

Die Anzahl von Unternehmen, die Data Warehouse Systeme einsetzen, ist in den letzten Jahren erheblich gestiegen. Inzwischen interessiert sich auch eine Vielzahl von mittelständischen Unternehmen für Data Warehouse Systeme¹. Dies liegt zum einen daran, dass unabhängig von der Unternehmensgröße die Erkenntnis gereift ist, dass das Unternehmen zwar über eine Vielzahl von Daten verfügt, ihnen aber nur eingeschränkte Mittel zur Verfügung stehen, um aus den Daten Informationen für gesicherte Entscheidungen zu gewinnen. Zum zweiten haben Data Warehouse Systeme einen Reifegrad und ein Preisniveau erreicht, das sie auch für mittelständische Unternehmen interessant und erschwinglich macht.



Data Warehouse für den Mittelstand

Methodische Vorgehensweise zur Analyse und Modellierung der Daten

Durch die genannten Entwicklungen erwachsen für Data Warehouse Systeme in mittelständischen Unternehmen breite Einsatzfelder. Bevor jedoch mit der Auswahl und Implementierung eines solchen Systems begonnen wird, muss eine Analyse und fachliche Modellierung von Data Warehouse Daten vorausgehen. Dies ist ein wesentlicher Schritt für eine erfolgreiche Einführung.

Die Analyse der Daten sollte dabei zielorientiert, systematisch und vollständig erfolgen. Zielorientiert bedeutet, zu ermitteln welcher Informationsbedarf besteht, um Unternehmensentscheidungen besser treffen zu können und welche Daten dafür in einem Data Warehouse notwendig sind. Das Ziel eines systematischen Vorgehens ist es, möglichst vollständig die wesentlichen

Daten eines Unternehmens oder Unternehmensbereichs zu erfassen. Dies erfordert Methodiken in der Erhebung, Dokumentation und Analyse, die einfach anzuwenden sind und einen leichten Überblick verschaffen, um eine Kontrolle über die Vollständigkeit zu ermöglichen.

Darauf aufbauend folgt die Phase der fachlichen Modellierung von *Data Marts* oder *Cubes*. Diese dienen zur flexiblen Auswertung wirtschaftsrelevanter Kennzahlen als Basis unternehmerischer Entscheidungen. Die Modellierung wird

1) Business Application Research Center: Computer Reseller News, 24/2007

2) Kimball, Ralph; Ross, Margy: The Data Warehouse Toolkit, Second Edition, New York 2002 (S. 395, 396)



genutzt, um aus fachlicher Sicht festzulegen, wie ein *Data Mart* oder *Cube* als Kern eines Data Warehouses aufgebaut sein sollte². Sie kennzeichnen einen mehrdimensionalen Datenbestand, der lediglich Teile der gesamten Datenmenge eines unternehmensweiten Data Warehouses enthält. Dabei kann ein Cube eine Untermenge eines Data Marts sein. Die entscheidungsrelevanten Auswertungen und Analysen werden in einem Data Warehouse auf Basis von Data Marts oder Cubes (im Weiteren nur noch



Prof. Dr. Siegbert Kern (FH Gelsenkirchen): Die korrekte Analyse und fachliche Modellierung der Data Warehouse Daten ist wesentlicher Schritt für eine erfolgreiche Einführung.

Data Marts genannt) durchgeführt. Das bedeutet, dass Daten oder Beziehungen, die dort nicht vorhanden sind, auch nicht ausgewertet und analysiert werden können. Deshalb ist es wichtig, aus fachlicher Sicht die Struktur von Data Marts zu konzipieren. Diesem Schritt muss eine zielgerichtete und umfassende Analyse der entscheidungsrelevanten Informationen vorausgehen. Dazu sollte eine geeignete Methode genutzt werden, um den Aufbau und Inhalt von Data Marts übersichtlich und eindeutig darstellen zu können.

Die Analyseergebnisse und die aus fachlicher Sicht konzipierten mehrdimensionalen Data Marts bilden die Voraussetzungen für eine erfolgreiche Einführung eines Data Warehouse Systems in Unternehmen. Im Folgenden wird eine am *Institut für Organisations- und Tech-*



Stefanie Detlefs (IOT GmbH): Eine einheitliche Semantik und eine klare Definition der Bezeichnungen, Beschreibungen, Regeln und Bedingungen von Kennzahlen und Merkmalen ist die Grundlage eines gelungenen Data Warehouses.

nikgestaltung entwickelte und erprobte methodische Vorgehensweise zur Analyse und Modellierung von Data Warehouse Daten (DW-Daten) vorgestellt, die die genannten Eigenschaften erfüllt.

Methodische Vorgehensweise

Ziel der Analyse und Modellierung von DW-Daten sollte es sein, ein ganzheitliches Bild über den Informationsbedarf im gesamten Unternehmen oder eines Unternehmensbereichs zu bekommen. In dieser Phase birgt ein ganzheitlicher Ansatz verschiedene Vorteile:

- ▶ Der Informationsbedarf einzelner Bereiche wird deutlich.
- ▶ Die Überlappungen von Informationsbedarfen zwischen den Bereichen werden erkennbar.
- ▶ Unternehmensweit kann ein erster Schritt zu einer gemeinsamen

und einheitlichen Sprachregelung erfolgen.

- ▶ Die Wiederverwendbarkeit von Quelldaten wird deutlich.
- ▶ Die Wirtschaftlichkeit eines Data Warehouses kann besser ermittelt werden, da die Anforderungen und der Nutzen aller Bereiche einfließen können.

