

# Das generierte Data Warehouse

Gregor Zeiler

Trivadis

Wien

## Schlüsselworte

Data Warehouse, Generator, DWH Lifecyclemanagement, Business Requirements Engineering.

## Einleitung

Die meisten Kosten für Konzeption, Implementierung und Betrieb eines Data Warehouses entstehen noch immer bei der Umsetzung von Standardaufgaben, die methodisch schon oft - vielleicht sogar im eigenen Hause - gelöst wurden. Im betrieblichen Alltag bleibt aber meistens, wenig Zeit für Vorgaben zur Standardisierung bzw. zur Qualitätssicherung und Dokumentation von derartigen Arbeitsschritten.

Vorhandene Datenintegrationswerkzeuge und DWH Funktionsbibliotheken unterstützen zwar bei der manuellen Implementierung vieler grundlegender technischer Aufgaben (SCD Handling, Quellsystemanbindung etc.). Der Graben zwischen fachlicher Anforderung und technischer Umsetzung ist aber nach wie vor breit und tief.

Wie weit Generator-Ansätze hier unterstützen können bzw. sollen, um den entsprechenden Nutzen zu stiften wird im gegenständlichen Vortrag erörtert. Wir betrachten dabei den ganzen Prozess von der Aufnahme der fachlichen Anforderungen über Aufbau und Betrieb des DWHs bis zur Definition des Reportings. Dabei steht aber nicht die Vollständigkeit des Ansatzes im Vordergrund sondern ganz pragmatisch dessen Effizienz und Effektivität.

## Die Erfolgsfaktoren von Data Warehouse-Lösungen

Agilität und Effizienz. Auf diese beiden Erfolgsfaktoren lassen sich die gestiegenen Anforderungen an Data Warehouse-Lösungen hinsichtlich schneller Einführung, begrenzter Budgets, einfacher Bedienung und effizienten Betriebes reduzieren. Sind die Grundlagen zu diesen Erfolgsfaktoren nicht ausreichend gegeben, schwindet das Vertrauen und die Akzeptanz bis zu kompletten Ablehnung und dem Misserfolg der BI-Lösung.

Mit **Agilität** verbinden wir folgende Eigenschaften:

*Neue Fachanforderungen rasch umsetzen zu können.* Dabei ist aber nicht nur die reine Implementierungszeit relevant, sondern die gesamte Durchlaufzeit von der Erarbeitung der Anforderung bis zur produktiv gesetzten Lösung. Denn das ist die Zeitspanne, die vom Bedarfsträger als Reaktionszeit wahrgenommen wird.

*Rasch und flexibel auf Änderungswünsche reagieren zu können.* Die Dynamik der Märkte erfordert oft kurzfristig das eigene Geschäft unter anderen/angepassten Perspektiven zu betrachten. Für die Data Warehouse-Lösung kann dies beispielsweise eine kurzfristige Anpassung von Dimensionen bzw. deren Struktur bedeuten. Auch hier ist die Zeitspanne von der Präzisierung des Änderungswunsches über die Analyse der Anpassungserfordernisse bis hin umgesetzten Lösung relevant.

*Skalierbarkeit in jeglicher Hinsicht.* Da die oben genannten Punkte auf sehr kurze Release-Zyklen abzielen, kommt es in der Praxis oft zu sogenannten „Schnellschüssen“, die nicht ausreichend erweiterbar oder skalierbar sind. Skalierbarkeit kann sich aber auch in einer Art Systematisierung

ausdrücken, die es erlaubt binnen kürzester Zeit eine neue Datenquelle einzubinden. Und Skalierbarkeit bedeutet natürlich auch mit den steigenden Datenvolumina entsprechend performant umgehen zu können.

Skalierbarkeit, der Anspruch auf Systematisierung und Standardisierung steht oft in krassen Widerspruch zu der Art, wie auf kurzfristige Umsetzungs- und Änderungsanforderungen reagiert wird. Um dauerhaft agil zu bleiben, ist Skalierbarkeit aber eine Grundbedingung, die unbedingt berücksichtigt werden muss.

Mit **Effizienz** verbinden wir folgende Eigenschaften:

*Kostengünstige(r) Umsetzung/Betrieb.* In Einzelfällen kann es vielleicht vorkommen, dass der Aufwand für die Umsetzung in keiner Relation zum Nutzen oder zur Dringlichkeit des Ergebnisses steht und daher auf „koste es was es wolle“ produziert werden kann. Die Regel ist aber hohes Kostenbewusstsein, sowohl für die Umsetzung, als auch für den Betrieb der Data Warehouse-Lösung.

