

Zweitveröffentlichung



Sinz, Elmar J.

Data Warehouse

Datum der Zweitveröffentlichung: 18.09.2024

Verlagsversion (Version of Record), Beitrag in Sammelwerk

Persistenter Identifikator: urn:nbn:de:bvb:473-irb-981616

Erstveröffentlichung

Sinz, Elmar J. (2006): „Data Warehouse“. In: Wirtschafts-Lexikon : das Wissen der Betriebswirtschaftslehre, 3. Controlling - Finanzdienstleister, Stuttgart: Schäffer-Poeschel, S. 1169–1176.

Rechtehinweis

Dieses Werk ist durch das Urheberrecht und/oder die Angabe einer Lizenz geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die für Sie geltende Gesetzgebung zum Urheberrecht und/oder durch die Lizenz erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

Für dieses Dokument gilt eine Creative-Commons-Lizenz.



Die Lizenzinformationen sind online verfügbar:

<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

D

Data Warehouse

Elmar J. Sinz

I. Begriffsbestimmung und Einordnung;
 II. Multidimensionales Datenmodell;
 III. Architektur; IV. Entwicklung und Betrieb.

I. Begriffsbestimmung und Einordnung

Mit dem Begriff Data Warehouse (DWH) wird eine speziell auf die Entscheidungsunterstützung im Management eines betrieblichen Systems ausgerichtete Datenbasis bezeichnet. Ein DWH ist im Allgemeinen unabhängig von den *Datenbanken* der operativen Anwendungssysteme (DB), welche die Abwicklung der laufenden Geschäftsprozesse unterstützen und als ERP-Systeme (Enterprise Resource Planning) bezeichnet werden.

Die Funktion eines Data Warehouse lässt sich anhand einer *regelkreisorientierten Betrachtung* eines betrieblichen Systems (z.B. Unternehmen, Unternehmensbereich, Unternehmensverbund) bestimmen (siehe Abb. 1). Danach besteht ein betriebliches System aus zwei Teilsystemen, einem Lenkungssystem und einem Leistungssystem (vgl. *Ferstl/Sinz* 2001, S. 4 f.). Das Lenkungssystem (Regler), auch als Führungssystem (vgl. *Küpper* 1997, S. 15) bezeichnet, dient der Planung, Steuerung und Kontrolle der betrieblichen Leistungserstellung. Das Leistungssystem (Regelstrecke) dient der Durchführung der Leistungserstellung.

Innerhalb des Lenkungssystems lassen sich weiter zwei Ebenen von Reglern unterscheiden:

- Aufgabe der GP-Lenkung (Regler) ist die unmittelbare Planung, Steuerung und Kontrolle des Leistungssystems (Regelstrecke) im Rahmen der laufenden Geschäftsprozesse. GP-Lenkung und Leistungssystem gehören zur Geschäftsprozessebene eines betrieblichen Systems.
- Aufgabe des Management (Regler) ist die Lenkung und Gestaltung des gesamten betrieblichen Systems. Hierzu gehören insbesondere die Lenkung der Geschäftsprozesse sowie die Gestaltung der Gesamtheit der Geschäftsprozesse (Regelstrecke) über Lenkungsziele bzw. Gestaltungsziele.

Auf beiden Regelkreisebenen erfassen die Regler die Zustände und Strukturen der jeweiligen Regelstrecke nicht direkt, sondern indirekt über eine Hilfsregelstrecke in Form eines Modells der jeweiligen Regelstrecke. Diese Hilfsregelstrecke wird auf der Geschäftsprozessebene durch die DB gebildet. Auf der Managementebene stellt das DWH die Hilfsregelstrecke dar. Das DWH enthält extrahierte und aufbereitete Daten aus den operativen Datenbanken sowie aus externen Datenquellen.

