

Einführung eines Data Warehouse in dezentral organisierten Unternehmen

Lutz Rothmann, GFT Technologies AG

Die Ersteinführung eines Data Warehouse stellt grundsätzlich eine besondere Herausforderung dar. Bei dezentral organisierten Unternehmen, die ein Konglomerat aus weitgehend selbstständigen Organisationen sind, die vielleicht sogar teilweise miteinander im Wettbewerb stehen, gilt es ein paar Besonderheiten zu beachten.

In diesem Artikel geht es um dezentral organisierte Unternehmen wie beispielsweise eine Profit-Center-Organisation. Als Profit-Center organisierte Unternehmenstöchter stehen zumindest teilweise im Wettbewerb zueinander. Sie werden aus einer Konzernzentrale gesteuert, behalten aber große Selbstständigkeit und eigene Entscheidungsfreiräume. Kosten und Gewinne werden über Schlüssel auf unterschiedliche Unternehmensbereiche verteilt. Die Schlüssel berücksichtigen oft Kennzahlen (KPIs), die die Leistungen der Unternehmensteile widerspiegeln. Unternehmens-eigner oder andere außenstehende Organe interessieren sich im Allgemeinen zunächst für eine Gesamtbetrachtung des Konzerns (Erfolg, Wachstum, Risiken). Im zweiten, detaillierteren Schritt werden einzelne Unternehmensteile betrachtet.

Die dezentralen Unternehmensstrukturen sind in nahezu allen Branchen vertreten. Bekannte Beispiele findet man in der Telekommunikation, in der Energiebranche, in den Bereichen Transport und Spedition, bei Banken und Versicherungen. Sie entstehen beispielsweise im Rahmen von Konsolidierungsbewegungen, ausgelöst durch Marktdruck oder Regulation, wo selbstständige Unternehmungen mit ähnlichen, nicht identischen Geschäftsmodellen zu Konzernen zusammengefasst werden. Solche Strukturen können auch bewusst organisch entstehen, etwa wenn Unternehmen in den internationalen Markt wachsen oder Vielmarken-Strategien fahren. Dezentral organisierte Unternehmen sind berichtsintensiv:

- Zentrale Unternehmensführung und externe Organe benötigen eine integrierte

und aggregierte Sicht auf das Gesamtunternehmen, um die Leistung des Gesamtkonzerns beurteilen zu können und die Strategie entsprechend daran auszurichten.

- Zentrale Unternehmensführung und externe Organe benötigen die Profit-Center-Sicht als Einzelbetrachtung (jedes Profit-Center für sich) und als Wettbewerbssicht (ein Profit-Center im Vergleich zu den anderen). Daraus können beispielsweise steuernde Maßnahmen wie Konsolidierungen oder operative Größen wie Verteilungsschlüssel abgeleitet werden.
- Die Unternehmenstöchter agieren selbstständig wie eigene Unternehmen mit eigenen Interessen und eigenen Perspektiven.

Der Bedarf nach Geschäftsdaten mit folgenden Eigenschaften liegt auf der Hand:

- Vergleichbarkeit (übergreifende Ordnungskriterien)
- Integrierbarkeit (übergreifende Metrik)
- Aggregierbarkeit (Wechsel in verschiedene Detail-Ebenen)

Eine gängige Lösung ist die Nutzung eines zentralen Data Warehouse, das Daten aus verschiedenen Unternehmensteilen regelmäßig automatisch aufbereitet und zentral bereitstellt. Die Beauftragung für ein solches Data Warehouse, welches auch übergreifende Analysen unterstützt, kommt üblicherweise aus der Unternehmenszentrale, während Profit-Center sich unter Umständen schnell und selbstständig

ihre eigenen Data Warehouses organisieren.

Die Herausforderung bei der Einführung eines zentralen Data Warehouse in dezentral strukturierten Unternehmen resultiert aus der historischen Eigenständigkeit der Unternehmenstöchter: Unterschiedliche Geschäftsmodelle führen zu unterschiedlichen Abbildungen der Transaktionen, auch wenn die erste Eingangsgröße (wie Kundenauftrag) und die daraus folgende finale Ausgangsgröße (etwa eine Lieferung) nicht davon abhängen, welche Unternehmensteile an der Abwicklung beteiligt sind.

