

16 Data Warehouse

16.1 Grundlagen

Data Warehouse Systeme und Data Mining Verfahren erfahren in den letzten Jahren als Management Informationssysteme (MIS) eine immer größer werdende Bedeutung. Die Information – speziell über Kunden – wird dabei als wichtiger Erfolgsfaktor erkannt, der das Bestehen in hart umkämpften Märkten ermöglicht.

Auf Grund einer zunehmenden Datenflut in den verschiedenen operativen Systemen (OLTP-Systeme, Online Transaction Processing) sehen sich viele Unternehmen mit den früheren Management Informationssystemen der „Executive Information Systems“ (EIS) jedoch nicht in der Lage, die relevanten Informationen zur Entscheidungsfindung aufzubereiten, weil diese lediglich standardisierte Auswertungen erlauben und ihren Schwerpunkt bei der Abwicklung von betrieblichen Prozessen haben.

Data Warehouse Systeme bilden als zentrale Datenbank die Grundlage für Analysetools und neuartige Decision Support Systeme (DSS) zu umfangreichen Auswertungen und multidimensionaler Datenanalyse (beispielsweise mit OLAP). Mit multidimensionaler Datenanalyse können die Daten, die aus unterschiedlichen operationalen Systemen stammen, gesamtheitlich aus beliebigen Blickwinkeln betrachtet und ausgewertet werden.

Data Mining Verfahren ergänzen die Verfahren der Decision Support Systeme, indem sie selbstständig Beziehungsmuster in den Daten erkennen; sie setzen sich aus bereits bekannten Verfahren der klassischen Marktforschung und der künstlichen Intelligenz zusammen (vgl. Kapitel 17). Sie generieren in einem Lernprozess an Hand der in Data Warehouse Systemen gesammelten Informationen neue Erkenntnisse für das Management.

Data Warehouse Systeme und Data Mining Verfahren werden für das Direktmarketing in Zusammenhang mit Customer Relationship Management und One-to-One Marketing bedeutsam. Dabei geht es im Allgemeinen darum, der steigenden Anzahl verschiedener Lebensstilgruppen, also Kunden, die in ihrem Kaufverhalten kritischer werden, gerecht zu werden oder in Angesicht einer häufig hohen Fluktuation innerhalb der Gesamtkundschaft Kundenbindung zu betreiben, um den Kundenlebenszyklus zu verlängern.

Der Kunde wird dafür mit Hilfe von Data Warehouse Systemen und Data Mining zum „transparenten Kunden“, soweit es die rechtlichen Restriktionen zulassen.

Es gilt, dem richtigen Kunden zur richtigen Zeit am richtigen Ort ein angemessenes Angebot zu machen.

16.2 Datenqualität

Ein Data Warehouse ist nur dann für das Direktmarketing nutzbar, wenn die Kundendaten möglichst aktuell, fehlerfrei und vollständig sind.

Die Qualität der Kundendaten wird nach vier Dimensionen bewertet:

- **Technische Datenqualität**
Die technische Datenqualität ist dann gegeben, wenn die Daten richtig übertragen wurden.
- **Formale Datenqualität**
Die Datentypen müssen den Vorgaben entsprechen; es sollen keine leeren Felder oder Leerzeichen auftreten.
- **Inhaltliche Datenqualität**
Die inhaltliche Datenqualität leidet, wenn die Datenbank falsche Anreden oder Fehler in den Adressen enthält. Eine nicht zeitnahe Aktualisierung und schlechte Pflege wirkt sich fatal auf die Qualität von Kundenselektionen aus.
- **Komplette Datenqualität**
Die Gesamtbetrachtung der Datenqualität beschäftigt sich beispielsweise mit der Frage, welche wichtigen Daten über die Kunden noch fehlen, um diese richtig bewerten zu können.

Eine empfehlenswerte Vorgehensweise zur Steigerung der Qualität von Data Warehouses könnte nach dem in der Abbildung 105 dargestellten Schema ablaufen.

1. Korrektur der Adressen	Vergleich der Adressen mit einer aktuellen Datenbank aller existierenden Postleitzahlen, Orte und Straßen
2. Prüfung der Existenz der Adressen	Vergleich der Adressen mit einer aktuellen Datenbank aller vorhandenen Häuser und Prüfung auf Existenz
3. Prüfung der Existenz der Person und Korrektur des Namens	Abgleich der Personen mit einer aktuellen Datenbank aller existierenden Menschen an den jeweiligen Adressen
4. Korrektur der Änderungen von Adressen	Abgleich der Adressen gegen eine aktuelle Datenbank mit allen postalischen Änderungen
5. Nachtragen der Adressen bei Umzügen	Vergleich der Personen mit den umgezogenen Menschen an Hand der Umzugsdatenbank
6. Sterbefälle	Abgleich der Personen in der Datenbank mit Spezialdatenbanken mit gestorbenen Menschen
7. Korrektur der Anrede (Geschlecht)	Vergleich des Geschlechts des Vornamens mit dem eingetragenen Geschlecht in der Referenz-Datenbank
8. Korrektur der Titel	Abgleich mit den eingetragenen Titeln in einer Referenz-Personendatenbank
9. Ergänzung bzw. Überprüfung von Telefonnummern	Ergänzung der Daten durch Spezialdatenbanken mit Telefonnummern, E-Mail-Adressen etc.

10. Ergänzung und Überprüfung des Alters	Ergänzung der Daten durch Spezialdatenbanken mit Altersdaten oder Schätzung durch Vornamensanalyse
11. Überprüfung der Berufsbezeichnungen etc.	Spezialprogramme, die beispielsweise durch Plausibilitätsprüfung 50-jährige Studenten erkennen
12. Auffinden von Inkonsistenzen in den Daten	Spezialprogramme, z. B. durch Zählungen, Gruppierungen, Klassenbildung etc.
13. Auffinden von Dubletten	Einsatz von entsprechenden Programmen zur Dubletten-Eliminierung
14. Auffinden von Haushalten	Programme zum Auffinden von Haushalten unter Nicht-Berücksichtigung des Vornamens
15. Eigene Spezialprogramme	Viele Fälle sind unternehmensspezifisch, z. B. unmögliche Produktkombinationen etc.

