

# Start small, think big

**Dr. Andrea Kennel  
InfoPunkt Kennel GmbH  
CH-8600 Dübendorf**

## **Schlüsselworte:**

Business Intelligence, Data Warehousing

## **Wieso wollten wir ein DWH?**

In einer Firma, die im Bereich Forderungsmanagement arbeitet, sammeln sich viele Daten über Fälle und Zahlungen an. Um die Ziele der Firma besser kontrollieren zu können, will man aus den vielen Daten Kennzahlen ermitteln, um beispielsweise den aktuellen Umsatz mit dem des Vormonats oder des Vorjahres zu vergleichen.

Das Bedürfnis nach Kennzahlen besteht aber in fast allen Abteilungen. So entstanden verschiedene Reports, die verschiedene Kennzahlen ausweisen. Dies führte, wie wohl in so mancher Firma, zu einem technischen und einem fachlichen Problem. Technisch wird die produktive Datenbank unnötig mit immer wieder sehr ähnlichen Abfragen und Berechnungen belastet. Fachlich liegt das Problem darin, dass die verschiedenen Abteilungen zu unterschiedlichen Kennzahlen kommen, weil die Kennzahlen in jeder Abteilung leicht anders definiert werden. So wird es schwierig, die Unterschiede der Kennzahlen zu finden und begründen.

Für beide Probleme kann ein Data Warehouse (DWH) die Lösung sein. Im DWH werden normalerweise nicht die Basisdaten, sondern direkt Kennzahlen gespeichert, so dass die Kennzahlen einheitlich definiert werden müssen. Weiter werden die Daten nur einmal aus der Produktion selektiert, ins DWH geschrieben und dann so oft gelesen wie gewünscht. Das Lesen der Daten belastet dann die Produktion nicht mehr.

## **Ziele des Projektes und gewählte Projektstruktur**

In wohl jedem DWH Projekt gibt es den Zielkonflikt zwischen einem schnellen ersten Resultat und einem DWH, das alle erdenklichen Anforderungen abdeckt.

In unserem Projekt haben wir uns bewusst dafür entschieden, im ersten Schritt nur einen Teil der Anforderungen abzudecken. Dies aber mit dem klaren Ziel, ein System zu bauen, das schrittweise ausgebaut werden kann. Wir haben unser DWH daher DWHPILOT getauft. So war auch für die Benutzer von vornherein klar, dass wir „nur“ einen ersten Schritt machen. Gleichzeitig haben wir aber die Architektur so gewählt, dass sie einfach ausbaubar ist. Dabei haben wir mit vereinheitlichten Dimensionen und Fakten gearbeitet, wie das Kimball [KIM02] als „Common Dimensions“ und „Common Fact“ empfiehlt.

Die Projektstruktur haben wir bewusst klein gehalten. Doch auch für ein kleines DWH-Projekt ist es wichtig, dass einige Rollen Personen zugeteilt werden:

In unserem Projekt war der Business Sponsor gleichzeitig ein sehr erfahrener Business Experte. Weiter brauchte auch der Projektmanager ein breites Business Wissen mit. Die Aufgaben des DWH-Architekten, -Umsetzers und -Educators wurden weitgehend von ein

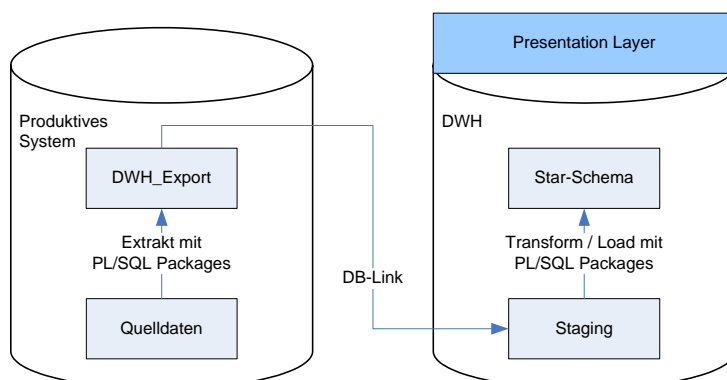
und derselben Person übernommen. Damit aber die Qualitätskontrolle nicht zu kurz kommt, hat vor allem der Projektmanager die Aufgabe des Reviewers übernommen. So konnten mit wenigen Personen alle Funktionen kompetent abgedeckt werden.

