

1 Abgrenzung und Einordnung

Im Bereich der auswertungsorientierten Informationssysteme gibt es nur wenige Begriffe, die seit den 90er Jahren häufiger und andauernder erwähnt und diskutiert wurden als der des Data-Warehouse-Systems. Viele Zeitungsartikel, Forschungsbeiträge und Produktinformationen propagieren zwar die Notwendigkeit eines Data-Warehouse-Systems, es geht aber selten eindeutig hervor, worin die Charakteristika und der Nutzen eines Data-Warehouse-Systems liegen. Die Verwendung des Begriffes ist derart vielseitig, dass es notwendig ist, nicht nur die Eigenschaften eines Data-Warehouse-Systems aufzuzeigen, sondern auch eine einheitliche Begriffsverwendung im Sprachgebrauch zu erreichen.¹

Die Vielseitigkeit des Data-Warehouse-Begriffes resultiert aus zwei grundlegenden Bereichen, die diesen Begriff geprägt haben: Auf der technischen Seite stehen die Grundlagen der Datenintegrationsmöglichkeiten und Datenbanksysteme, auf der Anwendungsseite finden sich die betriebswirtschaftlichen, wissenschaftlichen und technischen Anforderungen aus einer Nutzungsperspektive. Weiterhin ist der Einfluss der Marktanalysten, Beratungshäuser und Softwarefirmen nicht zu gering zu erachten. Es ist somit ein Muss, diese oft gegensätzlichen Gebiete in diesem Buch gleichermaßen zu betrachten.

Eine Lösung dieses Dualismus von Informatik und Betriebswirtschaft kann nur in der Kombination liegen. Das Data-Warehouse-System wird deshalb als ein System aus Datenbanken und Komponenten gesehen, das aus der technischen Sicht Daten aus verschiedenen Datenquellen integriert und aus der betriebswirtschaftlichen Sicht dem Anwender diese Daten zu Auswertungszwecken zur Verfügung stellt.

Weiterhin soll an dieser Stelle auch angemerkt werden, dass der Begriff »Data-Warehouse-System« zunehmend durch den Begriff »Business Intelligence« ergänzt, überlagert oder oft auch ersetzt wird. Business Intelligence ist als Erweiterung zum Data Warehousing zu sehen, da insbesondere die Anwendungen sowie die anwendungsseitigen Prozesse darunter verstanden werden. Das Data-

1. Leider muss auch nach drei Auflagen dieses Buches festgehalten werden, dass sich an dieser Tatsache nur wenig geändert hat.

Warehouse-System dient weiterhin der Integration eines zentralen, auswertungsorientierten Datenbestandes.

In Abschnitt 1.1 werden für das weitere Verständnis wichtige Definitionen und Abgrenzungen zu verwandten Bereichen gegeben. In Abschnitt 1.2 wird die lange Historie des Themengebietes sowohl von Anwendungs- als auch von Datenbankseite skizziert. Nachfolgend wird der Begriff »Business Intelligence« aufgegriffen (Abschnitt 1.3), um sowohl aktuelle Tendenzen als auch die Fokussierung dieses Buches herauszuarbeiten. Im daran anschließenden Abschnitt folgt ein Überblick über die Vielfältigkeit der möglichen Einsatzbereiche eines Data-Warehouse-Systems. Abschnitt 1.5 umreißt abschließend den Inhalt des Buches und schafft mit einem speziellen Anwendungsbeispiel die Basis für ein durchgängiges Beispiel.

1.1 Begriffliche Einordnung

Zu den drei konventionellen Produktionsfaktoren Boden, Arbeit und Kapital wird immer häufiger die Information als vierte Säule hinzugenommen. Informationen basieren auf Daten, die entweder aus einem Unternehmen selbst stammen oder extern zugekauft werden. Die Tatsache, dass Daten eine besondere Bedeutung zukommt, ist aber nicht allein im betriebswirtschaftlichen Kontext zu finden, sondern gilt ebenso für statistische, wissenschaftliche oder technische Anwendungen.

Verschiedenen Informationssystemen ist gemein, dass Daten erfasst und verwaltet werden. Daten in einem Datenbanksystem zu erfassen, ist an sich nichts Neues. In jedem Unternehmen werden Personaldaten eingegeben oder Verkäufe durch Scannerkassen erfasst. Die Verarbeitung und Verwaltung der Daten geschieht in der Regel aber autonom unter Verantwortung der jeweiligen Abteilung. Interessant wird es erst, Daten aus autonomen Quellen zu vereinen. Dieser Vorgang ist besonders schwierig, wenn heterogene Daten unterschiedlichster Qualität, in verschiedenen Datenformaten, in heterogenen Datenmodellen und Datenbanksystemen gehalten werden.

Zur Vollständigkeit soll an dieser Stelle ein Exkurs zu verschiedenen Integrationsmöglichkeiten gegeben werden. Grundsätzlich wird eine *Backend-Integration* von einer *Frontend-Integration* unterschieden. Während bei der Frontend-Integration die Daten und Applikationen aus verschiedenen Systemen nur durch eine gemeinsame Oberfläche integriert werden (Stichwort: Portal, [Lint00b]), werden bei der Backend-Integration entweder die Daten physisch oder virtuell integriert oder Applikationen über eine gemeinsame Schnittstelle zusammengebracht (Stichwort: Enterprise Application Integration (EAI), [Lint00a]). Das Data-Warehouse-System kann in dieser Klassifikation der Backend-Integration zugeordnet werden. Eine pauschale Bewertung kann leider nicht erfolgen, da die Vorteile jeder Integrationsart von dem jeweiligen Zweck und der Strategie abhängen. Anzustreben ist grundsätzlich eine möglichst frühzeitige Integration der Daten und Anwendungen, was jedoch immer einen mitunter beträchtlichen Aufwand nach sich zieht.