*Expertise und Ressourceneinsatz.* Steigen die Komplexität und auch der Umfang der Lösung, so ist entsprechende Expertise in geeignetem Ausmaß gefragt. Dies kann bei manuellen Vorgehensweisen durchaus überproportionalen Bedarf auslösen. Sind im eigenen Haus nicht ausreichend Ressourcen verfügbar, ist man auf den Zukauf von Expertise bzw. Ressource angewiesen.

*Erfüllung/Abbildung der Anforderungen.* Nicht selten finden Fachbereiche und IT keinen gemeinsamen Nenner in der Diskussion der Fachanforderung und den technischen Umsetzungsmöglichkeiten. Daraus spitzen sich dann meistens zwei mögliche Szenarien zu. Fachanforderungen, die sich nicht oder nur mit sehr großem Aufwand umsetzen lassen, oder Umsetzungsergebnisse, die nicht den Bedarf treffen. Beides ist keine befriedigende Lösung.

*Datenqualität und fristgerechte Daten-Verfügbarkeit.* Das Vertrauen in die Auswertungsgrundlage hängt maßgeblich von der gebotenen Datenqualität ab. Ist diese einmal in Mitleidenschaft gezogen schwindet rasch die Akzeptanz und ist mühsam wieder herzustellen. Ähnlich zu bewerten ist auch die Aktualität der Datenverfügbarkeit. Sind entscheidungsrelevante Informationen nicht fristgerecht vorhanden, ist der Nutzen der Auswertungsgrundlage für den Anwender nicht gegeben.

Dass es in Bezug auf Agilität und Effizienz in Unternehmen große Unterschiede gibt, belegt eine Studie des US-Marktforschers Aberdeen. Die Analysen zeigen, dass zum Beispiel für die Integration einer neuen Datenquelle die besten Unternehmen durchschnittlich 12 Tage benötigen, während die Schlechten durchschnittlich 143 Tage für die gleiche Aufgabenstellung aufwänden. Die Besten schaffen auch 93 Prozent der Informationen innerhalb des gewünschten Zeitfensters bereit zu stellen, während die Trödler nur 34 Prozent anbieten können.

Die Marktforscher haben sich die Ursachen dafür angesehen und zunächst einmal den Einsatz von Technologie verglichen. Dabei zeigen sich folgende Unterschiede: 75 Prozent der Besten arbeiten mit Data Integration Tools, jedoch nur 46 Prozent der Trödler. 66 Prozent der Klassenbesten setzen außerdem Data Modelling Software ein und nutzen Data Cleansing Tools. Nur rund 30 Prozent der Trödler tun dies. Das erklärt den Erfolg der Musterfirmen aber nur zum Teil. Denn über den Einsatz solcher Tools hinaus geben zwei Drittel der Besten an, die Nutzung von BI-Werkzeugen systematisch zu überprüfen. Gut jeder Zweite (51 Prozent) erklärt, die Erfordernisse der Anwender in einem formalisierten Prozess zu erheben.

## **Ein einfacher Generator ist eindeutig zu wenig**

Es ist unbestritten, dass Anforderungen wie rasche und kostengünstige Umsetzung förmlich nach Automatisierung und Generierung schreien. Noch dazu lassen sich speziell im Data Warehousing Regeln und Methoden zur Umsetzung sowie Architekturvorgaben relativ einfach standardisieren. Die Anzahl der wiederkehrenden Aufgaben ist im Vergleich zu anderen Entwicklungsprojekten überproportional hoch. Dem Generatorbau steht demnach nichts im Wege.

Wie viel Unterstützung der Generator bieten kann, hängt davon ab, in welchem Umfang die Ergebnisse des Generators bereits nutzbar sind bzw. welcher Aufwand notwendig ist um das Ergebnis nutzbar zu machen.

Bei Routineaufgaben ist der Erfüllungsgrad durch einen Generator meistens sehr hoch und mit wenig bis keinen Nacharbeiten verbunden. Stellen sich jedoch kniffligere Umsetzungs Herausforderungen, kann sich schnell das Problem ergeben, dass der Aufwand mit Generatoreinsatz höher ist als ohne diesen. Damit kann sich der generelle Einsatz von Generatoren in Frage stellen. Um dies zu verhindern, muss der Generator geeignete Möglichkeiten und Vorkehrungen bieten, auf individuelle Anforderungen reagieren zu können, ohne den generellen Ansatz zu gefährden.

Der Einsatz eines Generators für sich, adressiert aber nur einen Teil der oben genannten Erfolgsfaktoren. Sicherlich kann man durch den Einsatz eine Implementierung beschleunigen und kostengünstiger durchführen. Am besten geht das auch für ein Projekt auf der grünen Wiese.

Kommt der nächste Aspekt der Agilität, nämlich die Änderungsflexibilität zum Tragen muss der Generator auch den Umgang mit bereits im Betrieb befindlichen DWH-Komponenten beherrschen. Tut er das nicht, ist das Einsatzszenario des Generators auf die Erstimplementierung beschränkt. Anders gesprochen, wird dieser Erfolgsfaktor vom Generator dann nicht unterstützt.