Die Unterschiede zwischen DWH und DB ergeben sich aus den unterschiedlichen Aufgabenschwerpunkten von Management und GP-Lenkung. Eine der ersten Charakterisierungen der Merkmale eines DWH stammt von Inmon: »A data

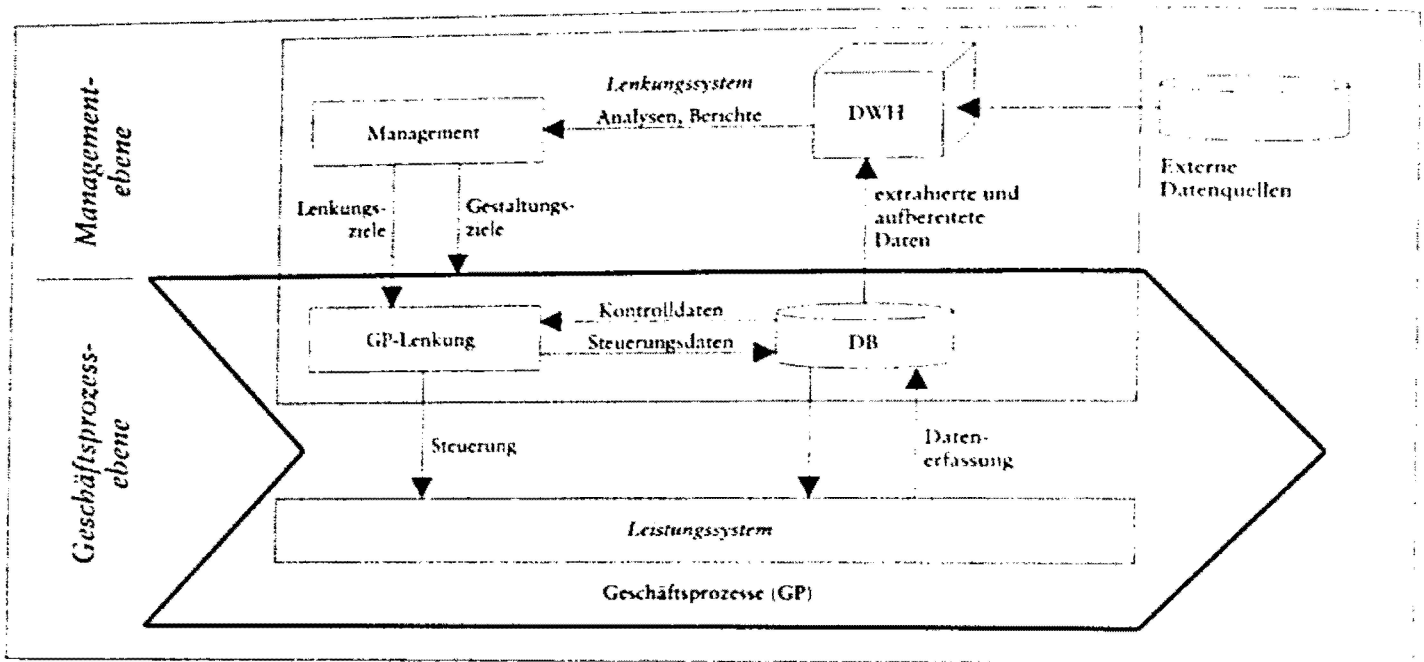


Abb. 1: Regelkreisorientierte Einordnung des Data Warehouse

warehouse is a subject oriented, integrated, non-volatile, and time variant collection of data in support of management's decisions« (Inmon 1996, S. 33). Dabei bedeuten (vgl. z.B. auch Muksch/Behme 2000, S. 9 ff.):

- *Subject oriented*: Aufbau und Inhalt eines DWH orientieren sich an den für die Unternehmensführung relevanten Entscheidungsfeldern. Im Gegensatz dazu unterstützen die DB die Durchführung der einzelnen betrieblichen Transaktionen im Rahmen von Geschäftsprozessen.
- *Integrated*: Ein DWH integriert und konsolidiert im Allgemeinen Daten aus mehreren DB und externen Datenquellen.
- *Non-volatile*: Daten, die einmal in das DWH eingestellt wurden, werden nicht mehr verändert oder gelöscht. Vielmehr werden die Inhalte des DWH zu festgelegten Aktualisierungszeitpunkten sukzessive erweitert. Im Gegensatz dazu werden Daten in den DB während der Durchführung einer betrieblichen

Transaktion laufend verändert und nach Abschluss der Transaktion zum Teil wieder gelöscht.