Ein Data-Warehouse-System ist aber nicht nur von diesem integrativen Aspekt geprägt, sondern zusätzlich vom Aspekt der Auswertung. Die Verwendung von Daten in operativen Anwendungen war lange Zeit geprägt von einer *transaktionalen* Verarbeitung mit vielen kurzen Lese- und Schreiboperationen. Im Gegensatz dazu steht beim Data-Warehouse-System eine eher vergleichende oder auswertende Verwendung der Daten im Vordergrund, bei der auf große Datenmengen lesend zugegriffen wird.

Einige Fragen müssen jetzt erlaubt sein: Was ist eigentlich ein Data-Warehouse-System und was zeichnet es aus? Was bedeutet der Begriff Data-Warehouse-System? Ist ein Data-Warehouse-System eine integrierte Datenbank oder eine Datenbasis zu Auswertungszwecken? Wo liegen die Gemeinsamkeiten der Einsatzbereiche? Die vielen Interpretationsmöglichkeiten machen es notwendig, einige Begriffe zu definieren.

1.1.1 Definitionen

Die Tatsache, dass dieses Themengebiet sowohl von der Anwendungsseite als auch der Informatikseite durch eigene Fachtermini geprägt ist, impliziert ein unterschiedliches Begriffsverständnis. Verschiedene Normungsgremien versuchen, diese Begriffe zu standardisieren. Diese Bestrebungen waren aber bislang wenig erfolgreich.

Eine der ersten Definitionen aus dem Umfeld des Begriffes Data-Warehouse-System wurde von Inmon geprägt:

»A data warehouse is a subject oriented, integrated, non-volatile, and time variant collection of data in support of management's decisions.« [Inmo96]

Ein »Data Warehouse« hat seiner Ansicht nach also vier Eigenschaften, die alle der Entscheidungsunterstützung dienen. Die Eigenschaften sollen hier kurz skizziert werden:

- *Fachorientierung* (engl. subject orientation):
Der Zweck der Datenbasis liegt nicht mehr auf der Erfüllung einer Aufgabe wie z.B. der Lohn- und Gehaltsabrechnung, sondern in der Möglichkeit, ganze Themenbereiche wie Produkte und Kunden auszuwerten.
- *Integrierte Datenbasis* (engl. integration):
Die Datenverarbeitung findet auf integrierten Daten aus mehreren Datenbanken statt.
- *Nicht flüchtige Datenbasis* (engl. non-volatile):
Die Datenbasis ist als stabil zu betrachten. Daten, die einmal in das Data-Warehouse-System eingebracht wurden, werden nicht mehr entfernt oder geändert.

■ *Historische Daten* (engl. time variance):

Die Verarbeitung der Daten ist so angelegt, dass vor allem Vergleiche über die Zeit stattfinden. Es ist dazu unumgänglich, Daten über einen längeren Zeitraum zu halten.

Diese Definition ist einerseits nicht aussagekräftig genug, um sie in der Praxis oder der Theorie verwenden zu können, andererseits ist sie so einschränkend, dass viele Anwendungsgebiete und Ansätze herausfallen. Eine neue Definition ist notwendig, um dieses Manko zu überwinden:

»Ein Data-Warehouse-System ist ein physisches Informationssystem, das eine integrierte Sicht auf beliebige Daten zu Auswertungszwecken ermöglicht.«

Aus der vermeintlich trivialen Forderung nach einer »physischen Integration zu Auswertungszwecken« entstehen Fragestellungen wie die Integration von Schemata und Daten aus unterschiedlichen Quellen. Diese Thematik ist zwar u.a. auch in *föderierten Datenbanksystemen* [Conr97] anzutreffen. Im Unterschied zu diesen bestehen zusätzliche Forderungen nach der *physischen* Integration und dem *Auswertungsaspekt*, der Kenntnisse und Denkweisen des Nutzers mit einbezieht.

Häufig wird diese Anforderung durch ein *multidimensionales Modell* [KRRT98] erreicht, das die Denkweise des Anwenders in *Dimensionen* und *Klassifikationshierarchien* widerspiegelt. Das multidimensionale Modell stellt im Gegensatz zu anderen Modellen besondere Strukturen und Auswertungsmöglichkeiten zur Verfügung, die schon bei der Modellierung einen Auswertungskontext schaffen.

Ein in diesem Zusammenhang wichtiger Treiber ist das *Online Analytical Processing* (OLAP, [CoCS93]), das eine explorative, interaktive Datenauswertung auf der Grundlage des konzeptuellen multidimensionalen Datenmodells darstellt. Weiterhin fällt oft das Stichwort *Data Mining*, darunter versteht man eine Suche nach unbekannten Mustern oder Beziehungen im Datenbestand des Data-Warehouse-Systems (Abschnitt 3.5.3). Obwohl ein Data-Warehouse-System keine notwendige Voraussetzung für Data Mining darstellt, kann ein Data-Warehouse-System als Ausgangspunkt für Data Mining verwendet werden.

Ein weiterer Unterschied eines Data-Warehouse-Systems gegenüber einem anderen Informationssystem liegt darin, dass die Daten in der Regel *nicht modifiziert* werden. Daten, die einmal in das Data-Warehouse-System übernommen wurden, dürfen nicht mehr verändert werden. Es können aber neue Daten in das Data-Warehouse-System aufgenommen werden, ohne die bereits vorhandenen zu überschreiben.