Abbildung 105: Steigerung der Datenqualität

16.3 Data Warehouse und Data Mining im Regelkreis des Database Marketings

Database Marketing basiert auf kundenindividuellen in einer Kundendatenbank gespeicherter Informationen und zielt auf einen Regelkreis in Form eines marktgerechten und marktgerichteten Entscheidungsfindungs- und Lernprozesses, in dem in einem Dialog mit dem Kunden immer wieder folgende Schritte durchlaufen werden (Link, Hildebrand, 1993, S. 30):

- Zieldefinition
- Strategie, weiteres Vorgehen
- Aktion
- Reaktion
- Analyse, Auswertung, Kontrolle

Das Ziel ist es, durch Direkt- oder Dialogmarketing, den Kontakt mit den Kunden bzw. Wunschkunden so eng wie möglich zu halten, um auf diese Weise langfristige Kundenbeziehungen aufzubauen.

Das Data Warehouse für Marketing (siehe Abbildung 106) wird jedoch nicht ausschließlich aus den Daten in der Marketingdatenbank, sondern aus allen relevanten unternehmensinternen und -externen Daten gespeist und ermöglicht eine gesamtheitliche Betrachtung des Kunden oder Wunschkunden zu Analysezwecken (Mentzl, Ludwig, 1998, S. 487).

Auch Daten, die nicht in der Marketingdatenbank enthalten sind, können für die Analyse wertvoll sein, beispielsweise Daten der Warenwirtschaft.

Während sich die unternehmensinternen Daten in den verschiedenen operationalen Systemen befinden, werden externe Daten wie soziodemografische Daten von Marktforschungsinstituten oder Adressdaten von Informationsbrokern häufig zugekauft, um Wunschkunden gezielt ansprechen zu können.

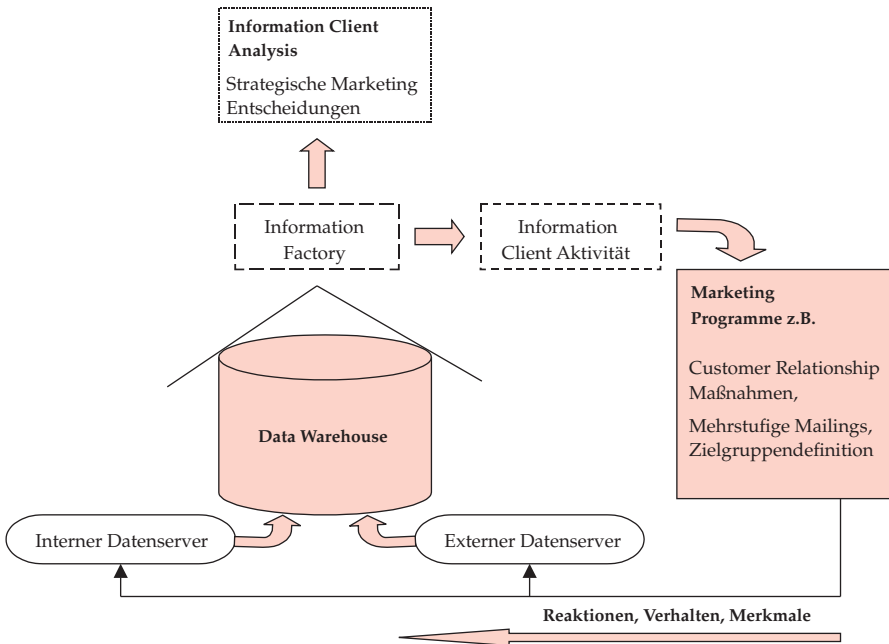


Abbildung 106: Logisches Modell eines Database Marketing Systems

Quelle: Mentzl, Ludwig, 1998, S. 487

Externe Daten können aus verschiedenen Datenquellen stammen. Sie werden auch dazu benutzt, bereits bestehende Kundendaten mit zusätzlichen Daten anzureichern, um daraus neue Informationen für Marketingmaßnahmen zu gewinnen.

Datenquellen für unternehmensexterne Daten:

- Marktforschung
- Erhebungen
- Transaktionen
- Adressdaten
- Luftbilder
- Register
- Statistik

Beispiel:

Die Deutsche Bank 24 verwendet als externe Datenquelle auch eine mikro-geografische Datenbank von microm (Mikromarketing-Systeme und Consult GmbH). Die Datenbank weist für jedes Haus in Deutschland den Anteil der darin existierenden Haushalte nach zehn Milieus und nach 38 MOSAIC-Typen aus, die Aussagen über Kaufkraft, Haushaltsgröße, Status, Familienstruktur und andere wichtige Hinweise geben. Durch einen Abgleich der Kundendatenbank mit diesen Daten werden noch genauere Informationen für Marketing-Maßnahmen gewonnen.

Nachdem externe und interne Daten über einen Datenserver in das Data Warehouse geladen sind, erfolgt in der Information Factory die inhaltliche Weiterverarbeitung für multidimensionale Betrachtung und besondere marketingsspezifische Auswertungen.

Die Information Client Analysis ermöglicht durch Decision Support Systeme (DSS), OLAP und Data Mining das Treffen strategischer Marketingentscheidungen, während der Information Client Activity Marketing Programme steuert, die in der Information Client Analysis beschlossen werden.

Diese Aktivitäten zielen direkt auf den Dialog mit dem Kunden oder potenziellen Kunden. An Hand von Responseelementen werden Verhalten und Merkmale erfasst und fließen als Reaktionen in das Data Warehouse zurück.

Sie führen in der Information Client Analysis durch Auswertungen zu einer besseren Selektionsgrundlage für die nächsten Marketing-Aktionen und zu einem Lernprozess. In der Information Client Activity werden die Reaktionen genutzt, um daraufhin den nächsten Schritt in einer vorab festgelegten Marketing-Aktion durchzuführen, beispielsweise eine Nachfassaktion des Call Centers.