- *Time variant*: Um Entwicklungen im Zeitablauf analysieren zu können, enthält ein DWH historisierte Daten. DB hingegen bilden im Wesentlichen die Zustände der betrieblichen Transaktionen zum aktuellen Zeitpunkt ab.

Ein DWH stellt einen wichtigen Baustein in einem Führungsinformationssystem (Executive Information System) dar, welches wiederum Teil eines umfassenden Management-Support-Systems sein kann (vgl. Chamoni/Gluchowski 1999, S. 8 f.).

Die Nutzung eines DWH folgt dem Konzept des *Online Analytical Processing* (OLAP). Charakteristisch dafür ist eine interaktive Datenanalyse, bei der der Entscheider in hierarchischen Klassifikationsstrukturen navigiert. Die Berichte enthalten überwiegend quantitative, verdichtete Kennzahlen in multidimensionaler Präsentationsform. Im Gegensatz dazu steht das Konzept des *Online Transaction Processing* (OLTP), die Durchführung be-

trieblicher Transaktionen mithilfe operativer Anwendungssysteme.

II. Multidimensionales Datenmodell

Bei operativen Anwendungssystemen erfolgt die Darstellung der Daten meist in relationaler Form. Die zugehörige Metapher ist die einer zweidimensionalen Tabelle. Die Zeilen der Tabelle stellen Entitäten dar (z.B. Kunden), die Spalten enthalten die Werte dieser Entitäten bezüglich ihrer einzelnen Attribute (z.B. Name).

Die Beschreibung eines DWH erfolgt in multidimensionaler Form (siehe Abb. 2). Die zugehörige Metapher ist die eines mehrdimensionalen Kubus (*Hypercube*) – aus Gründen der Veranschaulichung im Folgenden als dreidimensionaler Würfel dargestellt –, dessen Elemente Werte einer bestimmten entscheidungsrelevanten Kenngröße (auch Maßzahl, Variable, Fakt oder Metrik genannt; z.B. Umsatz) darstellen. Die Werte der Kenngröße sind entlang der Dimensionen des Hypercube klassifiziert (z.B. Produkt, Geographie, Zeit). Jeder Wert der Kenngröße ist jeder Dimension eindeutig zugeordnet (z.B. Umsatz: EUR 63,92, Produkt: Softwarepaket A; Filiale: München-Süd; Tag: 2000-06-13). Dimensionen können hierarchisch aggregiert sein. Zum Beispiel stellt Tag-Monat-Jahr eine zulässige Dimensionshierarchie dar. Tag-Kalenderwoche-Monat ist hingegen nicht zulässig, da eine Kalenderwoche in zwei Monate fallen kann und damit eine hierarchische Aggregation nicht möglich ist.

Im Allgemeinen sind mehrere Kenngrößen entscheidungsrelevant. Besitzen diese Kenngrößen identische Dimensionen, so können die zugehörigen