Für mittelständische Unternehmen ist die Betrachtung des Gesamtunternehmens immer zu empfehlen, da der Untersuchungsgegenstand einen Umfang hat, der eine Projektdauer von drei Monaten in der Regel nicht übersteigen lässt. Natürlich sieht die Sachlage bei Großunternehmen anders aus. Einerseits geht in Großunternehmen die Initiative erfahrungsgemäß von einzelnen Bereichen aus. Andererseits erreicht die Analyse des Informationsbedarfs für das Gesamtunternehmen einen kritischen Umfang, der zu längeren Projekten mit einem hohen Abstimmungsbedarf führen kann. Dies sollte möglichst vermieden werden, da solche Projekte in den wenigsten Fällen erfolgreich abgeschlossen werden.

Die methodische Vorgehensweise besteht aus vier Schritten, wie dies in Abbildung 1 dargestellt ist.

Schritt 1: Ermittlung des Informationsbedarfs

In der Analysephase muss zunächst der Informationsbedarf gedeckt werden. Unter Information verstehen wir hier zweckorientiertes Wissen³, wobei der Zweck der Information darin liegt, Entscheidungen und Handeln vorzubereiten. Hierzu sind vor allem zwei Methoden von Bedeutung:

Sammeln und Analysieren vorhandener Auswertungen und Berichte

Es kann davon ausgegangen werden, dass in allen Unternehmen vielfältige Berichte zu verschie-

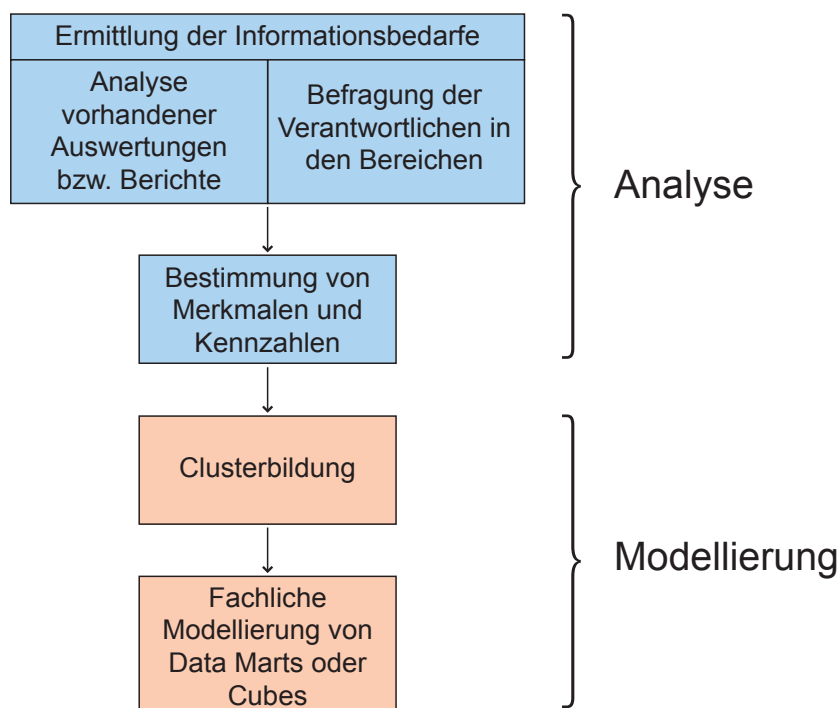


Abb. 1: Vier Schritte zum fachlichen DW-Modell

denen betriebswirtschaftlichen und auch technischen Fragestellungen bestehen. Diese Berichte und Auswertungen sollten systematisch je Organisationseinheit und Bereich gesammelt werden. Gleichzeitig sollten weitere Informationen zu diesen Berichten erfasst werden:

- ▶ Wer nutzt die Berichte – welche Organisationseinheit, welcher Verantwortliche, welcher Bereich?
- ▶ Für welche Entscheidungen liefern die Berichte wertvolle Informationen?
- ▶ Mit welchen IT-Systemen werden die Berichte erstellt?
- ▶ Wie häufig werden die Berichte erstellt (täglich, wöchentlich, monatlich, jährlich)?

Befragung der Verantwortlichen in den Bereichen

Die Ermittlung des Informationsbedarfs sollte sich nicht alleine auf vorhandene Berichte und Auswertungen stützen. Ein solches

Vorgehen enthält die Gefahr, dass keine qualitative Verbesserung des Prozesses zur Entscheidungsfindung erreicht wird. Deshalb sollten zusätzlich Interviews mit der Geschäftsleitung und den potentiellen Anwendern geführt werden. In diesen Gesprächen sollte herausgefunden werden, welche wesentlichen operativen und strategischen Entscheidungen wiederkehrend und ad hoc getroffen werden müssen. Die Entscheider sollten Aussagen darüber machen,

- ▶ welche Informationen sie benötigen und wünschen,
- ▶ wie schnell die Informationen zur Verfügung stehen sollen und
- ▶ welche Flexibilität in der Betrachtungsweise der Informationen notwendig ist,

um ihre Entscheidungen verbessern zu können. Hierbei bietet sich die Möglichkeit an, Interviews mit ein bis drei Personen oder Workshops mit einer Gruppe von Personen durchzuführen⁴. Solch eine Befragung ist ein wichtiger Schritt, um das Spektrum vom aktuellen

3) Wittmann, W.: Unternehmung und unvollkommene Information, Köln-Opladen 1959 (S. 14)

4) Kimball, Ralph, u.a.: The Data Warehouse Lifecycle Toolkit, Expert Methods for Designing, Developing, and Deploying Data Warehouses, New York 1998 (S. 98)

| | Merkmale | |
|----------------------|---------------------------|---------------------------|
| Bezeichnung | Merkmal 1 | Merkmal 2 |
| Quellsystem | SAP | Webshop |
| Auswertung | | |
| Bericht 1 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |
| Bericht 2 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |
| Informationsbedarf 1 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |

Tabelle 1: Schema für eine Merkmalsmatrix

| | Kennzahlen | |
|----------------------|---------------------------|---------------------------|
| Bezeichnung | Kennzahl 1 | Kennzahl 2 |
| Quellsystem | SAP | MS SQL Server |
| Auswertung | | |
| Bericht 1 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |
| Bericht 2 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |
| Informationsbedarf 1 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |

Tabelle 2: Schema für eine Kennzahlenmatrix

Ist-Zustand auf den zukünftigen Soll-Zustand für Auswertungen und Analysen fachlich zu erweitern. Die Einführung und Modellierung eines Data Warehouses darf sich nicht auf rein technische Verbesserungen des Berichtswesens beschränken, sondern sollte auch inhaltlich einen Fortschritt für die Entscheidungsunterstützung bringen.