Da eine Datenbankinstanz diese Eigenschaften in der Regel alleine nicht zur Verfügung stellen kann, werden mehrere Datenbanken mit spezifischen Verwendungszwecken für ein Data-Warehouse-System benötigt. Weiterhin umfasst das Data-Warehouse-System alle für die Integration und Auswertung notwendigen

Komponenten. Besonders hervorzuheben sind die Komponenten der *Datenintegration*, die für die Integration der Daten notwendig sind, die Komponenten zur *Auswertung* und die Komponenten zur Verwaltung des Data-Warehouse-Systems.

Der *Data-Warehouse-Prozess*, auch *Data Warehousing* genannt, beschreibt den *dynamischen* Vorgang, angefangen beim *Datenbeschaffungsprozess* über das Speichern bis zur Auswertung der Daten, d.h. den Fluss und die Verarbeitung der Daten aus den Datenquellen bis zum Auswertungsergebnis beim Anwender. Ein Data-Warehouse-System ist also mehr als die Summe seiner Komponenten, d.h., erst mit dem Data-Warehouse-Prozess kann es seine Aufgaben erfüllen. Für die exakten Definitionen sei auf das Glossar in Anhang B verwiesen.

1.1.2 Abgrenzung von transaktionalen Systemen

Die unterschiedlichen Anforderungen rechtfertigen es, von Gegensätzlichkeiten zwischen der *Datenintegration* und der *Auswertungsfunktion* zu sprechen. Die Vereinigung von Daten aus diversen Datenquellen, wie sie in vielen Unternehmen vorhanden sind, macht es für eine effiziente Verarbeitung notwendig, sich nicht nur um die Integration der Daten zu kümmern, sondern sie auch genau in der Form bereitzustellen, die der Anwender fordert. Die Frage nach den Anwenderwünschen wird hier deutlich: Für eine Auswertung sind nicht alle möglichen Daten notwendig, sondern nur genau die Daten seines Entscheidungsgebiets, für das er in adäquater Zeit Informationen benötigt. Der Anwender braucht also ein Informationssystem, das mehrere Datenquellen vereinigt und einen expliziten Bezug zum jeweiligen Anwendungsfall hat.

Im Gegensatz dazu stehen die transaktionalen Systeme, die oft auch als *Online Transactional Processing* (OLTP) bezeichnet werden. Es gibt Unterscheidungsmerkmale, die sich in die Bereiche *Anfragen*, *Daten* und *Anwender* klassifizieren lassen. Natürlich gibt es auch Mischformen, wie z.B. der Einsatz eines Data-Warehouse-Systems in einem operativen Produktionsprozess (Stichwort »Embedded Analytics« oder »Operational BI«), in der die OLTP- mit der OLAP-Welt verschwimmt.

Anfragen

Die Charakteristika von Anfragen haben einen großen Einfluss auf die Anfrageverarbeitung und Speicherform von Daten (Tab. 1–1). Der Fokus unterscheidet sich grundlegend darin, ob Daten in transaktionalen oder auswertungsorientierten Systemen verwaltet werden. Erstere lesen, modifizieren oder löschen Daten in kurzen und einfachen Transaktionen. Auswertungsorientierte Systeme hingegen gewinnen häufig Informationen aus den Daten, indem lange Lesetransaktionen in komplexen Anfragen erfolgen. Eine Anfrage im transaktionalen Betrieb betrifft meist nur wenige Datensätze in einem anfrageflexiblen Datenmodell, d.h., keine

Anfrage hat eine modellseitige Bevorzugung. Eine auswertende Verarbeitung betrifft Millionen von Datensätzen und findet oft ad hoc nach Benutzerwünschen in einem auswertungsorientierten Modell statt.

Anfragen	transaktional	analytisch
Fokus	Lesen, Schreiben, Modifizieren, Löschen	Lesen, periodisches Hinzufügen
Transaktionsdauer und -typ	kurze Lese-/Schreibtransaktionen	lange Lesetransaktionen
Anfragestruktur	einfach strukturiert	komplex
Datenvolumen einer Anfrage	wenige Datensätze	viele Datensätze
Datenmodell	anfrageflexibles Datenmodell	auswertungsbezogenes Datenmodell

Tab. 1-1 Gegenüberstellung der Anfragecharakteristika von transaktionalen und analytischen Anwendungen