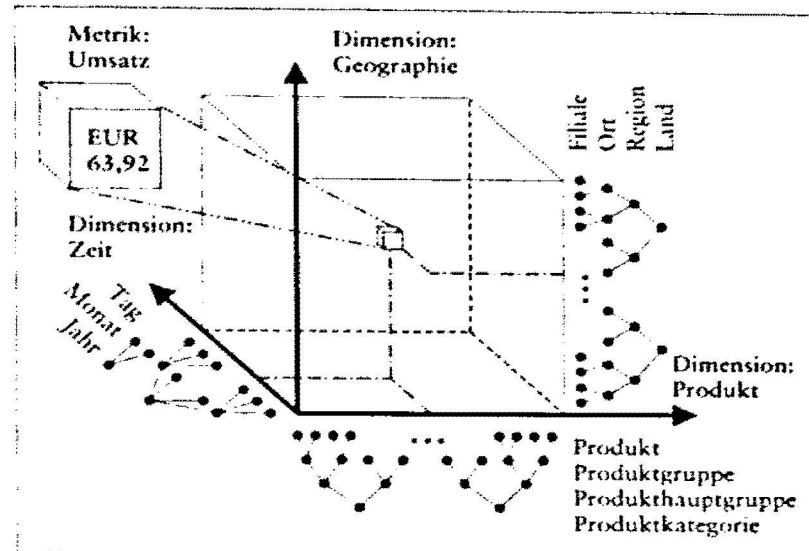


Abb. 2: Multidimensionale Datenstruktur eines Data Warehouse (Beispiel nach Bauer/Günzel 2001, S. 101 ff.)

Werte gemeinsam in einem Hypercube verwaltet werden. Unterscheiden sich die Dimensionen, so müssen mehrere Hypercubes aufgebaut werden.

Die in einem Hypercube befindlichen Werte einer Kenngröße werden als quantitative Daten, die Ausprägungen der einzelnen Dimensionen als qualitative Daten bezeichnet. Eine Anfrage an einen Hypercube bezieht sich auf quantitative Daten anhand vorgegebener qualitativer Daten. Hierzu stehen u.a. folgende Operationen zur Verfügung (vgl. z.B. *Holthuis* 2000, 154 ff.):

- *Selection, Slice* und *Dice*: *Selection* ermöglicht eine Auswahl von Werten anhand vorgegebener Kriterien (z.B. Finde die drei Produktgruppen mit den höchsten Monatsumsätzen in 11/2000). *Slice* und *Dice* stellen Spezialfälle der Operation *Selection* dar. *Slice* ermöglicht das »Herausschneiden einer Scheibe« aus dem Hypercube (z.B. alle Umsätze in den Dimensionen Produkt und Zeit für die Filiale München-Süd). Das Anfrageergebnis weist dabei eine gegenüber dem Hypercube um eins reduzierte Anzahl an