Die Zusammenführung der Ergebnisse aus beiden Methodiken macht deutlich, welche Bedeutung die erhobenen Berichte im Unternehmen haben, mit welchem Aufwand sie erstellt werden und in welchen Bereichen über das Bestehende hinaus Informationsbedarf besteht.

**Schritt 2:
Bestimmung von
Merkmalen
und Kennzahlen**

Im nächsten Schritt müssen die gesammelten Berichte und der In-

formationsbedarf analysiert werden. Dazu sollten die Berichte nach Kennzahlen und Merkmalen untersucht werden. Unter einer Kennzahl wird eine mengen- oder wertmäßige Information verstanden (z.B. Umsatz). Ein Merkmal ist eine beschreibende oder kategorisierende Information (z.B. Kundennummer).

Zur Dokumentation und Veranschaulichung der Ergebnisse sollte jeweils eine Matrix für Merkmale, deren Schema in Tabelle 1 deutlich wird sowie eine Matrix für Kennzahlen, deren Schema in Tabelle 2 dargestellt ist, aufgebaut werden. Darin sollten die Merkmale mit ihren Bezeichnungen und die analysierten Berichte und den Informationsbedarf aufgeführt und, wie in Tabelle 1 dargestellt, erfasst werden. Das *Quellsystem* gibt an, aus welchen IT-Systemen das Merkmal stammt. In der Zelle *vorhanden/nicht vorhanden* kann festgehalten werden, welches Merkmal in welcher Auswertung vorkommt. Für die Kennzahlen erfolgt die Dokumentation der Analyse in gleicher Form, wie in Tabelle 2 dargestellt.

Schon während der Aufstellung der Merkmals- und Kennzahlenmatrizen wird erkennbar, dass innerhalb des Unternehmens für die gleiche Kennzahl oder das gleiche Merkmal häufig unterschiedliche Begriffe und Abkürzungen verwendet werden. Es sollte bereits in dieser Phase versucht werden, eine einheitliche Semantik und Definition über die Bezeichnungen, Beschreibungen, Regeln und Bedingungen der Kennzahlen und Merkmale im Unternehmen herzustellen. Dazu sollte ein konsensfähiger Vorschlag

erarbeitet werden, der mit allen Beteiligten abgestimmt wird. Die Ergebnisse sollten in einem unternehmensweiten Data Warehouse Glossar dokumentiert werden.

Bei der Bestimmung von Kennzahlen und Merkmalen kann eine weitere Detaillierung durchgeführt werden, welche die spätere Modellierung wesentlich erleichtert. Merkmale können in Bezug auf ihre gegenseitige Abhängigkeit hin differenziert werden. Ein Merkmal, das von einem anderen Merkmal abhängig ist, wird als

Das IOT begleitet Sie durch den Prozess der systematischen Analyse und Realisierung logistischer und organisatorischer IT-Implementierung.

Zur Steuerung, Optimierung und Auswertung von Produktionsprozessen kommen Data-Warehouse-Anwendungen, SAP-Module und erprobte SAP-Eigenentwicklungen zum Einsatz.

10 Jahre Erfahrung und Entwicklung schaffen Sicherheit und Vertrauen bei unseren Kunden der mittelständischen Industrie.

Wir wissen worauf es Ihnen ankommt!



IOT - Institut für Organisations- und Technikgestaltung GmbH
 Buschgrundstraße 25
 45894 Gelsenkirchen
 Tel. +49 (0)209 359 763 - 0
 info@iot-online.de
 www.iot-online.de

Attribut bezeichnet. Beispielsweise ist das Attribut *Kundenname* vom Merkmal *Kundennummer* abhängig. Entsprechend können die Merkmale in unabhängige Merkmale und Attribute unterschieden werden.

Kennzahlen können in Basis-kennzahlen und berechnete Kennzahlen unterschieden werden. Bei Basiskennzahlen handelt es sich um Kennzahlen, die auf Basis von Daten der Quellsysteme gebildet werden, beispielsweise der fakturierte Umsatz Netto. Eine berechnete Kennzahl stellt mehrere Kennzahlen im Rahmen einer Formel zueinander in Bezug, zum Beispiel: Warenrohertrag als Summe aus fakturiertem Umsatz Netto abzüglich der fakturierten Menge multipliziert mit dem Einkaufspreis.