Je umfangreicher Methoden und Architekturansätze vom Generator unterstützt werden, desto geringer ist auch der Bedarf an tiefgreifender Expertise. Es ist einfacher die Parametrisierungsoptionen für verschiedene Umsetzungsszenarien zu verstehen, als detailliert über jeden erforderlichen manuellen Umsetzungsschritt Bescheid zu wissen.

Den größten Fortschritt erzielt man jedoch, wenn man die Tool-Unterstützung über die gesamte Prozesskette von der Anforderungsanalyse bis zum Generieren der Data Warehouse Komponenten erstrecken kann.

In der Praxis ergibt sich hier oft das Problem, dass Fachanforderungen relativ unstrukturiert in Fachkonzepten niedergeschrieben werden. Damit ein Generator zum Einsatz gebracht werden kann, muss der Input für den Generator aber in strukturierter Form vorliegen. Die Fachanforderungen sind daher erst in eine geeignete Form zu übersetzen. Dabei kann es zu Definitionslücken, oder auch zu Miss-Interpretationen kommen.

Eine Tool-Erweiterung, um auch das Business Requirements Engineering für die DWH-Lösung unterstützen zu können, ist daher naheliegend und hilft die Release-Zyklen kurz zu halten und gewährleistet die Treffsicherheit der Umsetzung in Bezug zu den Anforderungen.

Im Vortrag wird gezeigt welche Funktionen zur bestmöglichen Unterstützung der beiden Erfolgsfaktoren Agilität und Effizienz notwendig sind und wie diese in konkreten Projektbeispielen zum Einsatz kommen können.

## DWH-Lifecyclemanagement

Die folgende Abbildung zeigt Funktionsmodule und deren Zusammenwirken einer DWH-Lifecyclemanagement Lösung. Im Vortrag wird auf die Funktionsweise der einzelnen Module und das Zusammenwirken mit Oracle-Technologien eingegangen.

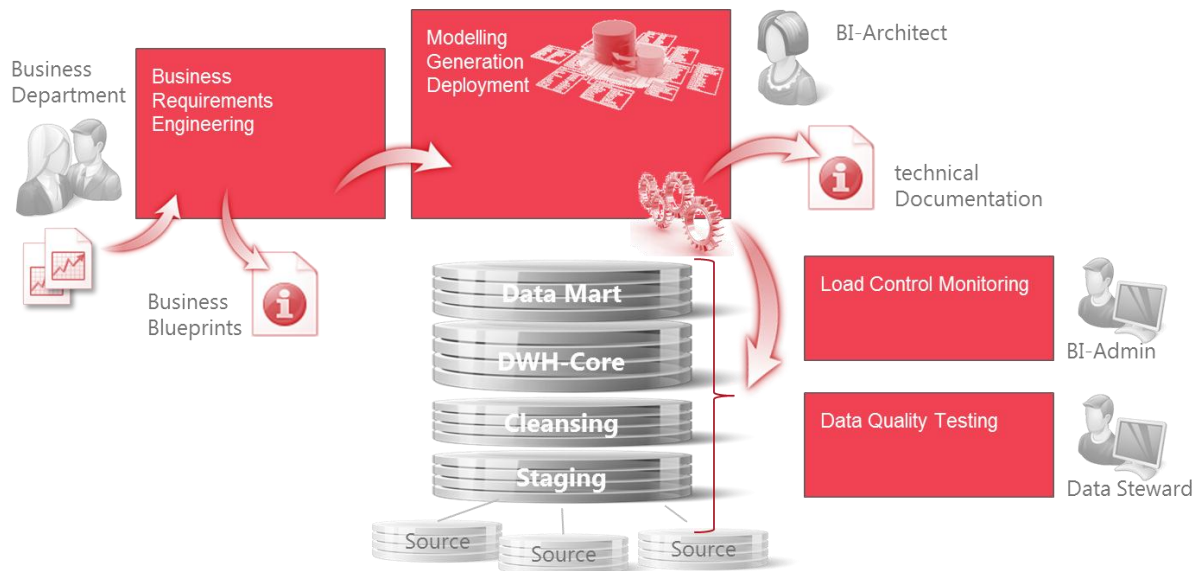


Abb. 1: Module DWH-Lifecycle Management

### Kontaktadresse:

Gregor Zeiler  
Trivadis Delphi GmbH  
Handelskai 94-96  
A-1200 Wien

Telefon: +43 (0) 1-332 35 31-DW  
Fax: +43 (0) 1-332 35 34  
E-Mail: [gregor.zeiler@trivadis.com](mailto:gregor.zeiler@trivadis.com)  
Internet: [www.trivadis.com](http://www.trivadis.com)