In einem fortlaufenden Dialog, auch in Form von mehrstufigen Mailings, soll durch Auswertung der Maßnahmen in der Information Client Analysis und durch deren Steuerung in der Information Client Activity die Ansprache an die individuellen Wünsche und Bedürfnisse der Kunden angepasst werden. Grundsätzlich stehen dem Unternehmen dafür sämtliche Instrumente zur gezielten Kundenansprache, beispielsweise auch Call-Center und Vertriebsmitarbeiter zur Verfügung, um Cross Selling Potenziale zu nutzen oder für einen Kunden oder potenziellen Kunden die geeigneten Produkte zu finden.

Beispiel:

Zur Steuerung der Marketing-Maßnahmen setzt die Deutsche Bank 24 das Tool Prime Vantage von Prime Response ein. Damit können Marketing-Kampagnen, nachdem die Maßnahmen auf den einzelnen Stufen festgelegt wurden, weitgehend automatisiert durchgeführt und ausgewertet werden.

Prime Vantage ermöglicht die Integration des Dialogs mit dem Kunden in den verschiedenen Marketing-Kanälen beispielsweise Mailing, Callcenter und Vertrieb, indem es automatisch an Hand der fest gelegten Auswertung der Response die entsprechende Person im Unternehmen (Vertrieb, Callcenter) informiert oder automatisch ein Mailing veranlasst.

16.4 Begriff des Data Warehouse

Unter einem Data Warehouse wird allgemein, in Analogie zum Begriff des Lagerhauses, eine zentrale Datenbank verstanden, auf deren Inhalte die Nutzer schnell, leicht und systematisch zugreifen können. Hierbei können alle verfügbaren internen und externen Datenquellen genutzt und die relevanten Informationen abgerufen werden (Wacker, 2002, S. 883 f.).

In großen Unternehmen existiert eine Vielzahl möglicher Datenquellen. Die Zielsetzung, für die Anwender alle relevanten Informationen schnell und leicht zur Verfügung zu stellen, stellt viele Unternehmen vor komplexe Aufgaben (Knowledge-Management). Es ist zu bedenken, dass ein Data Warehouse lediglich die relevanten Informationen beinhalten kann, die in Informationsbedarfsanalysen bestimmt wurden. Das Marketing-Data Warehouse umfasst nur Daten, die für das Marketing nützlich sind.

Das Data Warehouse enthält zudem vorwiegend aggregierte Daten.

Beispiel:

Vor diesem Hintergrund ist es zu erklären, dass die Deutsche Bank 24 ein Data Mart für das Customer Relationship Management entwickelt hat, obwohl ein Data Warehouse bereits vorhanden ist.

Data Marts sind „Mini-Data Warehouses“ für die einzelnen Geschäftsbereiche oder Fachabteilungen im Unternehmen, die weniger stark aggregierte Daten als das Data Warehouse verwenden und in der Regel ein Data Warehouse speisen (Kimball, Merz, 2000, S. 362).

„Mit dem Begriff Data Warehouse i. e. S. wird generell eine von den operationalen DV-Systemen isolierte Datenbank umschrieben, die als unternehmensweite Datenbasis für alle Ausprägungen managementunterstützender Systeme dient und durch eine strikte Trennung von operationalen und entscheidungsunterstützenden Daten und Systemen gekennzeichnet ist.“ (Mucksch, Brehme, 1998, S. 36).

Die Abbildung 107 zeigt den Aufbau einer Data Warehouse eines Versandhändlers. Diese Datenbank speist sich aus unterschiedlichen internen Systemen, wie dem Kundeninformationssystem, der Artikeldatenbank und der Auftragsabwicklung, und wird durch externe Daten, wie mikrogeografischen Analysen, angereichert.

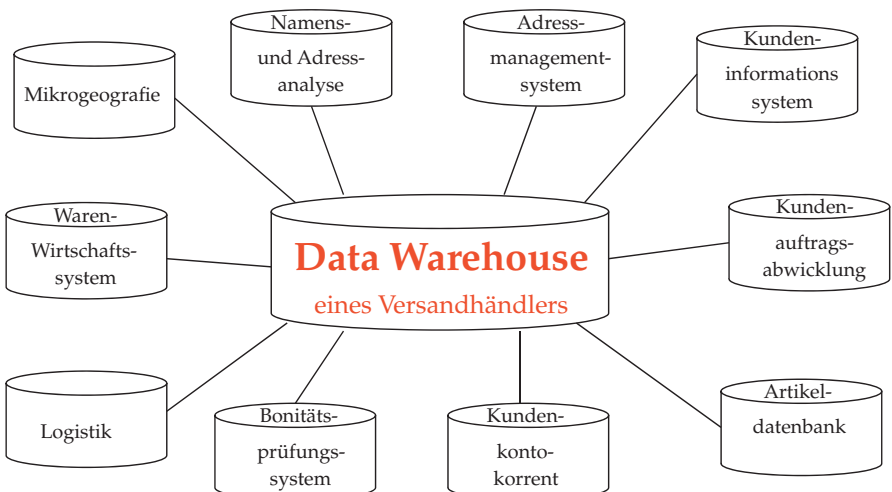


Abbildung 107: Data Warehouse eines Versandhändlers

16.5 Wesentliche Bestandteile eines Data Warehouse

Mit dem Ziel ein Data Warehouse aufzubauen, werden die strategisch relevanten Informationen aus den OLTP-Systemen und externen Quellen extrahiert, transformiert und anschließend in ein Data Warehouse Datenmodell geladen (siehe Abbildung 108).

Beispiel:

Bei der DEKA-Bank wurden die Anforderungen an das Data Warehouse und die relevanten abrufbaren Informationen in einer Informationsbedarfsanalyse der Fachabteilungen Marketing und Vertrieb und dem zuständigen Projektteam DekaData Bereich UI festgelegt.