Wird eine berechnete Kennzahl in die Matrix aufgenommen, ist zu prüfen, ob die benötigten Formelbestandteile bereits in der Matrix enthalten sind. Ist dies nicht der Fall, sind die benötigten Kennzahlen ebenfalls hinzuzufügen. Eine Zugehörigkeit wird zu den Berichten/dem Informationsbedarf angegeben, in denen die berechnete Kennzahl verwendet wird. Bei berechneten Kennzahlen ist es außerdem sinnvoll, noch eine weitere Unterscheidung vorzunehmen, die aus der Kennzahlenmatrix ableitbar ist. Zum einen unterscheidet man Kennzahlen, die in vielen Auswertungen vorkommen (berechnete Kennzahl – allgemein) und zum anderen Kennzahlen, die nur in einer oder sehr wenigen Auswertungen vorkommen (berechnete Kennzahl – spezifisch). Eine solche Differenzierung bietet einen wesentlichen Anhaltspunkt dafür, ob die Kennzahl bereits im Data Mart vorberechnet oder erst bei der Erstellung der Auswertung zur Laufzeit berechnet werden soll. *Berechnete Kennzahlen – spezifisch* sollten eher zur Laufzeit berechnet werden, da sie nur selten verwendet werden. Auf diese Weise kann eine unnötige Steigerung des Datenvolumens im Data Warehouse vermieden werden.

| | Merkmale | |
|----------------------|---------------------------|---------------------------|
| Bezeichnung | Merkmal 1 | Merkmal 2 |
| Quellsystem | SAP | Webshop |
| Merkmalstyp | Merkmal | Attribut |
| Auswertung | | |
| Bericht 1 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |
| Bericht 2 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |
| Informationsbedarf 1 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |

Tabelle 3: Erweitertes Schema für eine Merkmalsmatrix

| | Kennzahlen | |
|----------------------|---------------------------|---------------------------|
| Bezeichnung | Kennzahl 1 | Kennzahl 2 |
| Quellsystem | SAP | MS SQL Server |
| Kennzahlentyp | Basiskennzahl | berechnete Kennzahl |
| Auswertung | | |
| Bericht 1 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |
| Bericht 2 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |
| Informationsbedarf 1 | vorhanden/nicht vorhanden | vorhanden/nicht vorhanden |

Tabelle 4: Erweitertes Schema für eine Kennzahlenmatrix

Die erstellten Merkmals- und Kennzahlenmatrizen sollten um Merkmals- beziehungsweise Kennzahlentypen ergänzt werden (Tab. 3 und 4).

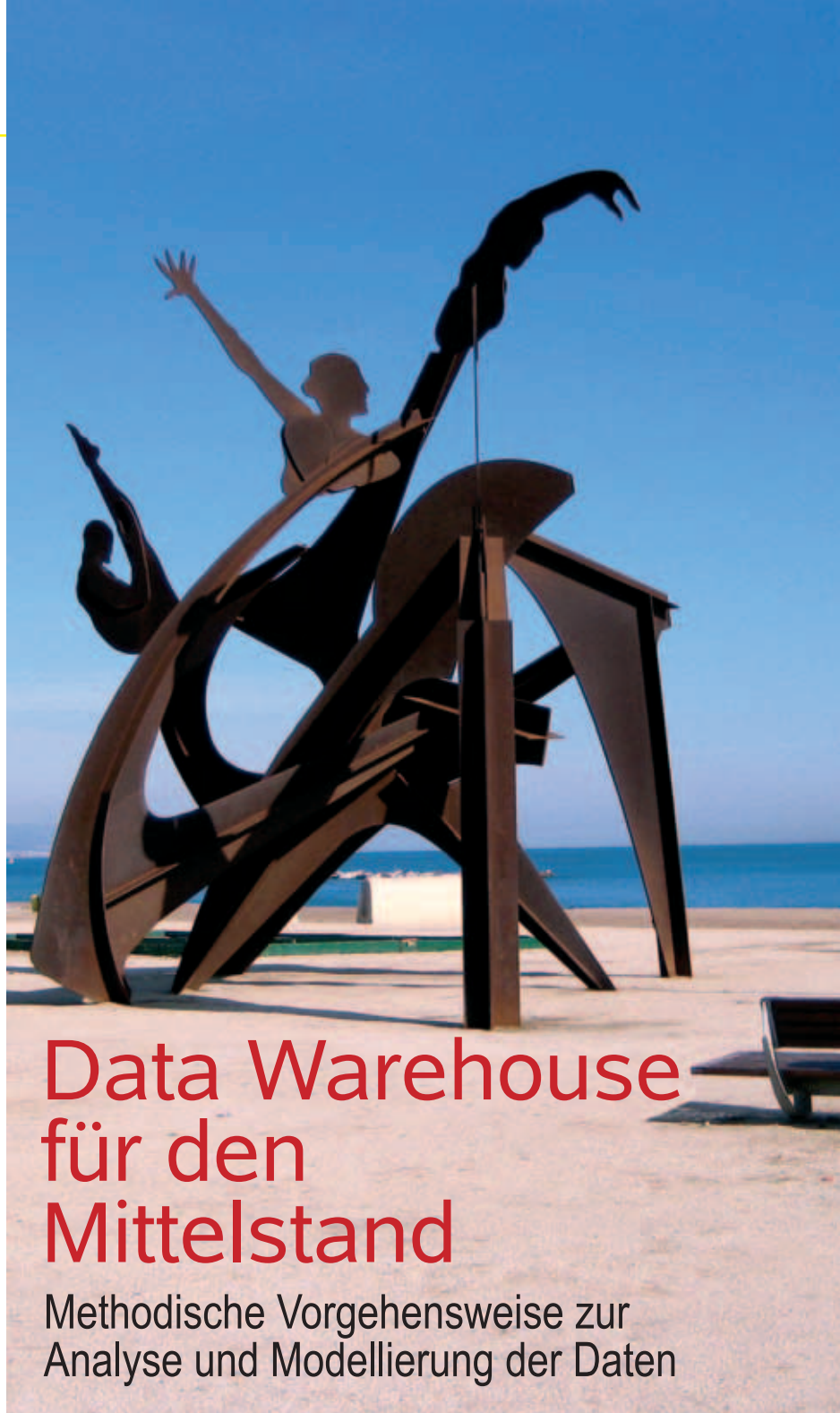
Mit Schritt 2 ist die Phase der fachlichen Analyse abgeschlossen. Die vorgestellte Methodik der Analyse und Darstellung enthält wesentliche Vorteile:

- ▶ Doppelte Kennzahlen und Merkmale können leicht entdeckt werden.
- ▶ Semantische Mehrdeutigkeiten bei Merkmalen und Kennzahlen werden aufgezeigt.
- ▶ Merkmale und Kennzahlen werden durch Typenbildung systematisiert.
- ▶ Es wird deutlich, in welchen Auswertungen und wie häufig ein Merkmal oder eine Kennzahl benötigt wird.
- ▶ Doppelte Berichte können leicht identifiziert werden.