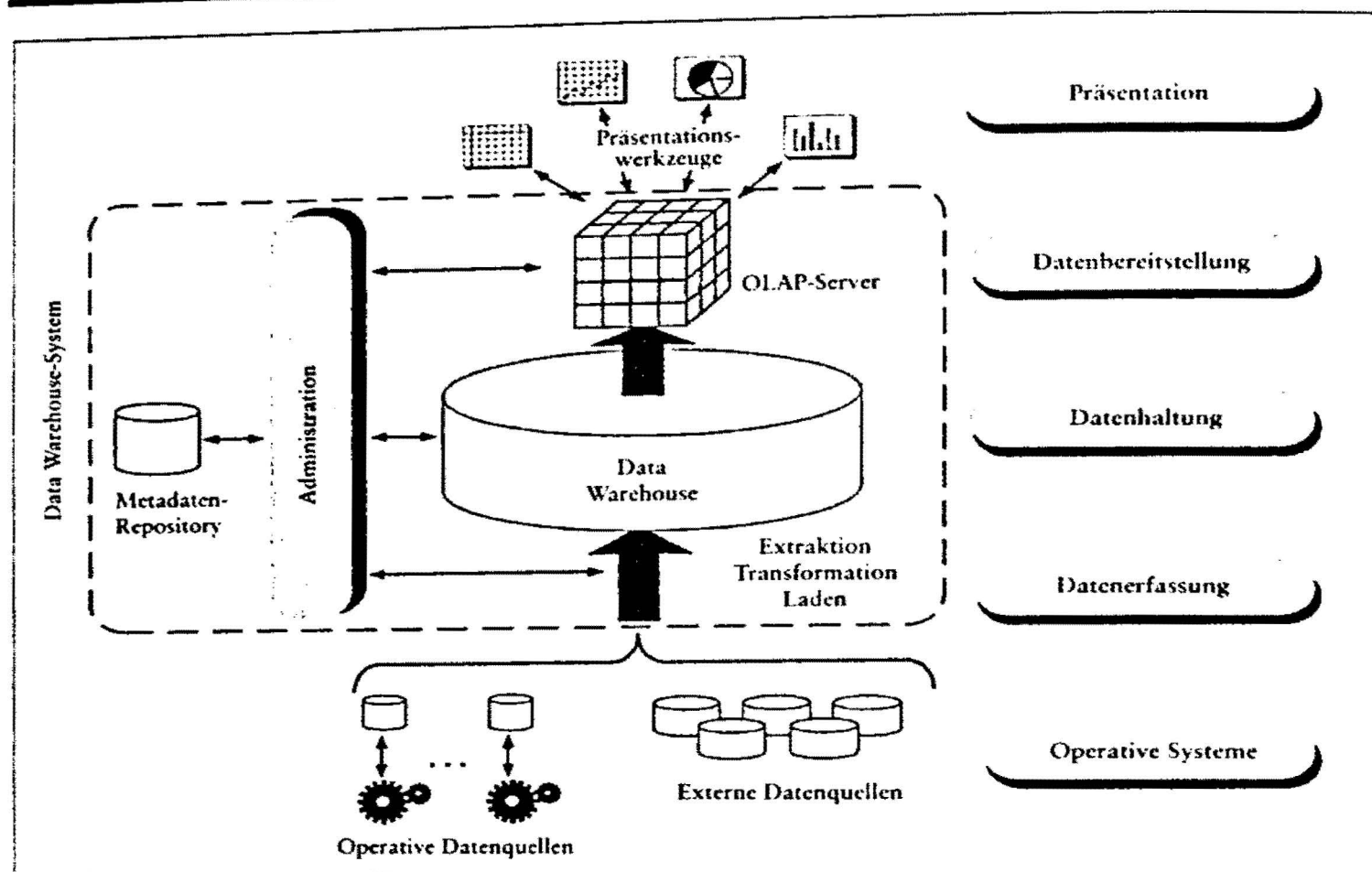


Abb. 3: Architektur eines Data Warehouse-Systems

Dimensionen auf (z.B. wird aus einem dreidimensionalen Würfel eine zweidimensionale Tabelle selektiert). Die Operation *Dice* selektiert einen Teilwürfel des Hypercube; die Anzahl der Dimensionen bleibt dabei unverändert (z.B. Umsätze aller deutschen Filialen im Jahr 2000).

- *Drill Down* und *Roll Up*: Navigieren innerhalb einer Dimensionshierarchie. *Drill Down* navigiert eine Ebene nach unten (z.B. von Umsätzen je Produktkategorie zu Umsätzen je Produkthauptgruppe oder von Monatsumsätzen zu Tagesumsätzen). *Roll Up* navigiert umgekehrt eine Ebene nach oben.
- *Rotate*: Drehen des Hypercube durch Vertauschen von zwei Dimensionen (z.B. Vertauschen von Zeit und Produkt). Dadurch ergibt sich eine veränderte Sicht auf die Datenstruktur, die

quantitativen Daten selbst bleiben unverändert.

- *Join*: Verknüpfen von zwei Hypercubes anhand einer oder mehrerer gemeinsamer Dimensionen. Im Allgemeinen erhöht sich dabei die Anzahl der Dimensionen des Anfrageergebnisses im Vergleich zu den verknüpften Hypercubes. Werden z.B. ein 4- und ein 6-dimensionaler Hypercube über 2 gemeinsame Dimensionen verknüpft, so besitzt das Anfrageergebnis $4 + 6 - 2 = 8$ Dimensionen.

III. Architektur

Ein DWH wird zusammen mit der für Aufbau, Betrieb und Nutzung erforderlichen Software als Data-Warehouse-System bezeichnet. Die idealtypische Archi-

tektur eines Data Warehouse-Systems ist in Abb. 3 aus funktionaler Sicht dargestellt (vgl. auch *Bauer/Günzel* 2001, 34 ff.; *Chamoni/Gluchowski* 1999, S. 268ff). Danach sind die Funktionen eines Data Warehouse-Systems auf drei hierarchisch angeordnete Schichten verteilt. Die untere und die obere Schicht besitzen Schnittstellen zur Daten bereitstellenden bzw. zur nutzenden Umgebung des Data Warehouse-Systems. Zusammen mit dieser Umgebung sind insgesamt fünf Schichten unterscheidbar, anhand derer die Funktionsweise eines Data Warehouse-Systems beschrieben werden kann:

