendet und darum braucht es an dieser Stelle ein Umdenken. Mehr noch als neue Werkzeuge braucht es neue

Methoden. Neue Werkzeuge erledigen die bestehende Arbeit schneller und besser. Neue Methoden gestalten die Arbeit um, so dass insgesamt weniger zu tun ist.

Dieses Umdenken hat stattgefunden. Es gibt neue Methoden für die Datenbereitstellung, für die Erstellung eines Data Warehouse, und von diesen Methoden hat in Deutschland in den letzten beiden Jahren sich vor allem Data Vault in Projekten bewährt.

Was macht Data Vault anders? Bei der Aufbereitung werden die Daten:

- ▣ **integriert:** Daten aus unterschiedlichen Datenquellen werden miteinander verknüpft
- ▣ **historisiert:** Daten werden dem Zeitablauf entsprechend gespeichert
- ▣ **konsolidiert:** gleiche Daten aus unterschiedlichen Quellen werden miteinander abgeglichen
- ▣ **interpretiert:** nicht zur Verfügung stehende Daten werden aus den vorhandenen Daten errechnet/ermittelt

Diese Arbeitsschritte sind heute nicht getrennt. Die Gründe dafür sind meist pragmatischer Natur oder weil eine schlechtere Performance befürchtet wird.

Standardisierung ermöglicht Automation. Eine Auftrennung der Verarbeitung nach genau diesen Gesichtspunkten erlaubt eine Standardisierung von Teilen der Verarbeitung. Und standardisierte Abläufe lassen sich automatisieren.

Die Trennung in Historisierung und Integration mit anschließender Konsolidierung und Interpretation (siehe Abbildung) erlaubt nun die Daten sehr schnell automatisiert zu historisieren und integrieren. Damit stehen die Daten bereit und können genutzt werden. Die Bereitstellung der Daten erfolgt ebenfalls standardisiert und auf die gleiche Art, egal ob die Daten mit oder ohne Konsolidierung/Integration bereitgestellt werden.

Schnelle Iterationen vermeiden Fehlmsetzungen. Die Qualität der so bereit gestellten Daten entspricht der Qualität der Datenquelle. Ohne Konsolidierung oder Interpretation sieht der Nutzer genau in welchem Zustand die Daten sind und bekommt einen genaueren Eindruck noch bevor alle Arbeiten abgeschlossen sind. Das ermöglicht ein frühes Feedback durch den Nutzer und unter Umständen eine Korrektur seiner Anforderungen an die Konsolidierungen noch bevor diese umgesetzt wurden. So lassen sich Fehlmsetzungen drastisch reduzieren.

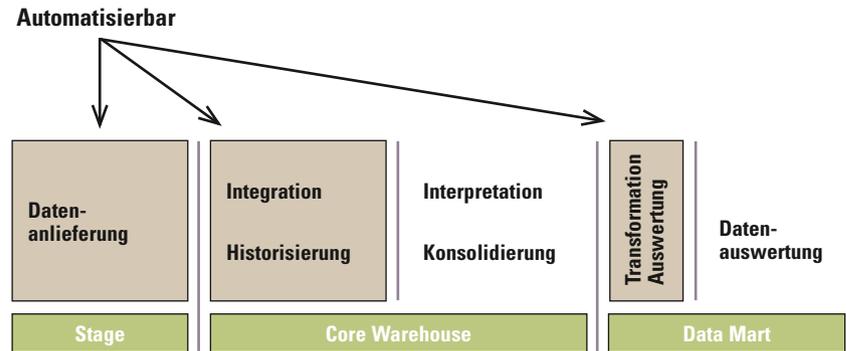
Begrenzen der Auswirkungen von Änderungen. Häufige Änderungen im Data Warehouse werden nicht durch die zu Grunde liegenden Daten verursacht. Diese ändern sich nur im Rahmen der Weiterentwicklung der liefernden Software oder bei Systemwechsel. Änderungen rühren vor allem aus den fachlichen Anforderungen. Durch die gewählte Reihenfolge der Datenaufbereitung müssen bei Änderungen an Konsolidierung und Interpretation nur klar umrissene Teilbereiche angepasst werden. Ein erneutes Laden der Daten aus der Datenlieferung entfällt. Die Starre der Datenmodelle im Data Warehouse löst sich auf und iterative beziehungsweise agile Ansätze sind möglich.