Dieser Beitrag wird in Heft 3-2008 fortgesetzt werden, mit den Schritten 3 – Clusterbildung – und 4 – Fachliche Modellierung von Data Marts und Cubes – sowie Ausführungen zur Modellierung. Interessenten, die den „EDV-Leiter“ nicht regelmäßig beziehen, können sich für eine Schutzgebühr von € 10,- ein Exemplar dieses Heftes reservieren lassen. Senden Sie dazu einfach eine E-Mail an service@edvleiter.com. Sie erhalten Heft 3-2008 dann unmittelbar zum Erscheinungstermin.

Prof. Dr. Siegbert Kern,
FH Gelsenkirchen
Dipl.-Inform. (FH) Stefanie Detlefs,
IOT GmbH

Data Warehouse Systeme sind im Mittelstand angekommen. Unabhängig von der Unternehmensgröße ist die Erkenntnis gereift, die im Unternehmen vorliegenden Daten gezielt für gesicherte Entscheidungen einzusetzen. Aufbauend auf Teil 1 (EDV-Leiter 2-2008) zeigen Prof. Dr. Siegbert Kern und Stefanie Detlefs die Clusterbildung und die fachliche Modellierung von Data Marts und Cubes – und verbinden diese mit verschiedenen Ansätzen zur Datenmodellierung¹.



Data Warehouse für den Mittelstand

Methodische Vorgehensweise zur
Analyse und Modellierung der Daten

Schritt 3: Clusterbildung

Mit der Clusterbildung beginnt – nach der Ermittlung des Informationsbedarfs und nach der Bestimmung von Merkmalen und Kennzahlen – die Phase der Modellierung. Ein Cluster ist ein abgegrenzter Bereich, dessen Bedarf an Auswertungen und Analysen vollständig durch die zugeordneten Berichte

und Informationen abgedeckt werden kann. Ein solcher Cluster kann nach unterschiedlichen Kriterien gebildet werden:

► Ähnlichkeiten von Berichten (z. B. Zusammenfassung von Berichten und Informationsbedarf, die eine hohe Übereinstimmung in den verwendeten Merkmalen und Kennzahlen besitzen)

► Funktionale Kriterien (z. B. Zusammenfassung von Berichten und Informationsbedarf nach Unternehmensfunktionen wie Vertrieb, Einkauf, Fertigung, Personal, u. a.)

► Prozessorientierte Kriterien (z. B. Zusammenfassung von Berichten und Informationsbedarf nach wesentlichen Prozessen in den Unternehmen, wie Auftragsabwicklung, Auslieferung an den Kunden, u. a.).

Für die Clusterbildung wichtig sind die in der Analysephase erhobenen Informationen und die erstell-

¹ Leser, die den ersten Teil dieses Beitrags versäumt haben, können Ausgabe 2-2008 direkt beim Verlag bestellen – service@edvleiter.com – oder den Beitrag im CASE-Archiv des EDV-Leiter online abrufen: www.edvleiter.com/html/intro_case-archiv.html.



Prof. Dr. Siegbert Kern: Wichtig ist, die Vollständigkeit der Datenbasis sicherzustellen. Auf diese Weise lassen sich zukünftige Änderungen vermeiden.



Stefanie Detlefs (IOT GmbH): Entdecken Sie beim Aufbau des Data Warehouses Fehler im Design, ist es durchaus sinnvoll, noch Änderungen in der Modellierung durchzusetzen.

ten Matrizen. Diese geben Auskunft über die Zuordnung von Kennzahlen, Merkmalen zu Berichten und den jeweiligen Informationsbedarf. Die in der Erhebung gewonnenen Erkenntnisse, welche Berichte und welcher Informationsbedarf von welchen Organisationseinheiten genutzt werden, geben Hinweise zur funktions- und prozessorientierten Clusterung.

Einem definierten Cluster müssen nun die Kennzahlen und Merkmale zugeordnet werden. Je nach Bildung der Cluster ist es möglich,

dass Kennzahlen und Merkmale in mehreren Clustern benötigt werden. Des Weiteren wird überprüft werden, inwieweit die zugeordneten Kennzahlen und Merkmale im Hinblick auf den geforderten Informationsbedarf innerhalb des Clusters vollständig sind. Dies muss im Rahmen eines Reviews mit den Verantwortlichen aus der Geschäftsleitung und den Fachbereichen erfolgen.

Anhand der Cluster, den aufgelisteten Auswertungen, zusätzlichem Informationsbedarf und den defi-

nierten Kennzahlen und Merkmalen, kann eine gezielte Abstimmung erfolgen. Sofern weitere Kennzahlen und Merkmale erforderlich sind, können diese einfach hinzugefügt werden. Damit kann in einem hohen Maß die Vollständigkeit der Datenbasis erreicht werden. Diese ist wichtig, um zukünftige Änderungen an der Datenbasis weitgehend zu vermeiden. Bei Veränderungen von Geschäftsmodellen und -prozessen können Anpassungen an der Datenbasis erforderlich werden.