Man betrachte – ohne zu tief in technische Details einzusteigen – eine gängige Data-Warehouse-Architektur, wie sie komplett als Oracle-Stack abgebildet werden kann (BI Enterprise Edition, Database, Linux, Warehouse Builder etc.). Das Data Warehouse soll nach dem klassischen „Hub and Spoke“-Ansatz aufgebaut sein (siehe Abbildung 1):

- Ein ETL-Framework (wie der Oracle Warehouse Builder oder Oracle Data Integrator und Shell-Skripte) extrahiert regelmäßig Daten aus verschiedenen bestandsführenden Systemen und überführt diese Daten in eine zentrale Datenbank.
- Innerhalb der zentralen Datenbank (beispielsweise von Oracle) werden die Daten weiterverarbeitet.
- Das ETL-Framework extrahiert Daten aus der zentralen Datenbank und überführt diese in spezielle dezentrale, auf BI-Anwendungen (beispielsweise Reporting) spezialisierte Datenbanken, sogenannte „Data Marts“.

- Ein (oder mehrere) Anwendungsserver wie die BI Enterprise Edition stellt ein Analyse-Portal zur Verfügung (Answers/Dashboards/Delivers/BI Publisher). Dieses wird über einen Web-Client vom Benutzer angesprochen, organisiert Daten und bereitet diese auf. Das Portal stellt auch administrative Dienste bereit.

Anhand dieser stark vereinfachten, plakativen Darstellung (siehe Abbildung 1) lassen sich – angelehnt am klassischen Vorgehen der Data-Warehouse-Implementierung – gängige Fragen stellen und mögliche Antworten aus der Perspektive der Profit-Center-Organisation diskutieren.

Fachliche Daten-Modellierung

Ein Blick auf das Datenmodell, welches dem zentralen Datenspeicher zugrunde liegt: Im Enterprise-Data-Warehouse dient es als Quellstruktur für die spezialisierten Data Marts. Letztere können zum Beispiel von einzelnen Organisationseinheiten bestellt und komplett nach ihren Bedürfnissen gestaltet werden. Ein Data Mart muss nicht zwingend Strukturen darstellen, die auf alle Unternehmensteile passen. Für die zentrale Datenbank gibt es zunächst zwei Alternativen:

- Datenmodell je Geschäftsmodellvariante (Quellsystem)
- Zentrales Datenmodell

Die erste Alternative hat den Vorteil des zunächst geringeren Modellierungsaufwands. Für jedes Quellsystem wird ein eigenes Modell (Schema) bereitgestellt. Die notwendige Integrations- und Transformationsarbeit für die Gesamtsicht wird komplett auf die Schnittstelle „Datenbank / Data Mart“ verlagert. Das kann eine sinnvolle Alternative sein, wenn das Thema „integrierte Darstellung“ grundsätzlich keine Rolle spielt. Innerhalb eines Unternehmens dürfte dies aber nicht der Fall sein. Man kann eine logische Schicht zwischen Datenbank und Data Marts vorsehen, sollte diese aber nicht grundsätzlich für die komplette Integration verwenden, sondern eher für die Anwendung fortgeschrittener Geschäftsregeln.

Die zweite Alternative – zentrales Datenmodell – ist im Allgemeinen zu bevorzugen. „Zentrales Datenmodell“ bedeutet nicht, dass die Daten aus den unterschiedli-