Daten

Ein Beispiel für einen transaktionalen Betrieb findet sich in der Personalabteilung. Anwender verwalten Daten über die Angestellten in einer Datenbank. Die Daten werden aus einer Produktionsumgebung durch ein Datenbanksystem verwaltet, um darauf transaktionale ein-/ausgabeorientierte Anwendungen auszuführen. Die Daten eines Data-Warehouse-Systems stammen ebenso aus der Produktionsumgebung, aber physisch aus einer oder mehreren operativen Datenbanken (Tab. 1-2). Daten in einem Data-Warehouse-System sind also abgeleitete Daten. In der transaktionalen Verarbeitung stehen Aufgaben wie die Konsistenzsicherung einzelner Transaktionen der Sammlung von Daten im Data-Warehouse-System gegenüber. Die Daten sind im transaktionalen Betrieb meist zeitaktuell, nicht abgeleitet, autonom, aus einer Datenquelle und dynamisch, d.h. von vielen Schreibvorgängen und ständigen Modifikationen betroffen. Die Auswertungsorientierung hingegen erfordert Daten, die konsolidiert, integriert, stabil und häufig aggregiert sind. Durch diese Stabilität der Daten und die Vereinigung mehrerer Datenquellen wächst das Datenvolumen von Mega- oder Gigabyte in transaktionalen Anwendungen auf Datenvolumina bis in den hohen Terabyte-Bereich bei analytischen Anwendungen an. Die Datenproblematik macht es auch notwendig, dass das Design der Datenbank sich dem Einsatzzweck anpasst, d.h., das Datenbankschema wandelt sich von der Anwendungsneutralität beim transaktionalen Bereich mit einer Anfrageflexibilität zu einer Auswertungsorientierung mit vorgedachten Auswertungspfaden. Weiterhin impliziert die Anwendung andere Zugriffsstrukturen: Im transaktionalen Fall finden weitgehend Einzeltupelzugriffe statt. Dagegen sind im auswertungsorientierten Fall häufig Suchläufe über den gesamten Datenbestand notwendig.

Daten	transaktional	analytisch
Datenquellen	meist eine	mehrere
Eigenschaften	nicht abgeleitet, zeitaktuell, autonom, dynamisch	abgeleitet, konsolidiert, historisiert, integriert, stabil
Datenvolumen	Megabyte – Gigabyte	Gigabyte – Terabyte
Zugriffe	Einzel tupelzugriff	Bereichsanfragen

Tab. 1–2 Gegenüberstellung der Datencharakteristika von transaktionalen und analytischen Anwendungen

Anwender

Der Anwender eines Data-Warehouse-Systems ist typischerweise nicht mehr der Sachbearbeiter, der Daten erfasst oder abfragt, oder die Anwender beispielsweise eines Flugbuchungssystems, sondern der Manager, Controller oder Analyst, der Daten in verdichteter Form zur Entscheidungsunterstützung benötigt (Tab. 1–3). Aus Organisationssicht reduziert sich die Anzahl der Anwender von Tausenden bei transaktionalen Systemen auf wenige Hundert im analytisch geprägten System. In letzter Zeit werden durch neue Anwendungsgebiete mit ständig wachsender Verarbeitungsgeschwindigkeit aber auch höhere Nutzerzahlen erreicht.

Eine explizite Nebenläufigkeit mit einem ausgefeilten Transaktionskonzept wird durch den lesenden Zugriff nicht mehr benötigt. In beiden Anwendungsfällen wird eine kurze Antwortzeit erwartet; bei einer Anwendung, die sehr große Datenmengen in komplexen Anfragen verwendet, kann eine Forderung nach Antwortzeiten im Sekunden- bis Minutenbereich zu einem erheblichen Problem werden.

Anwender	transaktional	analytisch
Anwendertyp	Ein-/Ausgabe durch Sachbearbeiter	Auswertungen durch Manager, Controller, Analysten
Anwenderzahl	sehr viele	wenige (bis einige Hundert)
Antwortzeit	ms – s	s – min

Tab. 1–3 Gegenüberstellung der Anwendercharakteristika von transaktionalen und analytischen Anwendungen

1.2 Historie des Themenbereichs

Die Data-Warehouse-Idee ist nicht neu: Entscheidungsträgern in verschiedenen Funktionsbereichen und auf verschiedenen Hierarchieebenen eines Unternehmens sollen im Moment des Informationsbedarfs alle Informationen zur Verfügung stehen, die sie benötigen. Die Informationsbereitstellung soll zeitnah, fehler-

frei, flexibel, ergonomisch, effizient, effektiv und inspirativ erfolgen. Die Bezeichnung *Managementinformationssystem* (MIS) wurde bereits Ende der sechziger Jahre geprägt und ist seither unter wechselnden Bezeichnungen Gegenstand intensiver Bemühungen in Wissenschaft und Praxis. Die Begriffe *Managementinformationssystem* (MIS), *Executive Information System* (EIS), *Führungsinformationssystem* (FIS), *Chefinformationssystem* (CIS), *Entscheidungsunterstützungssystem* (EUS), *Decision Support System* (DSS) usw. werden hier als weitgehend synonym aufgefasst. Jedes System dieser Kategorie bekommt sein individuelles Erscheinungsbild erst vor dem Hintergrund der spezifischen Anforderungen in Unternehmen. Die nuancierten Abgrenzungen, die teilweise angestellt werden, sind daher überflüssig. Sie tragen eher zur Verwirrung als zur Klärung bei, was durchaus auch die Intention mancher kreativer »Wortgestalter« sein mag [HiMo95a]. Fehlende Voraussetzungen, wie z.B.

- schnelle und flächendeckende Kommunikationstechnologien,
- grafische Benutzeroberflächen,
- ausreichende, kostengünstige und schnelle Datenspeicher,
- kostengünstige und leistungsfähige Prozessoren,
- große Datenbasen durch integrierte operative Systeme,

die die MIS-Ansätze der 60er-, 70er- und 80er-Jahre scheitern ließen, sind heute erfüllt.

Mangelte es früher neben der technischen Infrastruktur vor allem daran, dass keine ausreichende, elektronisch verfügbare und skalierbare Datenbasis vorhanden war, geht es in den meisten Unternehmen heute vor allem um die Frage, das vorhandene Potenzial zu erschließen, möglichst bevor es einem der Wettbewerber gelingt. Das Data Warehousing der 90er Jahre zielte in erster Linie darauf ab, alle entscheidungsrelevanten Informationen verfügbar zu machen, die bereits an irgendeiner Stelle im Unternehmen elektronisch gespeichert sind. Nicht zuletzt durch die zunehmende Verbreitung der geschäftsprozessorientierten Transaktionssysteme (z.B. SAP ERP) ist in Unternehmen ein großes Volumen an potenziell entscheidungsrelevanten Informationen vorhanden. Idealerweise sollen in einem Data-Warehouse-System darüber hinaus auch Informationen aus externen Informationssystemen integriert werden.