- *Operative Systeme*: Datenquellen für das Data Warehouse sind insbesondere die Datenbanken der operativen Anwendungssysteme sowie externe Datenquellen (vgl. z.B. *Fischer* 2001, S. 107 ff.), wie etwa öffentliche Datenbanken.
- *Datenerfassung*: Durchführung des ETL-Prozesses, bestehend aus den Schritten Extraktion, Transformation und Laden. Bei der Extraktion werden Rohdaten aus den Datenquellen selektiv übernommen und gefiltert. Bei der Transformation erfolgt eine syntaktische und semantische Bereinigung der Daten (vgl. z.B. *Kemper/Finger* 1999, 78 ff.). Beim Laden werden die bereinigten Daten organisatorisch aufbereitet (z.B. durch Sortieren, Berechnen von Aggregationen, Aufbau von Zugriffsstrukturen) und in das Data Warehouse übernommen.
- *Datenhaltung*: Verwaltung des Data Warehouse, d.h. der eigentlichen Datenbasis. Die Datenhaltung erfolgt analog zu den Datenbanken der operativen Systeme meist in relationaler Form unter Nutzung eines relationalen Datenbanksystems. Die Relationen des DWH lassen sich dabei als materialisierte, redundant gespeicherte Sichten auf die Relationen der DB (und auf externe Datenquellen) interpretieren. Im Gegensatz dazu werden innerhalb der operativen Anwendungssysteme virtuelle Sichten auf die DB verwendet, die jeweils aus den Basisrelationen der DB abgeleitet werden. Bei den Organisationsformen für die Datenhaltung dominiert das zentrale DWH, es kommen aber auch verteilte DWH zum Einsatz, die auf Basis verteilter Datenbanksysteme realisiert werden und die gegenüber der nächsthöheren Schicht einen logisch zentralisierten Datenbestand präsentieren. Von einem zentralen oder verteilten DWH zu unterscheiden sind *Data Marts*. Ein Data Mart ist ein thematisch oder bereichsspezifisch abgegrenztes Teil-DWH für eine bestimmte Organisationseinheit oder für ein spezielles Entscheidungsfeld. Eine Konsolidierung zwischen den Data Marts erfolgt i. Allg. durch Abstimmung mit einem zentralen DWH.
- *Datenbereitstellung*: Bereitstellung der Datenbestände für die nutzende Umgebung des DWH. Die Datenbereitstellung erfolgt in Form von multidimensionalen Datenstrukturen durch einen OLAP-Server. Bezüglich der Funktionsweise von *OLAP-Servern* lassen sich zwei Grundformen unterscheiden (vgl. z.B. *Chamoni/Gluchowski* 2000, S. 344 ff.). Beim relationalen OLAP (ROLAP) werden die auf dem Hypercube durchzuführenden Operationen in relationale Operationen transformiert (i.d.R. SQL), die meist direkt auf den Relationen des DWH ausgeführt werden. Beim multidimensionalen OLAP (MOLAP) werden zusätzlich zum DWH spezielle multidimensionale Daten- und Zugriffsstrukturen

aufgebaut, die den Hypercube physisch realisieren und auf denen die Operationen ausgeführt werden. Hierzu wird ein multidimensionales Datenbanksystem eingesetzt. Als Vorteil von ROLAP gilt insbesondere eine höhere Flexibilität der Datenstrukturen, als Vorteil von MOLAP eine schnellere Anfragebearbeitung.

- *Präsentation*: Darstellung und Auswertung der Anfrageergebnisse in Form von interaktiven Berichten und Diagrammen. Werkzeuge hierfür sind insbesondere OLAP-Standardberichte für bestimmte betriebliche Funktionsbereiche, OLAP-Berichtsgeneratoren sowie Erweiterungen von Standard-Bürosoftware (siehe z.B. *Chamoni/Gluchowski* 2000, S. 364 ff.).