Das Data Warehouse soll vorrangig zur Reporterstellung und zur multidimensionalen Datenanalyse eingesetzt werden. Die Zielsetzung ist das bisherige interne Reporting und externe Reporting an die Vermittler (Sparkassen und Landesbanken) der DEKA Fonds durch ein zeitgemäßes Online-Reporting zu ersetzen, um die Fachabteilungen in Marketing und Vertrieb zu entlasten. Zusätzlich soll dazu ein Online-Analysesystem zur Unterstützung der Fachabteilungen Marketing und Vertrieb mit multidimensionaler Datenanalyse eingerichtet werden. Das Data Warehouse wird zu diesem Zweck aus verschiedenen operationalen Datenquellen (den Host Systemen DekabankDepot (DBD), Emissionsgeschäft (EM) u. a.) gespeist. Das Extrahieren, Aufbereiten und das abschließende Laden (wird hier synonym mit Refresh verwendet) in ein Data Warehouse Datenmodell wird als Transformation bezeichnet und geschieht in der DEKA Bank täglich.

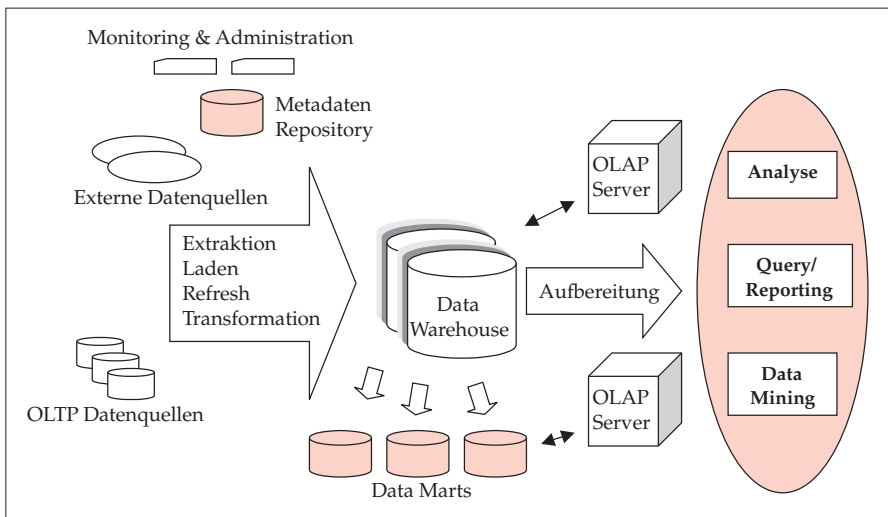


Abbildung 108: Die Architektur der Data Warehouses

Quelle: Surajit, Umeshwar, 2000, S. 518

Die Transformation der Daten setzt sich aus deren Filterung, Harmonisierung, Verdichtung und Anreicherung zusammen.

Filterung umfasst die Extraktion und Bereinigung der operativen Daten. Dabei werden aus Performancegründen lediglich die Daten aus den Quellsystemen extrahiert, die auch wirklich benötigt werden. Daraufhin folgt in der Harmonisierung die wirtschaftliche Abstimmung der gefilterten Daten. Dabei werden Disharmonien, die durch die Heterogenität der unterschiedlichen operativen Quellsysteme bedingt sind, beispielsweise unterschiedliche Kodierungen, Synonyme, Homonyme und Schlüsselverletzungen, beseitigt. Die Verdichtung betrifft die Aggregation gefilterter und harmonisierter Daten. In der Anreicherung werden aus Performancegründen auf Basis konsistenter Dimensionsdefinitionen Summenstrukturen erzeugt, denn die Anwender interessieren sich in der Berichterstellung und der multidimensionalen Datenanalyse vorwiegend für das Ermitteln von Kennzahlen.

Die Dimensionen werden im Data Warehouse Datenmodell getrennt von den Bestandsdaten, die in eine oder mehrere Fakttabellen überführt werden, in speziellen Strukturtabellen meist denormalisiert gehalten.

Es ist beim Transformationsprozess grundsätzlich zu beachten, dass die Data Warehouse Datenbank nicht überladen wird, denn dies würde zu erheblich längeren Ladezeiten führen und die Abfrageperformance der einzusetzenden Tools stark beeinträchtigen.

In der Metadaten Repository befindet sich die zentrale Sammlung der notwendigen Meta Daten, die benötigt werden, um die Daten aus den verschiedenen Quellsystemen zu extrahieren und in das zentrale Data Warehouse Datenmodell überführen zu können.

Im Bereich Monitoring und Administration wird diese Metadaten Repository verwaltet und die Transformation von einem Projektteam überwacht und gesteuert.

Ein wichtiger Aspekt eines Data Warehouses ist auch die Datenqualität, um keine verfälschten Ergebnisse zu erzielen. Fehler in der Datenqualität können verschiedene Ursachen haben, die in der Erfassung der Daten in den operationalen Systemen liegen, und werden nicht allein unter dem Aspekt der Bereinigung fehlerhafter Daten aufgefasst, sondern stellen einen Prozess dar, um die Qualität langfristig zu sichern und messbar zu machen (Häusler, 1988, S.75).

Beispiel:

In der Deutschen Bank 24 soll ausgehend von einer Initialanalyse der bestehenden Daten ein Datenqualitäts-Kreislauf initiiert werden. Beginnend mit der Mängelerkennung und dem darauf folgenden Reporting werden dafür zuständige Personen im Vertrieb informiert. Diese verteilen die Informationen über fehlerhafte Daten an entsprechende Mitarbeiter, die daraufhin die fehlerhaften Eingaben in dem jeweiligen Erfassungssystem korrigieren. Abschließend wird zur Kontrolle eine Überprüfung und Korrektur durch Filialmitarbeiter durchgeführt, um sicherzustellen, dass die Änderungen auch umgesetzt werden.

Ein Data Warehouse lässt sich auch in verschiedene Data Marts aufteilen.

Das Customer Relationship Management der Deutsche Bank 24 basiert auf einem Data Mart für ein Kundenanalyse- und Steuerungssystem, das als „DataSmart“ bezeichnet wird.