Schritt 4: Fachliche Modellierung von Data Marts oder Cubes

Basierend auf den ermittelten Merkmalen und Kennzahlen wird ein fachliches mehrdimensionales Datenmodell für einen Data Mart modelliert. Die in Schritt 3 ermittelten Cluster bilden dabei die Vorlage zum Aufbau eines oder mehrerer Data Marts. Hierbei entspricht in der Regel ein Cluster einem Data Mart. Es ist sinnvoll für jeden Data Mart ein eigenes Datenmodell zu entwickeln, da ein solches Datenmodell dabei hilft, die gesammelten Daten in eine übersichtliche Struktur zu bringen und bereits im Vorfeld Schwächen im Design zu entdecken. Des Weiteren ist das fachliche Datenmodell die Grundlage für die technische Umsetzung in einem Data Warehouse System.

Teilweise werden Fehler im Design auch erst beim Aufbau, also nach der eigentlichen Modellierungsphase, des Data Marts sichtbar. Auch zu diesem Zeitpunkt ist es durchaus sinnvoll, noch Änderungen in der Modellierung umzusetzen, da die endgültige Struktur meist über Jahre Bestand haben wird. Ein weiterer Vorteil eines Datenmodells liegt darin, dass es einen schnellen Überblick über die Zusammenhänge von Merkmalen, Kennzahlen und Dimensionen bietet und sich somit auch gut zur Dokumentation eignet.

Dimensionen sind eine fachliche Zusammenfassung von Merkmalen. Sie sollten derart gebildet werden, dass sie jeweils ein Themengebiet beschreiben. So ist es beispielsweise

sinnvoll, eine Dimension „Artikel“ zu bilden, in der Merkmale wie Artikelnummer und Artikeltyp zusammengefasst werden. Die einzelnen Dimensionen können bei der späteren Berichterstellung sinnvoll miteinander kombiniert werden. Prinzipiell sind auch Dimensionen mit nur einem Merkmal möglich. Da die Anzahl der Dimensionen allerdings in den meisten Systemen technisch begrenzt ist, sollte dieser Fall möglichst vermieden werden. In manchen Data Warehouse Systemen haben Dimensionen auch einen technischen Hintergrund, um bei Abfragen eine höhere Performance zu erzielen.

Das erstellte Datenmodell erleichtert erheblich sowohl die Auswahl eines Data Warehouse Systems als auch die Umsetzung in ein solches System. Mit den gewonnenen Erkenntnissen sind die wesentlichen fachlichen Anforderungen definiert, die mit angebotenen Data Warehouse Systemen abgeglichen werden können. Da bereits alle Zusammenhänge der einzelnen Elemente bekannt sind, erhält man eine höhere Planungssicherheit bei der Umsetzung. Durch ein bereits bestehendes mehrdimensionales Datenmodell lässt sich der Umsetzungsaufwand reduzieren. Durch eine sorgfältige Umsetzung können mögliche Designfehler im Modell vermieden werden.

Prinzipiell lässt sich zusammenfassen, dass vor der Umsetzung in ein Data Warehouse System auf jeden Fall ein Datenmodell entwickelt werden sollte, da eine strukturierte und vor allem vollständige Daten-

basis die Grundvoraussetzung für die gewünschten Auswertungen und Analysen darstellt.

Durch die Komplexität von mehrdimensionalen Datenmodellen ist es sinnvoll, grafische Modellierungsmethoden zur übersichtlichen Darstellung anzuwenden. Während es zur Erstellung von operationalen Systemen einige standardisierte Modellierungsmöglichkeiten gibt – *Entity Relationship Modell (ERM)* und in den letzten Jahren verstärkt auch *Unified Modelling Language (UML)* –, hat sich im Umfeld der Data Warehouse Modellierung noch kein Modellierungsansatz durchsetzen können. Nachfolgend geben wir daher einen kurzen Überblick über die Möglichkeiten, die zur grafischen Darstellung eines mehrdimensionalen Datenmodells für ein Data Warehouse dienen können².

Ansätze auf Basis des Entity Relationship Modells

Das ERM ist eine der am weitesten verbreiteten Methoden zur Erstellung von Datenmodellen. Im Laufe der Zeit wurde es um einige Elemente erweitert, die es ermöglichen auch komplexere, multidimensionale Datenmodelle zu erstellen. Dabei werden mehrere Ansätze verfolgt, die alle auf dem klassischen ERM basieren.

► Multidimensionales Entity Relationship Modell (ME/RM)

Erweiterung der ER-Notation um die drei Elemente „Faktenrelation“, „Dimensionsebene“ und „hierarchische Beziehung“.

► Modifizierte Objekttypenmethode (OTM)

OTM wurde entwickelt, um eine formale Begriffsbildung zu erreichen. Für die Modellierung von Data Warehouse Systemen wurden die Sprachkonstrukte „Objekttyp“, „Konnexion“ (Beziehungstyp für die Verbindung von Objekten) und „Aggregation“ eingeführt. Zusätzlich wurde ein neuer komplexer Objekttyp eingeführt, der hierarchische Elemente zusammenfasst.

2) Weitere Informationen und Beispiele zu den einzelnen Ansätzen finden sich in den folgenden Quellen: Muksch, H.; Behme, W.: Das Data Warehouse-Konzept. Architektur – Datenmodelle – Anwendungen, 4. vollständig überarbeitete und erweiterte Auflage, Wiesbaden 2000, S. 191ff, und Egger, N.; Fiechter, J.-M.; Rohlf, J.: SAP BW Datenmodellierung, Bonn 2004, S. 109ff.
3) Egger, N.; Fiechter, J.-M.; Rohlf, J.: SAP BW Datenmodellierung, Bonn, 2004, S. 114.
4) Egger, N.; Fiechter, J.-M.; Rohlf, J.: SAP BW Datenmodellierung, Bonn, 2004, S. 115.
5) Zu beziehen sind die Shapes unter www.hahneonline.de/shapes.html. Eine Beschreibung der einzelnen Komponenten findet sich in folgender Quelle: Hahne M.: Grafische Repräsentation mehrdimensionaler Datenmodelle des SAP Business Information Warehouse; cundus AG – www.cundus.de/downloads/MKW2004_Hahne_paper.pdf. Einen guten Einstieg in die Modellierung mit ADAPT bietet: Bulos, D. Forsman, S.: Getting Started with ADAPT™, OLAP Database Design, Symmetry Corporation – www.symcorp.com/downloads/ADAPT_white_paper.pdf.