Querschnittlich zu den Schichten sind Funktionen zur Administration des Data Warehouse-Systems vorgesehen. Diese greifen auf Meta-Daten zurück, die in einem zugehörigen Repository verwaltet werden. Diese Meta-Daten beschreiben insbesondere die Datenstrukturen des DWH sowie die verschiedenen Transformationen von der Datenerfassung bis zur Datenbereitstellung.

IV. Entwicklung und Betrieb

Die Entwicklung eines umfassenden Data Warehouse-Systems ist ein komplexes Unterfangen und mit entsprechend hohen Projektrisiken behaftet. Zur Bewältigung dieser Komplexität und zur Reduzierung der Risiken wird im Allgemeinen ein evolutionäres, prototypisches Vorgehen gewählt, dessen Grundstruktur aus den Phasen Analyse, Design und Implementierung besteht, die mehrfach zy-

klisch durchlaufen werden (vgl. *Wieken* 1999, S. 233 ff.):

- *Analysephase*: Ein erster Aufgabenschwerpunkt ist die Informationsbedarfsanalyse. Dabei sind der aus den Managementaufgaben ableitbare objektive Informationsbedarf, der subjektive Informationsbedarf der Entscheider, die tatsächlich zu erwartende Informationsnachfrage und das potenzielle Informationsangebot zu analysieren und aufeinander abzustimmen (vgl. *Biethahn/Muksch/Ruf* 1994, S. 7 f.). Aufbauend darauf erfolgt die Entwicklung von Nutzungsszenarien für das DWH. Dazu werden die ermittelten Informationsbedarfe klassifiziert und geeignet gebildeten Nutzergruppen zugeordnet. Abschließend erfolgt die fachliche Definition der Hypercubes mit Kennzahlen und Dimensionen.
- *Designphase*: Kern der Designphase ist der softwaretechnische Entwurf der multidimensionalen Datenstrukturen. Aufbauend auf der fachlichen Definition der Hypercubes und einer Entscheidung zwischen ROLAP und MOLAP sind Maßnahmen zur Optimierung des Anfrageverhaltens zu treffen. Diese betreffen insbesondere die Vorstrukturierung und Vorverdichtung der Daten des DWH einschließlich der Definition von Standardberichten. Weiter wird das Anwendungsdesign durchgeführt. Dieses umfasst das Layout und die Struktur von Berichten, die Definition der Nutzerberechtigungen sowie die Planung von Hilfsmitteln zur Nutzerunterstützung.
- *Implementierungsphase*: Diese Phase dient der Implementierung des Data Warehouse-Systems auf der Basis einer ausgewählten Systemplattform (Daten-

banksystem, OLAP-System) einschließlich der Schnittstellen zu den operativen und externen Datenquellen sowie der Realisierung der gesamten Anwendungsumgebung.

Das evolutionäre Vorgehen bei der Entwicklung eines Data Warehouse-Systems ermöglicht eine sukzessive Umsetzung der Anforderungen der Entscheider, die in vielen Fällen erst während der Nutzung des Systems zutage treten. Ein sukzessiver Aufbau kann auch zunächst die Entwicklung isolierter Data-Marts vorsehen, die dann später zu einem umfassenden Data Warehouse-System integriert werden.

Voraussetzung für eine erfolgreiche Entwicklung und einen effektiven Betrieb eines Data Warehouse-Systems ist neben der Beachtung von fachlichen und systemtechnischen Faktoren eine hinreichende Berücksichtigung der relevanten Rahmenbedingungen. Hierzu gehören sowohl rechtliche Regelungen (z.B. Datenschutz und Mitbestimmung bei der Speicherung personenbezogener Daten) als auch »weiche Faktoren« wie Akzeptanz und Vertrauen aller Beteiligten. Muss z.B. bei einer verteilten Organisation mit (teil-) autonomen Organisationseinheiten ein Entscheider davon ausgehen, dass Daten seines Autonomiebereichs auch externen Personen zugänglich sind, so besteht die Gefahr, dass die Datenbereitstellung für das DWH verzögert oder unterlaufen wird. Ein Lösungsansatz besteht hier im Aufbau eines verteilten DWH, dessen Verteilungsstruktur sowie die Inhalte der einzelnen Teil-DWH mit der Organisations- und Führungsstruktur der verteilten Organisation abgestimmt sind (vgl. *Sinz/Böhnlein/Plaha et al.* 2001).