Was sich im Laufe der MIS-Bemühungen der 70er Jahre zunächst als Utopie abzeichnete, nämlich das Konzept einer völligen Integration in eine EDV-Lösung mit Zugriff auf eine Datenbank, erhält durch den Fortschritt in der Informationstechnologie im Gewand des Data Warehousing eine Renaissance. Entscheidender Unterschied dabei ist allerdings, dass es sich beim Data-Warehouse-System erstens um redundant gehaltene Daten handelt, die von den Quellsystemen losgelöst sind, und zweitens, dass es sich nur um einen Teil der Daten handelt, der dem jeweiligen Auswertungszweck dient. Auf dieser Datenbasis setzen dann funktionsspezifische und benutzeradäquat gestaltete Managementinformations-

systeme aller Facetten bzw. analytische Informationssysteme aus anderen betrieblichen Bereichen auf.

Neben der Einordnung zu Informationssystemen ist es ebenso schwierig, ein Data-Warehouse-System von anderen Ansätzen abzugrenzen. Die Integration von Daten aus vielen heterogenen, autonomen und verteilten Quellsystemen in einem Data-Warehouse-System kann mit der Integration von Daten in dem Gebiet der *Mehrrechnerdatenbanksysteme* verglichen werden. Eine Unterscheidung eines Data-Warehouse-Systems von anderen Datenbankansätzen kann nach der *Homogenität*, der *Kopplung* oder der *räumlichen Verteilung* der Datenbanksysteme getroffen werden [Rahm02]. Aus diesen Kriterien werden dann die unterschiedlichen Datenbanksystemansätze der *parallelen*, *verteilten* oder *föderierten* Datenbanksysteme abgeleitet. Dem Data-Warehouse-Ansatz am nächsten kommt der föderierte Ansatz [Conr97], da dort ein neues konzeptuelles Schema gebildet wird, die Quellsysteme autonom bleiben und das originäre konzeptuelle Schema beibehalten wird. Die Unterschiede zum Data-Warehouse-System liegen darin, dass das konzeptuelle Schema einem speziellen Auswertungszweck dient, die Daten redundant vorgehalten werden und kein schreibender Zugriff auf die Quellsysteme gefordert wird. Auf Ebene der physischen Verarbeitung der Daten liegen die Vorfahren von Data-Warehouse-Systemen im Bereich der *Statistical and Scientific Databases* (SSDB, [Mich91]), die ihren Schwerpunkt auf der Masendatenverarbeitung zur statistischen Auswertung haben.

1.3 Einordnung und Abgrenzung von Business Intelligence

Ausgehend von der Diskussion der Historie kann und muss nach 20 Jahren des Begriffs »Data-Warehouse-System« die Frage gestellt werden, ob nicht die Bezeichnung »Business Intelligence« geeigneter wäre.

Um darauf eine eindeutige Antwort zu geben, bedarf es auch hierbei einer einheitlichen und präzisen Definition. Stattdessen hat sich auch hier eine Vielfalt von Definitionen, Beschreibungen und Vorstellungen breit gemacht, die von unterschiedlichen Interessengruppen (Marktforscher, Wissenschaftler, Hersteller) geprägt wird. Auch in diesem Umfeld existieren bereits einige Konsolidierungsansätze wie z.B. in [KeMU07]. Dort wird definiert:

»Unter Business Intelligence (BI) wird ein integrierter, unternehmensspezifischer, IT-basierter Gesamtansatz zur betrieblichen Entscheidungsunterstützung verstanden.

- BI-Werkzeuge dienen ausschließlich der Entwicklung von BI-Anwendungen.
- BI-Anwendungssysteme bilden Teilaspekte des BI-Gesamtansatzes ab.«

([KeMU07])

Aus diesen und weiteren Ansätzen können zwei Tendenzen gegenüber dem Begriff »Data-Warehouse-System« abgeleitet werden:

- Ausweitung des Themas Integration – neben Daten werden auch Strategien, Prozesse, Anwendungen und Technologien integriert.
- Ausweitung des Themas Auswertung – neben der Auswertung der Daten wird die Erzeugung von Wissen über Status, Potenziale und Perspektiven im Geschäftsumfeld gefordert.

Damit soll für dieses Buch als Abgrenzung festgehalten werden, dass das Data-Warehouse-System mit seiner Definition, Architektur und den verwendeten Mechanismen weiter Bestand hat. Das Data-Warehouse-System kann somit als Teil von Business Intelligence gesehen werden. Erst die Erweiterung der Integrationsmöglichkeiten und die Hinzunahme des Wissensmanagements führt zum Business Intelligence. Ein wesentlicher Aspekt des Business Intelligence ist Ausrichtung am betriebswirtschaftlichen Anwendungsfeld, indem es den analytischen Prozess, der Unternehmens- und Wettbewerbsdaten in handlungsgerechtes Wissen für die Unternehmenssteuerung überführt, darstellt. Dadurch kann Data Warehousing in die Business-Intelligence-Organisation eingebettet werden. In einigen Firmen wird dem bereits Rechnung getragen, indem Data Warehouse Competence Center zu BI Competence Center (BICC) (Abschnitt 11.2) wurden.