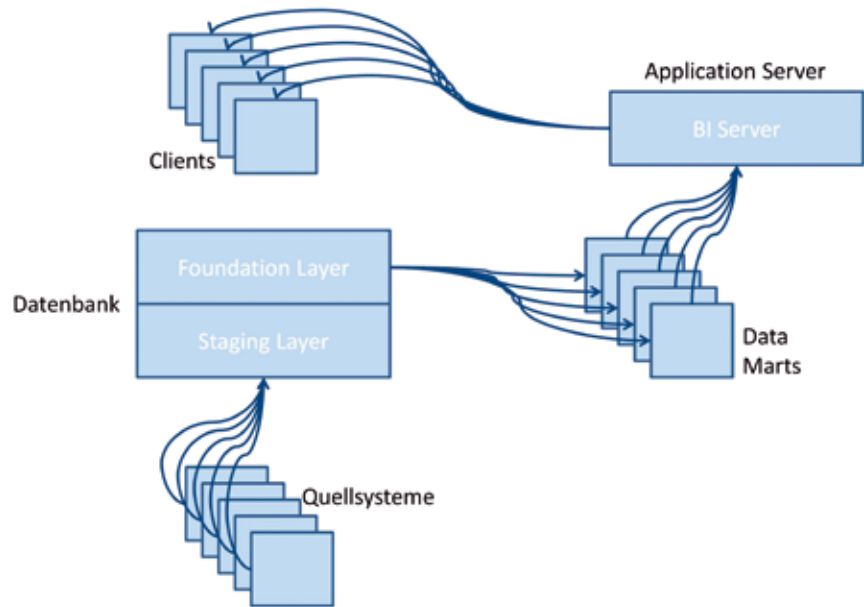


Abbildung 1: Data-Warehouse-Referenzarchitektur

chen Unternehmensbereichen untrennbar miteinander vermischt werden müssen. Es besagt, dass man die Daten in eine einheitliche Struktur (Modell) überführt. Ähnlich wie bei der Historisierung können die Daten nach ihrem Ursprung gekennzeichnet sein. Das lässt die Option offen, aus dem zentralen Datenpool Data Marts exklusiv mit Daten einer bestimmten Tochterorganisation zu erstellen – etwa für deren eigenes Data Warehouse.

Die Erstellung des zentralen Datenmodells ist insbesondere bei dezentral organisierten Unternehmen sehr aufwändig. Die einzelnen Organisationseinheiten lösen sich nicht ohne Weiteres von ihren vertrauten Sichtweisen auf Prozesse, Transaktionssysteme und Datenmodelle. Es ist an dieser Stelle essenziell, alle betroffenen Fachbereiche bei der Modellentwicklung einzubinden und ein strikt zielorientiertes, am konkreten fachlichen Bedarf ausgerichtetes Modell zu entwickeln. Auch wenn die Fachbereiche letztendlich nicht direkt mit der zentralen Datenbank arbeiten, müssen sie im Laufe der Implementierung des Data Warehouse laufend eingebunden werden, informieren und entscheiden.

Dieses Modell muss detailliert ausgearbeitet, dokumentiert und akzeptiert sein. Es dient über das initiale Implementierungsprojekt hinaus als „Wörterbuch“ zwi-

schen technisch und fachlich orientierten Betroffenen des Data Warehouse. Darüber hinaus unterstützt es grundsätzlich im Dialog zwischen Technik und Fachbereich, etwa bei Konzeption und Anforderungs-Analyse.

Datenbeschaffung

Dezentral organisierte Unternehmen werden oft von einer Vielzahl von Dienstleistern unterstützt, die Soft- und Hardware-Systeme betreiben. Beispiele für extern betreute Komponenten sind Speicherplatz, Betriebssysteme, Datenbanken und Anwendungen. Das sind Systeme, die die Quell-Landschaft des Data Warehouse betreffen. Kommunikation ist in diesem heterogenen Umfeld je nach vereinbartem Servicelevel zeitlich und finanziell sehr aufwändig.