Literatur

- Bauer, Andreas/Günzel, Holger:* Data-Warehouse-Systeme. Architektur, Entwicklung, Heidelberg 2001.
- Biethahn, Jörg/Muksch, Harry/Ruf, Walter:* Ganzheitliches Informationsmanagement. Band 1: Grundlagen, München et al., 3. A., 1994.
- Chamoni, Peter/Gluchowski, Peter:* On-Line Analytical Processing (OLAP), in: Das Data Warehouse-Konzept. Architektur – Datenmodelle – Anwendungen, hrsg. v. *Muksch, Harry/Behme, Wolfgang*, Wiesbaden, 4. A., 2000, S. 333–376.
- Chamoni, Peter/Gluchowski, Peter:* Analytische Informationssysteme – Einordnung und Überblick, in: Analytische Informationssysteme. Data Warehouse, On-Line Analytical Processing, Data Mining, hrsg. v. *Chamoni, Peter/Gluchowski, Peter*, Berlin et al., 2. A., 1999, S. 3–25.
- Chamoni, Peter/Gluchowski, Peter:* Entwicklungslinien und Architekturkonzepte des On-Line Analytical Processing, in: Analytische Informationssysteme. Data Warehouse, On-Line Analytical Processing, Data Mining, hrsg. v. *Chamoni, Peter/Gluchowski, Peter*, Berlin et al., 2. A., 1999, S. 261–280.
- Ferstl, Otto K./Sinz, Elmar J.:* Grundlagen der Wirtschaftsinformatik. Band 1, München et al., 4. A., 2001.
- Fischer, Cai:* Externe Daten als Achillesferse von Data-Warehouse-Projekten – Probleme und Lösungsansätze, in: Data Warehouse Managementhandbuch. Konzepte, Software, Erfahrungen, hrsg. v. *Schütte, R./Rotthowe, T./Holten, R.*, Berlin et al. 2001, S. 107–118.
- Holthuis, Jan:* Grundüberlegungen für die Modellierung einer Data Warehouse-Datenbasis, in: Das Data Warehouse-Konzept. Architektur – Datenmodelle – Anwendungen, hrsg. v. *Muksch, Harry/Behme, Wolfgang*, Wiesbaden, 4. A., 2000, S. 149–214.
- Inmon, William H.:* Building the Data Warehouse, New York et al., 2. A., 1996.
- Kemper, Hans-Georg/Finger, Ralf:* Konzeptionelle Überlegungen zur Filterung, Harmonisierung, Verdichtung und Anreicherung operativer Datenbestände, in: Analytische Informationssysteme. Data Warehouse, On-Line Analytical Processing, Data Mining, hrsg. v. *Chamoni, Peter/Gluchowski, Peter*, Berlin et al., 2. A., 1999, S. 77–94.

- Küpper, Hans-Ulrich*: Controlling. Konzeption, Aufgaben und Instrumente, Stuttgart et al., 2. A., 1997.
- Muksch, Harry/Behme, Wolfgang*: Das Data Warehouse-Konzept als Basis einer unternehmensweiten Informationslogistik, in: Das Data Warehouse-Konzept. Architektur – Datenmodelle – Anwendungen, hrsg. v. *Muksch, Harry/Behme, Wolfgang*, Wiesbaden, 4. A., 2000, S. 3–80.
- Sinz, Elmar J./Böhnlein, Michael/Plaha, Markus et al.*: Architekturkonzept eines verteilten Data-Warehouse-Systems für das Hochschulwesen, in: Information Age Economy, hrsg. v. *Bull, H.-U./Huther, A./Reitwiesner, B.*, Heidelberg 2001.
- Wieken, John-Harry*: Der Weg zum Data Warehouse, München et al. 1999.

HWU, 4. Aufl. 2002