1.4 Verwendung von Data-Warehouse-Systemen

Nahezu überall, wo Daten gespeichert werden, entsteht auch der Wunsch, diese auswerten zu können. Die Anwendungsgebiete sind entsprechend weit gestreut und reichen von betriebswirtschaftlichen Aufgabenstellungen über wissenschaftliche Auswertungen bis hin zu technischen Belangen. Kapitel 13 dieses Buches bietet einen Einblick in eine Auswahl von Data-Warehouse-Projekten und gibt somit einen detaillierteren Eindruck über Einsatzpotenziale von Data-Warehouse-Systemen.

Im Folgenden werden Anwendungsfälle und verschiedene Anwendungsbereiche unterschieden. Die häufigsten Einsatzfelder finden sich im betriebswirtschaftlichen Bereich.

1.4.1 Anwendungsfälle

Die Motivation zum Aufbauen und Betreiben eines Data-Warehouse-Systems kommt aus dem Bedürfnis heraus, Daten aus operativen Systemen aufzubereiten und auszuwerten, um dadurch Erkenntnisse zu erlangen. Diese Erkenntnisse sind oft darauf ausgerichtet, dem Unternehmen oder der Organisation einen Wettbewerbsvorteil zu verschaffen. Es kann auch durch externe Vorgaben ein Zwang bestehen, bestimmte Datenströme im Sinne eines Data-Warehouse-Systems

homogen und integriert aufzubereiten und anschließend zu überwachen, um beispielsweise Muster in Überweisungsdaten zu identifizieren und so Anti-Geldwäsche-Auflagen zu erfüllen. Methodisch lassen sich die Anwendungsfälle in Kategorien einordnen. In den letzten Jahren jedoch ist eine Konvergenz der Anwendungsarten zu beobachten, die eine klare Abgrenzung mittlerweile in vielen Fällen unmöglich macht. Gerade die Weiterentwicklung von Weboberflächen hat dazu geführt, dass beispielsweise vormals rein informationsorientierte Dashboards Möglichkeiten anbieten, Auswertungen durchzuführen und aggregierte Kennzahlen so weit aufzubrechen und zu kombinieren, dass OLAP-ähnliche Auswertungen durchgeführt werden und die reine Informationsversorgung mit wenigen Arbeitsschritten zu einer Auswertung wird.

Informationsorientierte Anwendungsfälle

Den Schwerpunkt bildet hier die Versorgung von Informationsempfängern mit statischer Information – die Informationen sind vordefiniert und werden meist in einer festen Frequenz zusammengestellt und geliefert. Als Werkzeuge kommen hier Standardberichte (oft auch nach wie vor als PDF- oder Word-Dokument bzw. in Papierform) zum Einsatz. Mit der zunehmenden Verbreitung von Webtechnologie und der damit verbundenen Integration der Data-Warehouse-Auswertungsschicht in z.B. das Firmen-Intranet werden hier auch vermehrt Cockpits und Dashboards verwendet, die Unternehmenskennzahlen anzeigen. Beispiele für informationsorientierte Anwendungsfälle sind wöchentliche Berichte zu Umsatz und Lagerbeständen oder ein Controlling-Cockpit eines Bereichsleiters, dem die Vertriebsergebnisse des laufenden Monats gezeigt werden.

Planungsorientierte Anwendungsfälle

Bei diesen Anwendungen handelt es sich um einen Sonderfall der informationsorientierten Anwendung. Zusätzlich zu den Istdaten werden hier für Planungs- und Überwachungszwecke die Planwerte für definierte Unternehmenskennzahlen vorgehalten und verglichen. Mithilfe derselben Werkzeuge (Standardbericht, Cockpit/Dashboard) kann so der Erfüllungsgrad überwacht werden und beispielsweise mittels Ampeln oder Tachometern angezeigt werden. Auch können Warnmechanismen manuell (durch Anzeige) implementiert werden. Die Überwachung von Budgets ist hier als eines von vielen Beispielen zu nennen.

Plan- und Istinformationen sind untrennbar miteinander verzahnte Managementaspekte. Damit sind beide unverzichtbare Bestandteile analytischer Informationssysteme. Ohne dass vorher Ziele in Plangrößen konkretisiert wurden, ist später eine steuerungseffektive Beurteilung der Istsituation nicht möglich. Über die Leistung in einem Betrachtungszeitraum lassen sich führungsrelevante Beurteilungen und darauf aufbauende Steuerungsmaßnahmen nur durch den Vergleich von Plan- und Istinformationen treffen. Es gilt die Aussage: Planung ohne

Kontrolle ist sinnlos, und Kontrolle ohne Planung ist unmöglich [Wild81]. Beispielsweise ist der führungsrelevante Informationsgehalt durch das Aufzeigen eines Abwärtstrends in den Istzahlen für sich alleine relativ gering. Eine unmittelbare fundierte Bewertung ist nicht möglich, ohne den Sachverhalt vorher im Planungsprozess vor dem Hintergrund der unternehmensspezifischen Prämissen zu durchdringen, einerseits durch die Zielplanung und andererseits durch die Maßnahmenplanung.