Die Zielsetzung ist es, Beratung und Vertrieb im Vergleich zum bisher eingesetzten Data Warehouse, das vielen Anwendern dient, anwenderspezifischere Analysen, Reports und Auswertungen zu ermöglichen und die Steuerung der Kampagnen mit Prime Vantage weitgehend zu automatisieren. Die zentrale Gruppe Marketinganalyse soll mit Hilfe der OLAP-Anwendungen entlastet werden, indem das System selbstständig über das Intranet Analysen durchführen kann und die bisherigen Reports und das Wissensportal für Vertrieb, Zentrale und Regionen erweitert werden.

Das Data Mart erhält dazu detaillierte anwenderbezogene Daten aus den operationalen Datenbanken zu Spareinlagen, Depotbeständen und Zahlungsverkehr der Kunden und bereits bestehende Daten aus dem Data Warehouse. Die operationalen Datenbanken erhalten wiederum relevante Daten aus dem Data Mart und das Data Warehouse erhält die Daten, die auch für das Management interessant sind. Zusätzlich wird das Data Mart mit externen mikrogeografischen Daten von microm angereichert, die für jedes Haus in Deutschland den Anteil der darin existierenden Haushalte nach zehn Milieus und nach 38 MOSAIC Typen ausweisen.

Das Data Warehouse bzw. die Data Marts speisen die OLAP-Systeme, welche die Daten für Analysen, Reporting und Data Mining aufbereiten und hierfür erweiterte Analysefunktionalitäten bereitstellen. In der Deutschen Bank 24 wird als OLAP-Tool Cognos Powerplay eingesetzt und für die Analyse SAS Enterprise Miner.

16.6 Datenmodellierung im Data Warehouse

Ein Ziel der Datenmodellierung für das Data Warehouse ist es, die auszuwertenden Daten in eine integrierte und konsistente Form zu überführen. Während bei den relationalen Datenbanken der operativen Systeme der Fokus auf den Transaktionen des Betriebs liegt, die möglichst effizient abgewickelt werden und durch Normalisierung wenig Redundanzen (mehrfach geführte Daten) in der Datenhaltung beinhalten sollen, verfolgt die Datenmodellierung für das Data Warehouse den Zweck der multidimensionalen Datenanalyse (Kirchner, 1998, S. 160).

Unter Normalisierung wird der Prozess zur Vermeidung von Redundanzen verstanden, indem die Daten, die sich ursprünglich alle in einer Tabelle befinden, in verschiedene Tabellen eingeteilt werden, die in einer Relation zueinander stehen.

In der multidimensionalen Datenanalyse geht es beispielsweise um folgende Fragestellungen (Jarke, 2000, S. 87):

- Wie hoch war der Umsatz differenziert nach Regionen und Produktkategorien in den letzten Jahren?
- Wie viele Produkte wurden zu den östlichen Regionen im letzten Quartal in welchen Mengen verschickt, die größer als zehn waren, und wie groß war der Deckungsbeitrag, der damit realisieren werden konnte, im Verhältnis zu denen, deren Mengen kleiner als zehn war?

- Wie korrelierten die Ausgaben für Direktmarketing mit deren Erfolg im letzten Jahr?

Die Datenmodelle der operativen Systeme erweisen sich für diese Fragestellungen jedoch als ineffizient bezüglich Laden und Abfragen der Daten.

Zusätzlich werden die relationalen Datenmodelle der OLTP-Systeme den Anforderungen der multidimensionalen Sichtweise der Anwender wenig gerecht, da sie durch Normalisierung in viele Tabellen eingeteilt sind, die für den Laien schwer ersichtlich sind.

Daher wird bei der Datenmodellierung für das Data Warehouse bzw. das Data Mart der umgekehrte Weg der sogenannten Denormalisierung beschritten, wofür bewusst Redundanzen in Kauf genommen werden. Das bekannteste Schema für multidimensionale Datenmodelle ist das einfache Star-Schema. Dieses besteht aus einer Faktentabelle, die alle auswertbaren Kennzahlen enthält; für jede Dimension, nach der ausgewertet werden kann, wird eine Tabelle angelegt. Die Dimensionstabellen repräsentieren dabei die fachlichen Zusammenhänge, nach der im Data Warehouse multidimensionale Datenanalyse betrieben wird (Wieken, 1999, S. 173).

Die Dimensionen sind beim Star Schema sternförmig um eine Faktentabelle angeordnet und sind in denormalisierter Form gespeichert, das heißt alle Elemente und Attribute einer Dimension werden in einer Dimensionentabelle gespeichert.

Andere Datenmodelle sind das Galaxy-Schema (wird synonym mit dem Multi-Star-Schema verwandt) und das Snowflake-Schema. Beim Galaxy-Schema werden die Kennzahlen mit unterschiedlichen Dimensionen in separaten Faktentabellen zusam-