Grenzen der Automatisierung. Damit sind bis auf die Konsolidierung und die Interpretation der Daten alle Teile der Datenaufbereitung standardisiert. Für diese können lediglich Teile standardisiert werden. Eine grundsätzliche Automatisierung von Konsolidierung und Interpretation ist leider (noch) nicht möglich.

Die Kommunikation mit dem Fachbereich verändert sich. Die Kommunikation mit dem Fachbereich verändert sich mit diesem Vorgehen grundsätzlich. Der Aufwand für die Bereitstellung der Daten ohne Konsolidierung ist linear und lässt sich einfach berechnen. Erweiterungen dieses Standards sind Extras und sind als

Automatisierung durch Trennung der Abläufe

Quelle: MID



Die Trennung in Historisierung und Integration mit anschließender Konsolidierung und Interpretation erlaubt die Daten sehr schnell automatisiert zu historisieren und integrieren.

solche zu berechnen. Die Aufwände der IT werden transparenter und verständlicher. Das schafft Vertrauen.

Werkzeuge unterstützen. Die Automatisierung mit Data Vault folgt einem klaren Muster. Abläufe werden standardisiert und umgesetzt. Die standardisierten Bereinigungsprozesse werden dann mit Hilfe von Metadaten an die konkreten Daten angepasst.

Ein Werkzeug pflegt die Metadaten und generiert die konkreten Aufbereitungsprozesse. Die am Markt befindlichen Werkzeuge erledigen diese Arbeiten ganz unterschiedlich. Neben den großen und relativ teuren Werkzeugen zur Automatisierung haben sich noch günstige Erweiterungen von Datenmodellierungswerkzeugen durchgesetzt. Bei der Toolauswahl sollte jedoch die Unterstützung von Methode und Projektbedürfnissen im Vordergrund stehen.

Neue Methoden bringen moderne Möglichkeiten. Mit Data Vault kommen zusätzlich zum agilen Vorgehen und der erhöhten Geschwindigkeit neue Möglichkeiten hinzu. Ein Umstieg auf tägliches oder mehrfach tägliches Laden ist ohne Aufwand möglich, die Methode sieht dies von vornherein vor.

Die geladenen Daten haben alle Daten zu ihrer Herkunft, das komplette

Data Warehouse ist auditfähig. Es können die Datentransformationen für jeden einzelnen Satz nachverfolgt werden. Das ist nicht nur für Themen wie BCBS239 wertvoll.

Geschwindigkeit und Agilität. Eine Veränderung in der Methodik muss nicht immer auf einmal erfolgen. Data Vault eignet sich auch hervorragend, um ein Data Warehouse sukzessive umzustellen. So können neue Wege ausprobiert und der Erfolg unter Beweis gestellt werden. Nach und nach wird das Warehouse auf die neue Logik umgestellt.

Die Veränderung wirkt dann stückweise und in dem Maße, wie die Umstellung erfolgt ist, nimmt die Entwicklungsgeschwindigkeit zu. Die Kommunikation mit dem Fachbereich ändert sich und so können alte Strukturen aufbrechen und den Weg freimachen für konstruktive Zusammenarbeit und ein noch erfolgreicheres Warehouse.

Michael Müller



Michael Müller, Dipl.-Inf. (FH), ist Principal Consultant bei der MID GmbH und beschäftigt sich seit 2000 mit Business Intelligence, Data Warehousing und Data Vault. Seine Schwerpunktthemen sind Architekturen, Modellierung und modellgetriebene

Automation für Business Intelligence.