Hieraus leitet sich die Forderung ab, dass in einem Data-Warehouse-System, das die Datenbasis für ein Managementinformationssystem bildet, für führungsrelevante Istzahlen auch Planzahlen gespeichert werden können. Planzahlen müssen dazu im gleichen Bereich generiert werden, in das auch die Istzahlen aus den operativen Systemen extrahiert werden. Dies ist effizient zu realisieren, wenn Planzahlen direkt während des Planungsprozesses in das Data-Warehouse-System geschrieben werden können.² Neben der Sicherung der Datenqualität durch die Verwendung und Pflege eines einzigen Datenmodells besteht dadurch der Vorteil, dass alle am Planungsprozess beteiligten Personen Zugriff auf den jeweils aktuellen Planungsstand haben. Zudem können mehrere Planungsvarianten top-down und/oder bottom-up parallel entwickelt, analysiert und diskutiert werden.

Data-Warehouse-basierte Planungssysteme verbessern also vor allem die Koordination des im Gegenstromverfahren ablaufenden Planungsprozesses unabhängig vom Fokus der Planung. Gleichzeitig wird die Datenqualität hinsichtlich der Kongruenz von Plan- und Istzahlen sichergestellt. Die konkrete Ausgestaltung von Informationssystemen für bestimmte Arten von Planungen (z.B. Absatz- oder Kostenstellenplanung) orientiert sich an Abschnitt 1.4.4.

Im Gegensatz zu auswertungsorientierten Anwendungen benötigen planungsorientierte Anwendungen auch Schreibzugriffe aus Anwendersicht auf ein Data-Warehouse-System. Dies ist notwendig, da der Prozess der Unternehmensplanung in der Regel nicht über ein lokales operatives Informationssystem durchgeführt wird, sondern den integrierten Datenbestand des Data-Warehouse-Systems verwendet. Bei auswertungsorientierten Anwendungen arbeiten die Nutzer meist völlig unabhängig voneinander mit dem System. Demgegenüber besteht zwischen den Nutzern bei planungsorientierten Anwendungen eine Abhängigkeit in Form des Planungsprozesses.

Die Unternehmensplanung als systematische Zukunftsgestaltung des Unternehmens ist kein klassisches Anwendungsgebiet für OLAP und Data-Warehouse-System, da diese Technologien anfangs eher für die Auswertung von vergangenheitsorientierten Daten konzipiert waren. Bedingt durch Verbesserungen und Erweiterungen der OLAP-Produkte eignen sich viele von ihnen inzwischen allerdings auch als Planungsinstrumente.

2. Hier sei bemerkt, dass ein direktes Aufnehmen von Daten in das Data-Warehouse-System ohne den Datenbeschaffungsprozess problematisch werden kann, aber aus Performanzgründen oft gemacht wird (siehe auch Abschnitt 7.2.3).

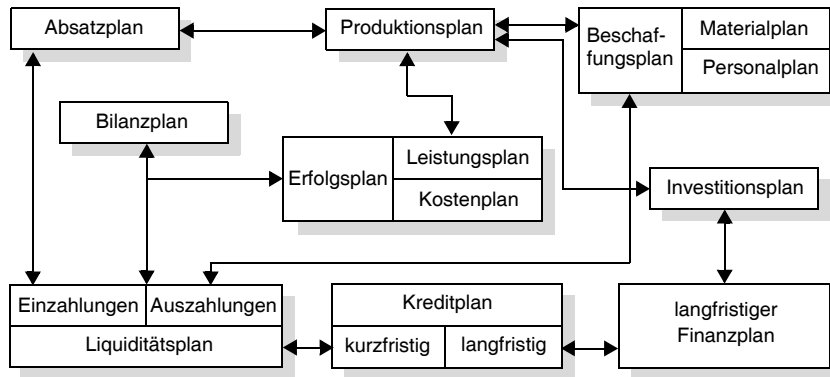


Abb. 1–1 Integrierte Unternehmensplanung [Arbe70]

Abbildung 1–1 zeigt die vereinfachte Darstellung eines integrierten Planungsrahmens mit den Interdependenzen zwischen den unterschiedlichen Teilplänen. Die relevanten Teilpläne sind aus finanzieller Sicht Absatz-, Investitions-, Beschaffungs- und Personalplan. Der Produktionsplan kann in der Regel für die Finanzplanung vernachlässigt werden, da er nur unternehmensinterne Wertbewegungen abbildet. Die Exaktheit der Pläne ist ausschlaggebend für die Qualität der Finanzplanung, die das letzte Glied im sukzessiven Planungsprozess darstellt und somit auf die vorlaufenden Pläne angewiesen ist.

Auswertungsorientierte Anwendungsfälle

Im Gegensatz zu den informationsorientierten und planungsorientierten Anwendungsfällen wird eine auswertungsorientierte Verwendung meist auf Anfrage durchgeführt. Mithilfe von angepassten oder eigens erstellten Ad-hoc-Berichten werden komplexe Anfragen beantwortet, die sich oft auf Informationen aus verschiedenen Datenquellen und Funktionsbereichen beziehen. Auch kommen hier Werkzeuge aus dem Bereich Online Analytical Processing (OLAP) zum Einsatz. Ein Beispiel eines solchen Anwendungsfalls aus der Versicherungswirtschaft ist die Auswertung der eingereichten Versicherungsschäden in Abhängigkeit von Schadensart, Alter und Geografie, um eine aufgefallene Häufung von Schadensmeldungen zu untersuchen.

Im Folgenden werden einige ausgewählte Anwendungen vorgestellt.

■ Erlös-, Marketing- oder Vertriebs-Controlling:

Der Absatz von Gütern und Dienstleistungen ist charakterisiert durch eine große Vielfalt möglicher Vorgehensweisen. Allein eine Entscheidungsmatrix aus einer regionalen und variantenorientierten Marktsegmentierung, die mit entsprechenden kommunikativen und physischen Distributionsaufwendungen gekoppelt ist, generiert ein hohes Datenvolumen an Planungs-, Kontroll- und Hochrechnungsinformationen.