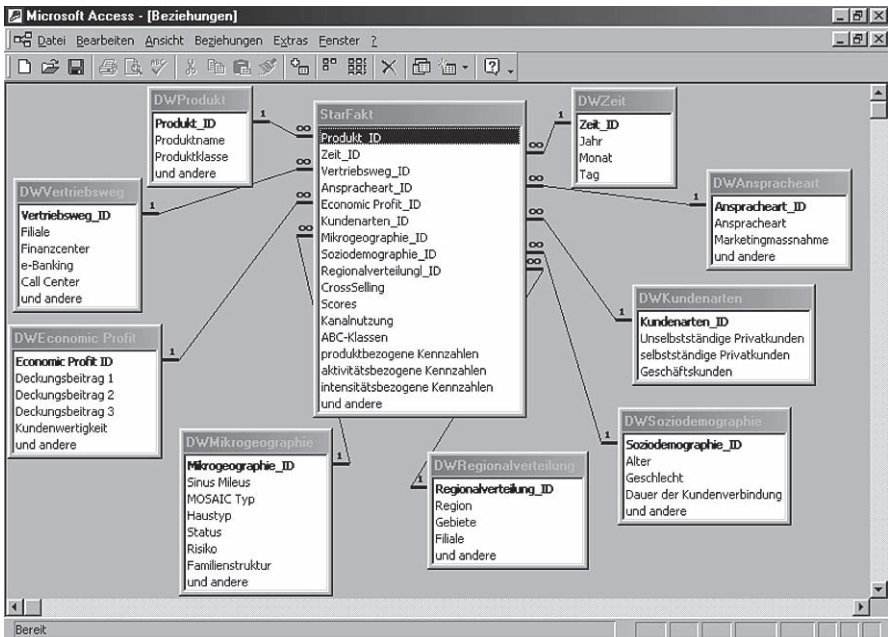


Abbildung 109: Star Schema zur Datenmodellierung

mengefasst, das heißt für jede Dimension eine Fakttable. Dadurch werden mehrere durch gleiche Dimensionen verbundene Star-Schemata dargestellt. Die Dimensionen werden ebenfalls denormalisiert in Dimensionentabellen gespeichert. Im Vergleich zum Star-Schema benötigt es weniger Speicherplatz, weil mehrere Fakttabellen weniger Redundanzen hervorrufen. Allerdings ist die Architektur für den Anwender nicht so übersichtlich. Beide Modelle haben jedoch auch den Nachteil, dass die einzelnen Hierarchien der Dimensionen (das sind die Verdichtungsstufen, nach denen ausgewertet werden kann) nicht explizit ersichtlich sind.

Beim Snowflake-Schema jedoch werden die Dimensionen zusätzlich in verschiedene Tabellen normalisiert gespeichert, die die Hierarchien abbilden. Die Anordnung der Dimensionen und Fakten kann dabei prinzipiell als Star- oder Galaxy-Schema angelegt werden.

Im Vergleich zu den anderen beiden Verfahren benötigt das Snowflake-Schema durch Normalisierung auch weniger Speicherplatz.

In der Praxis können jedoch auch andere Formen existieren.

16.7 Top-Down oder Bottom-Up-Ansatz von Data Warehouses

Grundsätzlich können Data Warehouses Top-Down oder Bottom-Up erstellt werden.

Während beim klassischen Top-Down Ansatz zum Aufbau eines Data Warehouse auf Data Marts verzichtet werden kann, werden beim Bottom-Up Ansatz häufig Data Marts errichtet, die ein zentrales Data Warehouse speisen (Beyer, 2000, S. 55 f.).

Der wesentliche Unterschied liegt darin, dass beim Top-Down Ansatz überwiegend Datenmodelle zentral nach dem Informationsbedarf des Top Managements geschaffen werden und demzufolge das Data Warehouse meist sehr statische Strukturen aufweist, während beim Bottom-Up Ansatz die Bedürfnisse der einzelnen Fachabteilungen in der Informationsbedarfsanalyse zur Datenmodellierung der Data Marts berücksichtigt werden und erst darauf folgend das Data Warehouse errichtet wird. Der Bottom-Up Ansatz führt daher durch die Errichtung von Data Marts zu einem größeren Nutzen für das gesamte Unternehmen und zu einer flexibleren Datenhaltung und -bereitstellung.

Mit Data Marts können auch Vertriebsmitarbeiter, Mitarbeiter im Call-Center und andere arbeiten, um gezielt One-to-One Marketing zu betreiben, womit sich die Bedeutung für das Direktmarketing erheblich steigert. Darüber hinaus führt die Errichtung von Data Marts zu erheblichen Kosten und Zeitvorteilen (Beyer, 2000, S. 57).

Beispiel:

Das Data Warehouse Projekt der DEKA Bank kann in diesem Zusammenhang auch als Data Mart betrachtet werden, da es vorwiegend für die Fachabteilungen Marketing und Vertrieb errichtet wird und kein unternehmensweites Data Warehouse im eigentlichen Sinn darstellt.

16.8 Einsatz im Internet und Intranet

Bei der Nutzung des Data Warehouses bzw. der Data Marts im Internet oder Intranet handelt es sich oft um dynamische Berichterstellungen im Intranet, beispielsweise in der DEKA Bank, und der Nutzung eines Online-OLAP-Würfels in der Deutschen Bank 24. Dies führt zu flexibleren Einsatzmöglichkeiten für die Mitarbeiter, die sich nicht ständig im Unternehmen aufhalten, beispielsweise Vertriebsmitarbeiter. Darüber hinaus entfallen die sonst anfallenden Softwareinstallationen auf den Einzelplatz PCs. Dynamische Abrufe des Data Warehouse im Internet werden durch Technologien wie Java Applets und Servlets ermöglicht, die in die statischen HTML-Seiten eingebunden werden (Wieken, 1999, S. 91).

Die neuen Business Intelligence Tools von Oracle, Actuate, Cognos und andere ermöglichen ebenso eine Erstellung dynamischer Komponenten, die im Internet/Intranet eingesetzt werden. Voraussetzung hierbei ist die Installation auf dem Webserver.

Das webbasierte Data Warehouse oder Data Webhouse wird im Vergleich zu den bisher dargestellten Data Warehouse Systemen um die Clickstream-Daten (Daten des Surfverhaltens) der Besucher auf der Website erweitert. Mit neuen Tools wie Oracle Intelligent Webhouse kann der Clickstream überwacht und analysiert werden, indem auf die Logfiles des Webserver zugriffen wird. Diese Daten werden dazu benutzt, mögliche Kunden zu identifizieren, zu segmentieren und daraufhin in Gruppen einzuteilen, denen mögliche Einnahmen und Kosten zugewiesen werden können. Den Clickstream Daten wird dabei eine weitaus größere Aussagekraft zugewiesen, als den herkömmlichen Datenquellen, weil sie sich näher am Verhalten der Kunden orientieren, das heißt das Kundenverhalten auf der Website kann beinahe zeitgleich über das Internet genau nachvollzogen und bearbeitet werden.