► **Datenmodellierung Data Warehouse – ein Lösungsvorschlag des Guide/Share Europe**

Bei diesem Modellierungsansatz steht die Modellierung von Dimensionshierarchien und Faktenentitäten im Vordergrund.

Ansätze auf Basis des objektorientierten Modells

Der objektorientierte Ansatz der Datenmodellierung ist wesentlich neuer als die ER-Modellierung. Er erfreut sich großer Beliebtheit und hat sich relativ schnell verbreitet. Neben den reinen Daten, wie auch bei der ER-Modellierung, werden hier zusätzlich die Methoden betrachtet, die zum Datenzugriff oder zur Datenbearbeitung dienen.

Im Zusammenhang mit der Erstellung von Datenmodellen im Data Warehouse Umfeld sind die folgenden Ansätze von Bedeutung:

► **Object Oriented Analysis (OOA)**

„Die Klasse „Würfeld“ ist das zentrale Element des Ansatzes, deren Instanzen mit den für sie relevanten Dimensionen in Verbindung stehen. Ausprägungen der Klasse „Würfeld“ stehen mit Ausprägungen der Klasse „Kenngröße“ über Objekte der Klasse „Atom-Datum“ in Verbindung...“³.

► **Object Modeling Technique (OMT)**

Bei diesem Ansatz wird zwischen Makro- und Mikrosichten unterschieden. Die Makrosicht wird dabei in Daten-, Funktions-, Organisations- und Geschäftsprozessansichten unterteilt. Die Mikrosichten sind wiederum ein Unterpunkt der Datensicht.

► **Unified Modeling Language (UML)**

UML hat sich als Standard zur Modellierung von objektorientierten Systemen durchgesetzt. „Entworfen wird ein Klassenmodell für betriebswirtschaftliche Kennzahlen und Dimensi-

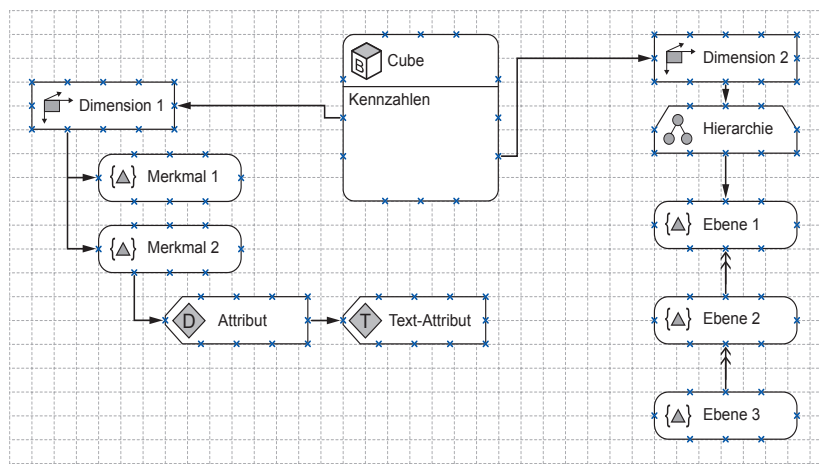


Abb. 1: Elemente der ADAPT-Modellierung

onen jeweils mit den entsprechenden Funktionen zur Verwaltung der später zu erzeugenden Objekte.“⁴

Application Design for Analytical Processing Technologies

Obwohl die Modellierungstechniken ERM und UML weit verbreitet sind und es einige Ansätze gibt, um Data Warehouse Modelle zu erstellen, gibt es keinen Ansatz, der die Modellierung mit hundertprozentiger Zufriedenheit unterstützt. Aus diesem Grund wurde die Modellierungssprache *Application Design for Analytical Processing Technologies*

(ADAPT) entwickelt. In Abbildung 1 sind die wesentlichen Modellierungselemente von ADAPT dargestellt. Auch wenn diese Modellierungssprache nicht die ultimative Lösung darstellt, eignet sich ADAPT im Umfeld der mehrdimensionalen Datenmodelle dennoch besonders gut für die OLAP-Modellierung, also für Datenmodelle, in denen die OLAP-Funktionalitäten abgebildet werden sollen. Darüber hinaus handelt es sich um eine leicht verständliche Modellierungssprache, die durch ihre Einfachheit eine hohe Akzeptanz bei den Mitarbeitern der Fachabteilungen zur Folge hat. Ein besonderer Vorteil von

ADAPT ist, dass für eine rechnergestützte Modellierung die Elemente in Form von Schablonendateien (Shapes) für *Microsoft Visio* zur Verfügung stehen.⁵

Abbildung 2 (Seite 36) zeigt beispielhaft einen Ausschnitt des im Rahmen eines Praxisprojektes konzipierten Modells eines Data Marts zur Auswertung und Analyse von Vertriebsdaten. Hier sind alle Dimensionen, Merkmale und Kennzahlen abgebildet, die der Data Mart „Vertrieb“ beinhalten wird.

Das IOT begleitet Sie durch den Prozess der systematischen Analyse und Realisierung logistischer und organisatorischer IT-Implementierung.

Zur Steuerung, Optimierung und Auswertung von Produktionsprozessen kommen Data-Warehouse-Anwendungen, SAP-Module und erprobte SAP-Eigenentwicklungen zum Einsatz.