### **Spezifikationsphase: Vorgehen nach Kimball**

Bei der Spezifikation ist es gemäss Kimball [KIM02] wichtig, dass Dimensionen und auch Fakten einheitlich definiert werden. Konkret bedeutet das, dass eine Dimension für mehrere Fakten gebraucht werden kann. Wenn also Erträge und Kosten je Fall berechnet und ausgewertet werden können, so muss dafür dieselbe Dimension genutzt werden. Nur so ist es möglich, auch über eine Gruppierung von Fällen den Ertrag den Kosten gegenüber zu stellen. Auch Fakten müssen vereinheitlicht werden: wenn ein Faktum "Umsatz" benannt wird, muss dies für alle Auswertungen und vor allem für alle Abteilungen dieselbe Bedeutung haben. Die Definition der vereinheitlichten Dimensionen verlief erstaunlich einfach. Die Definition der Fakten war schwieriger. Hier zeigte sich, dass nicht alle Abteilungen die vermeintlich gleiche Kennzahl auch gleich definieren. So mussten teils neue Namen für Kennzahlen definiert werden, die präziser sind. Weiter wurde in der Spezifikation klar festgelegt, wie die Kennzahlen fachlich und technisch definiert sind. Diese Spezifikation wurde von allen Beteiligten abgenommen bevor sie umgesetzt wurde.

### **Implementationsphase: Architektur, Tooleinsatz**

Für die Wahl der Architektur war für uns wichtig, dass im produktiven System keine Daten gelöscht werden, also die komplette Historie enthält. So dient das DWH nur für die Auswertung, nicht aber gleichzeitig für die Archivierung. Konkret heisst das, dass jederzeit alle Daten des DWH aus der Produktion geladen werden könnten. Somit kann auch die Architektur etwas vereinfacht werden, da nicht zwischen dem eigentlichen Datenlager (data storage) und den Marts für die Auswertung unterschieden werden muss. So wurde folgende Architektur gewählt:



*Abb. 1: Architektur der DWH-PILOTEN*

Die Logik zur Ermittlung der Kennzahlen ist im Quellsystem implementiert. Die Daten für das DWH werden dann in speziellen Export-Tabellen gespeichert.

Diese Tabellen werden über DB-Link ins DWH geholt und dort in die Marts geschrieben. So wurden die Fakten und Dimensionen, die sich aus der Spezifikation ergaben, direkt als Marts im Star-Schema umgesetzt.

Durch die klare Aufteilung in Extrakt und Load konnte die Komplexität reduziert werden. So ist beispielsweise die Logik, wie die Dimensionen genau gelesen werden müssen, im Extrakt verpackt und die Versionierung, dort wo erforderlich, im Load.

Als Präsentationsschicht haben wir Business Objects gewählt, da dieses bereits im Einsatz ist.

### **Testphase und Schulung**

Für die Testphase wurden Reports erstellt, die teils gleiche Zahlen darstellen wie bestehende Reports. So war es möglich Differenzen zu finden und zu analysieren. Dabei konnten in einigen Fällen Missverständnisse behoben und das DWH korrigiert werden. In anderen Fällen mussten wir feststellen, dass die alten Reports in wenigen Details nicht präzise waren und das DWH verlässlichere Zahlen liefert.

Gleichzeitig mit dem Testen wurden auch weitere Reports aufgebaut und die Benutzer geschult. Dabei wurde unterschieden zwischen Benutzern, die vordefinierte Reports mit Parametern benötigen, und Benutzern, die selbst neue Reports zur vertieften Datenanalyse erstellen müssen.

Weiter wurde das tägliche Laden der neuen Daten bewusst erst während der Testphase fertig implementiert. So konnte Wartezeit verhindert werden, respektive die Tests früher begonnen werden.

### **Betrieb inkl. Optimierung**

Während der Testphase wurden nicht nur die geladenen Daten geprüft, sondern auch das tägliche Laden der Änderungen, der sogenannte Delta-Load. Da wir für die Tests aber nicht die Produktion, sondern die Testumgebung als Quelle zur Verfügung hatten, konnten die Delta-Loads nur simuliert werden.