Mit XML (Extensible Markup Language) wurde ein neues Standardformat für den Austausch von Daten im HTML-Format geschaffen. Damit können die Informationen über die Besucher der Website gezielt im Unternehmen über das Intranet und damit auch an die verschiedenen Customer Relationship Instrumente beispielsweise in das Callcenter verteilt werden.

16.9 Auswertungen mit Querying und Reporting

Mit Hilfe von Querying (= SQL-Selektionsabfragen) können erforderliche Daten in Datenbanken nach beliebigen Kriterien, sofern vorhanden, selektiert und eingesehen werden. Dabei handelt es sich um Datenbankabfragen, die in der Regel in SQL geschrieben sind. SQL beinhaltet einen einfachen aber umfangreichen Befehlsschatz zur Datenbankmanipulation. Der notwendige SQL-Code für Selektionsabfragen kann mit den gängigen Tools auch visuell mittels einer grafischen Oberfläche erstellt werden, ohne dass Programmierkenntnisse erforderlich sind (Kirchner, 1998, S. 165). Selektionsabfragen stellen die erste Stufe der Datenanalyse dar, weil die folgenden Verfahren auf der Ergebnismenge von Abfragen aufbauen.

Beispiel:

Mit Hilfe von Abfragen kann beispielsweise der Zusammenhang zwischen sozialem Status und Kreditkartenbesitz betrachtet werden, sofern die Informationen in der Datenbank vorliegen. Die Deutsche Bank 24 kam dabei zu dem Ergebnis, dass bei steigendem Status die Besitzwahrscheinlichkeit für eine Kreditkarte steigt.

Berichte (Reporting) bereiten die Ergebnismenge der Selektionsabfragen grafisch auf. Zusätzlich zu den Daten der Abfrage können auch Formelfelder für Berechnungen und Diagramme eingefügt werden. Berichte können vom Anwender bei entsprechender Berechtigung dynamisch generiert werden. Sie können danach als sogenannte Standardreports auf ein Netzlaufwerk im Unternehmen gespeichert werden, im HTML-Format ins Intranet geladen werden oder per E-Mail im Unternehmen an bestimmte Personen in einem geeigneten Format beispielsweise MS-Excel verteilt werden.

Neue Ansätze ermöglichen durch Technologien wie der Programmiersprache Java und Microsoft Active X Dokumenten dynamische Berichterstellung auf einer Website im Intranet.

Abfragen und Berichte werden im Direktmarketing für Häufigkeitsauswertungen eingesetzt, die Antworten auf Fragen liefern, wie beispielsweise:

- Wie hoch war die Response von Marketing-Maßnahme X?
- Welche Kunden haben sich letzten Monat beschwert?
- Wie viele Kunden haben Produkt A gekauft?
- ...

Sie bilden aber auch die Grundlage für verschiedene Scoringverfahren, wie die ein-dimensionale Kundenqualifizierung und das Kundenportfoliomodell.

16.10 Auswertungen mit OLAP

Mit OLAP (Online Analytical Processing) ändert sich der Blickwinkel auf die zu untersuchenden Daten. Die Datenanalyse erfolgt nicht mehr datensatzorientiert, sondern im multidimensionalen Raum, der auch als Würfel dargestellt wird. Darin sind die Datenobjekte über verschiedene Dimensionen, die eine hierarchische Dimensionsstruktur aufweisen, definiert (Düsing, 1998, S. 296).

Die Anwender können analysierend durch die Daten navigieren, um Trends zu entdecken, Ausnahmen herauszufinden und die zu Grunde liegenden Daten anschauen, um damit Geschäftsentwicklungen ihres Unternehmens zu erkennen. OLAP ist eine Abfragemethode, die Endbenutzern einen mehrdimensionalen schnellen Zugriff und eine benutzerfreundliche interaktive Analyse von Daten ermöglicht.

Die Abbildung 110 zeigt eine Tabelle, in der der Absatz einzelner Produkte (Spalten) an die unterschiedlichen Kunden (Zeilen) dargestellt ist. Die einzelnen Monate eines Jahres lassen sich über die dritte Dimension, die hintereinander liegenden Blätter, analysieren.

In Abbildung 111 sind diese Informationen in einem OLAP-Würfel dargestellt. Er enthält den Absatz eines Unternehmens mit den folgenden Dimensionen:

- Zeit (Monate eines Jahres)
- Region (Verkaufsbezirke)
- Produkt (Sortiment, Warengruppe)

Beispiel:
 Die Deutsche Bank 24 verspricht sich durch Einführung von OLAP-Würfeln für Beratung und Vertrieb innerhalb des Projekts „DataSmart“, dass diese selbstständig Aussagen über die Produktaffinität, Anspracheart, den vom Kunden präferierten Vertriebsweg, Ertrag und Risiko von Marketing-Aktionen und Loyalität der einzelnen Kunden treffen können.

Kunden	Produkt 1	Produkt 2	...	Produkt m
1	5%	39%		
2	23%	4%		
3	42%			
...				
Kunde(n)	<div style="border: 1px solid gray; padding: 5px; width: fit-content; margin: 0 auto;"> Produktsicht, z.B. Mailing </div>			

Kundensicht
z.B. Filiale

Abbildung 110: Produkt- und Kundensicht in der Segmentierung
 Quelle: Nitsche, 2001, S. 136

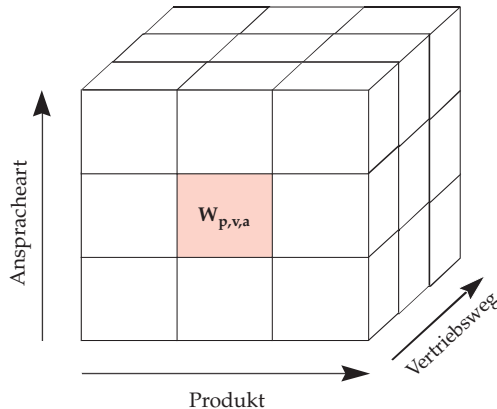


Abbildung 111: OLAP-Würfel

Allgemeines Ziel der multidimensionalen Datenanalyse ist es, durch folgende Operationen die Masse der Daten zu aussagekräftigen Kennzahlen zu verdichten (Jarke, 2000, S. 90).