Erschwert wird die Situation dadurch, dass – bei Beauftragung des Data Warehouse durch die Unternehmenszentrale – der Auftraggeber die Daten aus den Quellsystemen nicht kennt und daher nicht weiß, welche Daten er benötigt. Ein Ansatz, um Zeitverluste durch komplexe Abstimmungen im Rahmen zu halten, ist die Beauftragung einer „Maximal-Extraktion von Daten. Dieser „Schrotflintenansatz“ geht davon aus, dass man mit gesundem Fachverstand in einer sehr frühen Implementierungsprojektphase recht zügig und sicher einen Anteil an für das Data Warehouse

Unsere Inserenten

Cirquent GmbH www.cirquent.de	U 2
Hays AG www.hays.de	Seite 15
Hunkler GmbH & Co. KG www.hunkler.de	Seite 21
Huss-Verlag GmbH www.logistik-heute.de	Seite 17
ORACLE Deutschland B.V. & Co. KG www.oracle.com	U 4
PROMATIS software GmbH www.promatis.de	Seite 7
OPITZ CONSULTING GmbH www.opitz-consulting.de	Seite 3
Retarus GmbH www.retarus.com	Seite 4
WIN-Verlag GmbH & Co. KG www.digital-business- magazin.de	Seite 27

unwichtigen Daten im Quellsystem identifizieren kann. Unter der Voraussetzung, dass Lieferformat und -struktur der Daten sowie Sicherheitsaspekte geklärt sind und die Quellsysteme inklusive der Netzwerke in der Lage sind, die Daten zu vereinbarten Terminen zu liefern, kann die Arbeit an der Datenlieferung von den einzelnen Organisationseinheiten autark und parallel zum Restprojekt erfolgen.

Der Ansatz ist vertretbar, da der Aufwand für die Entwicklung der Extraktion (das „E“ in „ETL“) im Verhältnis zur Entwicklung der Beladung recht gering ist. Durch die Ausweitung der Extraktion auf aktuell noch nicht notwendige, aber später vielleicht interessante Daten werden die Projektaufwände normalerweise nicht spürbar erhöht. Die Selektion und Transformation der Daten erfolgt innerhalb des Data-Warehouse-Bereichs. Der Vorteil dieses Ansatzes liegt neben der frühen parallelen

Abwicklung im Projekt darin, dass in künftigen Ausbaustufen des Data Warehouse die notwendigen Quelldaten schon bereit liegen. Ein Änderungsauftrag für die Extraktion ist nicht notwendig. Der Nachteil liegt im höheren Datendurchsatz im Netzwerk und in dem höheren Speicherbedarf auf verschiedenen Stationen zwischen Quellsystem und „Endlager“.

Datenbeladung

Die Gestaltung der Datenbeladung inklusive Abbildung der Daten (Mapping), Skript-Erstellung, Performance-Optimierung etc. ist der aufwändigste Teil der Implementierungsprojekte. Hier kommen bislang unentdeckte Missverständnisse und Lücken zum Vorschein und Vertreter unterschiedlichster Gruppenzugehörigkeit müssen Details miteinander abstimmen. Die gute Qualität des zentralen Datenmodells und dessen Entwicklung bewährt sich besonders an dieser Stelle.

Als Besonderheit der dezentral organisierten Unternehmen taucht die Frage auf, wie stark die Daten im Zielsystem zu integrieren sind und wo die Integration stattfindet. Es gibt zwei Extrempositionen:

- Die Daten sind vollständig integriert; die Spezifika der Quellsysteme sind im zentralen Datenspeicher nicht mehr erkennbar
- Die Daten werden voneinander getrennt gehalten; Extraktion und Beladung sind theoretisch reversibel

Bei der vollständigen Integration werden beim Import der Quelldaten stärkere Bereinigungsprozesse angewandt. So werden beispielsweise Dubletten identifiziert und bereinigt, Stammdaten definiert, identifiziert und vereinheitlicht. Den Daten ist im zentralen Data-Warehouse-Speicher nicht mehr ihr systemischer Ursprung anzusehen. Das setzt eine fortgeschrittene Reife der Organisation voraus. Die zu betrachtenden Kenngrößen müssen in allen Unternehmensteilen übergreifend bekannt, verfügbar, definiert und akzeptiert sein. Es gibt dort eine einheitliche, akzeptierte Metrik.