10 Jahre Erfahrung und Entwicklung schaffen Sicherheit und Vertrauen bei unseren Kunden der mittelständischen Industrie.

Wir wissen worauf es Ihnen ankommt!



IOT - Institut für Organisations- und Technikgestaltung GmbH
 Buschgrundstraße 25
 45894 Gelsenkirchen
 Tel. +49 (0)209 359 763 - 0
 info@iot-online.de
 www.iot-online.de

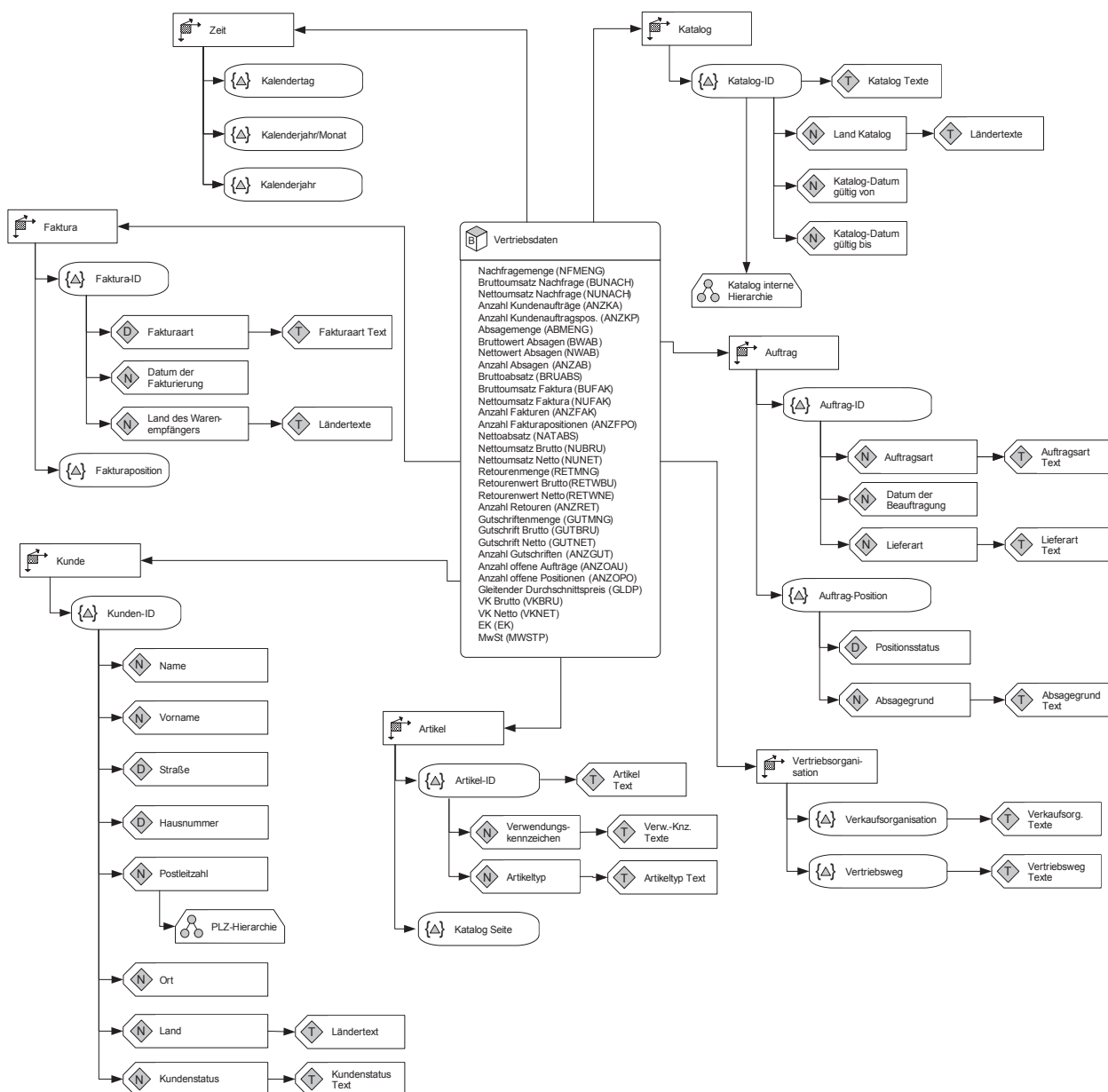


Abb. 2: Beispielhaft: Ein Datenmodell eines Data Marts „Vertrieb“

Die verwendeten Bezeichnungen sind bereits unternehmensweit harmonisiert und in einem Glossar näher beschrieben und definiert.

Zusammenfassung

Die vorgestellte methodische Vorgehensweise zur Analyse und Modellierung von Data Warehouse Daten wurde im Rahmen wissenschaftlicher Arbeiten entwickelt und in verschiedenen Praxisprojekten erprobt. Mit ihrer Anwendung wird

eine zielorientierte, systematische und vollständige Analyse erreicht. Zwischen Analyse und Modellierung existiert kein Bruch, sondern bereits in der Analyse werden die Grundlagen für eine spätere fachliche Modellierung geschaffen. Die Vorgehensweise ist unabhängig von spezifischen Data Warehouse Systemen und kann in jeder Art von Unternehmen eingesetzt werden. Die angewandten Methodiken sind bewusst einfach und übersichtlich

gehalten. Damit werden sie leicht erlernbar und bieten sich insbesondere für mittelständische Unternehmen an, die nicht über die personellen Ressourcen von Großunternehmen verfügen, um den Informationsbedarf eines Unternehmens zu analysieren und zu modellieren.

Prof. Dr. Siegbert Kern,
FH Gelsenkirchen
Dipl.-Inform. (FH) Stefanie Detlefs,
IOT GmbH