- **Slice und Dice**

Slicing ist die Auswertung eines OLAP-Würfels durch Festlegung auf eine Kategorie oder Hierarchiestufe einer Dimension. Dadurch ergibt sich bei insgesamt drei Dimensionen eine zweidimensionale Betrachtung auf die anderen Dimensionen, die dann einer Kreuztabelle gleicht. Der Ausschnitt dieses Würfels wird als Slice bezeichnet.

Dice beinhaltet das Austauschen der gegenübergestellten Dimensionen. Der Würfel wird damit gezielt „gedreht“, so dass die sichtbare zweidimensionale Matrix die gewünschten Dimensionen enthält.

Slice und Dice bietet damit die Möglichkeit, beliebig den Blickwinkel auf die zu analysierenden Daten zu ändern. Der OLAP-Würfel (siehe Abbildung 111) könnte durch Slicing zu der Darstellung aller Produkte nach Anspracheart auf das letzte Jahr beschränkt sein. Mit Dicing könnte die Sichtweise zu Vertriebsweg nach Produkten geändert werden.

- **Drill down und Roll up**

Durch Drill down kann die dargestellte Ansicht verfeinert werden, indem die Darstellung in der nächsten niedrigeren Hierarchiestufe einer Dimension erfolgt beispielsweise von Produktgruppen zu einzelnen Produkten. Roll up bezeichnet den umgekehrten Weg.

- **Screening (oder Selection, Filtering)**

Screening betrifft das Setzen eines Filters in einer Dimension für die weitere Analyse, beispielsweise kann der Würfel auf Produkte beschränkt werden, die einen negativen Deckungsbeitrag haben.

- **Scoping**

Scoping beschränkt wie das Screening die Sicht auf die Datenbankobjekte, jedoch erfolgt dies zum Aufbau des zu analysierenden Würfels und damit in einer anderen Stufe der multidimensionalen Datenverarbeitung. Es könnte damit vorab bestimmt werden, dass der Würfel nur Produkte enthält, deren Marktanteil sich im letzten Jahr negativ entwickelt hat.

Beispiel:

Abbildung 112 zeigt ein typisches Beispiel für eine OLAP-Auswertung.

In dem Würfel sind die abgesetzten Mengen eines Unternehmens differenziert nach den drei Kriterien aufgeführt:

- Monate des Jahres
- Verkaufsbezirke des Unternehmens
- Produktgruppen des Unternehmens

In diesen dreidimensionalen Würfel werden die Absatzzahlen als viertes Kriterium hinein gegeben und durch Farbverläufe verdeutlicht (beispielsweise

von dunkelrot für starke Absatzverluste bis dunkelgrün für deutliche Absatzgewinne).

Durch diese grafische Darstellung lassen sich markante Entwicklungen besonders übersichtlich verdeutlichen.

- **Produktmanager**

Der Produktmanager für das Produkt B hat nun die Möglichkeit, sich „eine Scheibe aus dem Würfel herauszuschneiden“ (slice) und damit nur seinen Verantwortungsbereich über alle Monate und alle Verkaufsgebiete zu betrachten. Durch drill down ist ihm eine tiefere Auswertung möglich, so kann er beispielsweise alle unterschiedlichen Packungsgrößen analysieren.

- **Gebietsmanager**

Der Gebietsmanager für das Verkaufsgebiet 5 schneidet sich ebenfalls seine Scheibe aus dem Würfel und analysiert die Verkaufszahlen aller Produkte über alle Monate nur in seinem Gebiet. Eine tiefere Auswertung führt ihn dann zu allen Filialen und Verkaufsmitarbeitern.

- **Finanzleiter**

Der Finanzleiter analysiert die Geldströme über alle Gebiete und Produkte in einem Monat und verfeinert seine Auswertung bis auf die Ebene von Tagen.

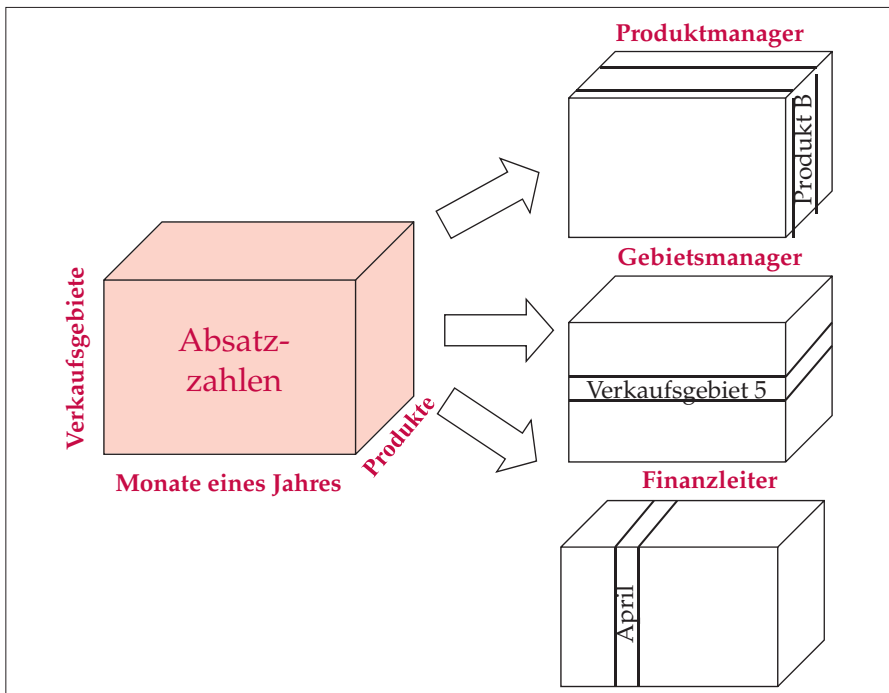


Abbildung 112: OLAP-Auswertung