Es gibt dezentral orientierte Unternehmen ohne allgemein akzeptierte Metrik. Das kann beispielsweise an Unterschieden

bei den Geschäftsmodellen der Töchter liegen. In diesen Unternehmen geht der Verabschiedung von Kennzahlen ein oft sehr diskussionsintensiver Einigungsprozess voraus. Ein entsprechend hoher und schwer kontrollierbarer Abstimmungsaufwand ist auch beim Entwurf der Datenintegration zu erwarten. Entwirft man das Data Warehouse nun so, dass die Datensätze nach Quellsystem getrennt in das Ziel-Datenmodell einsortiert werden, kann man den hohen Abstimmungsaufwand bei Gestaltung der Beladung deutlich entschärfen. Man hat dann eine Integration in zwei Schritten:

1. Umsortierung vom Quellsystem in das zentrale Enterprise-Schema
2. Transformation und Bereinigung vom zentralen Enterprise-Schema in die integrierten Data Marts

Die Integration kann über eine Zwischenschicht, den „Business Logic Layer“, zwischen dem zentralen Enterprise-Schema und den Data Marts erfolgen, da dieselbe Integrationslogik wohl auf mehrere unterschiedliche Data Marts angewandt werden wird. Dieser „Business Logic Layer“ muss nicht unbedingt nur aus Transformations-Routinen bestehen, er kann – je nach Design – die Ergebnisse auch dauerhaft speichern und vorhalten. Vorteile der grundsätzlich integrierten Datenhaltung sind einerseits der oft deutlich geringere Speicherbedarf und die einfachere (billigere, schnellere, robustere) Bereitstellung der Daten. Die „geschichtete“ Haltung der Daten – Datensätze nach Quellsystem getrennt – verbraucht normalerweise deutlich mehr Speicher und führt zu einer komplexeren Data-Warehouse-Architektur. Andererseits stehen die Daten bei Bedarf sauber getrennt zur Verfügung. Jedes Datum ist konkret auf die Quelle zurückzuführen und bietet damit erweiterte Analysemöglichkeiten. Diese können beispielsweise für Qualitätsanalysen genutzt werden, die auf Eigenheiten der Quellsysteme abzielen (Vergleich der Quellsysteme, Vergleich Data Warehouse mit Quellsystem). Damit können neue, übergreifende Kennzahlen entwickelt und verifiziert werden.

Die getrennte Haltung hilft auch, einem Data-Warehouse-„Wildwuchs“ entgegen-

2 Ausgaben
gratis

Lösungen für Geschäftsprozesse

**Immer die Nase vorne
mit einem persönlichen Abonnement**

www.digital-business-magazin.de/abo



zuwirken. Die Tochterunternehmen können Berichte oder Data Marts mit ausschließlich ihren Daten anfordern. Das erlaubt unter Umständen schon die Nutzung von „Quickwins“, bevor alle Diskussionen zur Daten-Integration abgeschlossen sind. Welche der genannten Alternativen bevorzugt wird, hängt letztendlich davon ab, ob sich der offensichtliche Mehraufwand für die getrennte Haltung rechnet. Ist diese Voraussetzung dafür, ein zentrales Data Warehouse implementieren zu können? Werden die erweiterten Analysemöglichkeiten gewinnbringend genutzt?

Lessons Learned

Kommunikation ist (fast) alles. Das ist eine nicht gerade neue Erkenntnis im Data-Warehouse-Umfeld, in dem es auch darum geht, unterschiedliche Informationsquellen zu integrieren, wirtschaftliches Optimierungspotenzial zu entdecken oder Erfolge zu messen. In dezentralen Unternehmen sind die individuellen Ziele der beteiligten Parteien oft deutlich schwieriger zu harmonisieren als in straff zentral organisierten Unternehmen (siehe Abbildung 2).

Einzelne Organisationseinheiten stehen im Wettbewerb miteinander und besitzen großes Geschick darin, eigene Interessen zu wahren. Deren weitgehende unternehmerische Selbstständigkeit lockert den in stärker zentral gesteuerten Unternehmen vorherrschenden Konsenszwang. Das führt im Data-Warehouse-Projekt zu hohem Abstimmungsaufwand und erzwingt hohe Flexibilität sowie Kompromissbereitschaft in der Projektabwicklung.