Dies hatte den Nachteil, dass insbesondere die Performance nicht richtig getestet werden konnte. So gab es dann auch anfangs im Betrieb ein paar kniffligen Performance-Probleme zu lösen. Vor allem stolperten wir über das Problem der Bind-Variablen. Hier weiss der Optimizer nicht genau, welche Werte den Bind-Variablen übergeben werden und geht von einem Durchschnitt aus. Für den Delta-Load ist aber klar, dass mit „between“ auf etwa einen Tag eingeschränkt wird. Da wir aber dieselbe Prozedur auch für den gesamten Load über eine grössere Zeitperiode brauchen, war es nicht möglich mit Hints zu arbeiten. So sahen wir uns gezwungen, für solche Fälle in PL/SQL mit dynamischen SQLs zu arbeiten. Dann können die Parameter direkt in das SQL eingebunden werden und der Optimizer findet einen schnelleren Weg.

### **Fazit organisatorisch**

Es hat sich gezeigt, dass es einfacher ist erste Anforderungen an ein DWH zu formulieren, wenn nicht zu viele Personen mitreden müssen. Sicher ist es aber wichtig, die richtigen Personen an Bord zu haben, so dass schon am Anfang zentrale Fakten und Dimensionen ins DWH integriert werden können.

Es lohnt sich also, nicht zu viele, aber die richtigen Personen einzubeziehen.

Weiter hatten wir in unserem Projekt den Vorteil, dass es vom Topmanagement abgesegnet und auch klar unterstützt wurde, nicht aber im Rampenlicht stand. So konnte ohne grossen Druck, aber mit der nötigen Unterstützung gearbeitet werden.

Da mit einem Reporting Tool Reports relativ einfach über Parameter gesteuert oder bearbeitet werden können, war der Nutzen eines DWH für die Benutzer bereits während der Testphase klar sichtbar und spürbar..

So wandelte sich die anfängliche Skepsis einiger Benutzer in Unterstützung. Erfahrungen für das weitere Vorgehen konnten gesammelt werden. Die Richtung ist klar, der Nutzen auch.

### **Fazit technisch**

Der Ansatz, bei der Architektur keine Abstriche vorzunehmen und von Anfang an mit einer üblichen DWH-Architektur zu arbeiten, war sicher richtig. Einerseits ist es nun relativ einfach möglich, den Piloten zu einem grösseren DWH auszubauen. Andererseits waren bereits in den ersten Anforderungen komplexere Aufgaben zu lösen. Da hilft es, wenn Fakten und Dimensionen klar getrennt geladen werden können und das Laden in die Schritte Extraktion, Transformation und Ladeprozess unterteilt werden kann. So wird die Komplexität zerschlagen und auch die Fehlersuche vereinfacht.

Ein weiterer wichtiger Punkt ist die Performance. Bei der Implementierung achteten wir von Anfang an konsequent auf Performance, was sich bewährt hat.

### **Zusammenfassung**

Der Ansatz „Start small, think big“ hat sich in unserem Projekt sicher gelohnt. Damit aber ein Projekt bewusst klein startet, braucht es Mut. Weiter ist es auch wichtig, dass eine Firma den Mut hat, neue Konzepte konsequent umzusetzen. Mein Auftraggeber ist hier als mittelgrosses Unternehmen wohl führend. Werden doch in verschiedenen Projekten konsequent neue Konzepte eingesetzt, dies mit Erfolg.

Dank des kleinen Starts mit einem Piloten war es möglich, bereits 6 Monate nach dem KickOff den DWHPILOT produktiv zu setzen und zu nutzen. Dank der klaren Konzepte ist es aber auch möglich, den Piloten nun schrittweise zu einem umfangreicheren DWH auszubauen. Dies ist wichtig, da bekanntlich ein DWH, das genutzt wird, immer auch neue Anforderungen erfüllen muss. Frei nach dem Motto mit dem Essen kommt der Hunger.

So habe ich mit diesem Projekt auch die Erfahrung gemacht, dass man eben Essen geben muss, solange der Hunger noch geweckt werden kann. Sprich: die ersten Erfolge sollen schnell sichtbar sein, dann bringt das Projekt schnell Nutzen und erhält damit auch Unterstützung.

### **Literatur:**

[KIM02] The Data Warehouse Toolkit. Ralph Kimball. Wiley 2002. ISBN 0-471-20024-7

### **Kontaktadresse:**

#### **Dr. Andrea Kennel**

InfoPunkt Kennel GmbH  
Bahnhofstr. 48  
CH-8600 Dübendorf

Telefon: +41(0)4-820 71 40  
E-Mail: [anrea@infokennel.ch](mailto:anrea@infokennel.ch)  
Internet: [www.infokennel.ch](http://www.infokennel.ch)