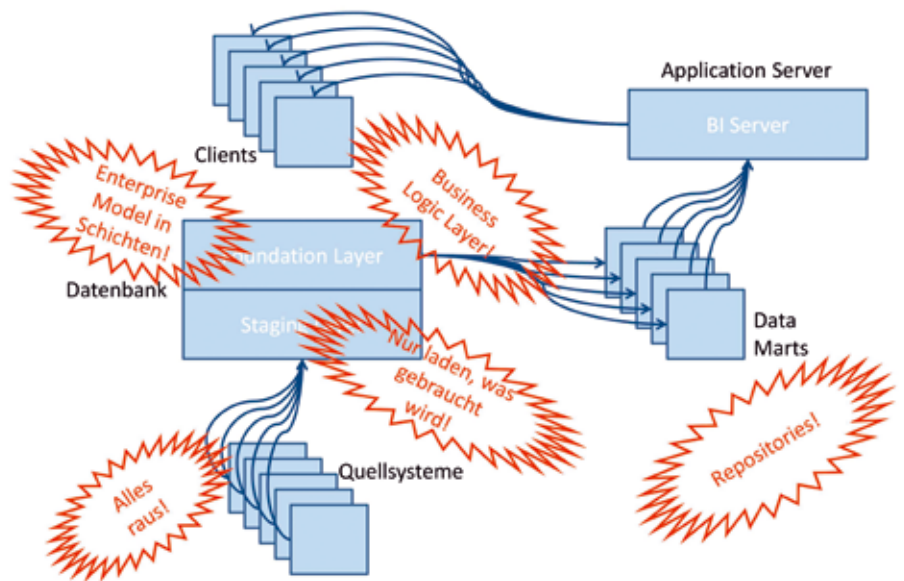


Abbildung 2: Merkmale zur Data-Warehouse-Struktur bei dezentral organisierten Unternehmen

Um ein Data-Warehouse-Projekt trotzdem plan- und steuerbar zu halten, ist einerseits eine iterative Vorgehensweise notwendig, die zwar das Gesamtziel vor Augen hat, aber laufend für den Kunden offensichtlich verwertbare Zwischenergebnisse bietet. Schließlich steht vor einem ersten praktischen Einsatz eine hohe Investition. Andererseits muss die technische Data-Warehouse-Plattform diese Flexibilität und Kompromissbereitschaft auch unterstützen. Das war auch in einem Projekt des Autors einer der wesentlichen Gründe für die Auswahl des Oracle-Stacks als BI-Plattform.

Die Werkzeuge der Oracle Business Intelligence Suite und deren konsequente Wei-

terentwicklung in Richtung der „Fusion“-Architektur sind ausreichend flexibel und leistungsfähig, um im großen Unternehmen eine zentrale Data-Warehouse-Plattform zunächst als Lösung mit relativ schmal definiertem Einsatzspektrum zu implementieren und anschließend schrittweise zu einer globalen „Allround“-Lösung auszubauen. Ein benutzerfreundliches Analyse-Portal bietet dem Benutzer dabei eine stabile Oberfläche, die die umfassenden Arbeiten „unter der Motorhaube“ zuverlässig verbirgt. Neue Erweiterungen können ohne großen Aufwand freigeschaltet werden.

Lutz Rothmann

Lutz.Rothmann@gft.com

Herausgeber:
DOAG Deutsche ORACLE-
Anwendergruppe e.V.
Tempelhofer Weg 64, 12347 Berlin
www.doag.org

Verlag:
DOAG Dienstleistungen GmbH
Fried Saacke, Geschäftsführer
info@doag-dienstleistungen.de

Chefredakteur (VisdP):
Wolfgang Taschner
redaktion@doag.org

Chefin von Dienst (CvD):
Carmen Al-Youssef
office@doag.org

Titel, Gestaltung und Satz:
Claudia Wagner
DOAG Dienstleistungen GmbH

Titelfoto: Fotolia

Anzeigen: CrossMarketteam
Ralf Rutkat, Doris Budwill
www.crossmarketteam.de
Mediadaten und Preise finden Sie unter
www.doag.org/publikationen/

Druck:
adame Advertising and Media GmbH Berlin
www.adame.de