Nach Ablauf einer Betrachtungsperiode bleibt oft unklar, ob der mit dem Marketing- und Vertriebsapparat erzielte Erfolg nicht größer gewesen wäre, wenn z.B. mehr Geld für Werbung und/oder sonstige Marketingmaßnahmen ausgegeben worden wäre. Vielleicht würden sogar ein kleinerer Aufwand für Werbung und ein kleinerer Vertriebsapparat mehr Gewinn bringen. Das Gleiche gilt für die Fragen »Sortiment ausweiten oder verkleinern«, »Zielgruppen breit oder fokussiert wählen«, »Absatzgebiete global ausweiten oder regional begrenzen« sowie »Zahl der Absatzwege verringern oder ausweiten«. Die Informationsversorgung des Managements zur Lösung dieser Fragen ist Aufgabe des Marketing-Controlling.³

Durch den Einsatz analytischer Informationssysteme auf der Basis einer Data-Warehouse-Architektur ergeben sich deshalb große Chancen, sich einen Marktvorsprung vor der Konkurrenz zu sichern. Neben der grundlegenden Dimension Zeit hat das Marketing-Controlling mindestens fünf weitere Dimensionen: Auftragsgröße, Kundengruppe, Absatzregion, Produktsortiment, Vertriebsweg.

Dem Marketingmanagement eines Unternehmens steht damit prinzipiell ein großes Reservoir an potenziellen Steuerungsgrößen zur Verfügung, auf deren Grundlage der Absatz von Produkten und Dienstleistungen effizienter, effektiver und nicht zuletzt vor allem kundenorientierter gestaltet werden kann. Abhängig ist dies davon, wie geschickt die jeweiligen Führungspersonen in der Lage sind, die Informationsflut zu filtern und richtig zu interpretieren. Schon Wild stellte 1971 eine Verdopplung des Informationsvolumens alle 6–7 Jahre fest [Wild71].

Zur Auswertung der gesammelten Daten bietet sich die Data-Warehouse-Technologie an. Allein die elektronischen Verkaufssysteme eines Versandhauses generieren täglich ca. 500.000 Verkaufsdatensätze, was einem Volumen an operativen Daten von ca. 1,5 Gigabyte pro Tag entspricht. Über umfangreiche, auf Modellannahmen basierende Transformationsprozesse werden daraus Managementinformationen erzeugt, die Aussagen darüber enthalten, welche Angebote an welchem Tag, zu welcher Zeit, in welcher Kategorie und von welchen Kundensegmenten genutzt wurden. Diese informationsorientierte Anwendung der Data-Warehouse-Technologie wird um Auswertungsmöglichkeiten ergänzt, um z.B. Kunden für neue oder ergänzende Angebote zu identifizieren, den Einfluss der Sortimentsgestaltung und Regalbestückung auf das Kaufverhalten zu simulieren oder andere interessierende Sachverhalte zu untersuchen. Gerichteten, d.h. hypothesengetriebenen, Auswertungsverfahren werden an dieser Stelle ungerichtete Auswertungsverfahren der Mustererkennung zur Seite gestellt. Data Mining bezeichnet die Nutzung von Methoden der Statistik, des maschinellen Lernens und der neuronalen Netze

3. Synonyme Bezeichner sind Vertriebs-Controlling oder Erlös-Controlling.

zur Identifikation von Mustern in Datenbeständen. Diese Muster können wiederum eingesetzt werden, um Klassifikations-, Segmentierungs- und Assoziationsaufgaben zu lösen. Ein typisches Anwendungsbeispiel für den Einsatz von Data Mining findet sich im folgenden Abschnitt zum Kampagnenmanagement.

■ Kennzahlensysteme:

Betriebswirtschaftliche Kennzahlen sind ein wesentlicher Inhalt von auswertungsorientierten Anwendungen. Sie dienen dazu, betriebliche Sachverhalte in konzentrierter Form wiederzugeben. Die wichtigsten Eigenschaften von Kennzahlen sind ihre Zweckorientierung, da sie quantifizierbare Informationen für Entscheidungssituationen enthalten. Kennzahlensysteme führen Kennzahlen, die sachlich und sinnvoll zueinander in Beziehung stehen, meist in einer hierarchischen Form zusammen, die in einer Spitzenkennzahl gipfelt. Sie ermöglichen damit eine zusammenhängende Betrachtung von Funktionen oder Prozessen inner- und außerhalb des Unternehmens. Eines der bekanntesten Kennzahlensysteme dieser Art ist das *Du Pont-System of Financial Control*, das den *Return on Investment* (ROI) als Spitzenkennzahl besitzt [Webe95].

Ein aktuelles Konzept für ein erfolgsfaktororientiertes Kennzahlensystem, das zurzeit in vielen Veröffentlichungen in Verbindung mit Data-Warehouse-Systemen genannt wird, ist die *Balanced Scorecard*. Der Kern dieses Konzepts besteht darin, erfolgsfaktororientierte Kennzahlen aus einzelnen Bereichen nicht isoliert zu betrachten, sondern immer den Wirkungszusammenhang in mehreren Sichten kombiniert darzustellen. Insbesondere sollen Unternehmensstrategien mithilfe von auf Erfolgsfaktoren basierenden Kennzahlen beurteilt werden. Durch die Balanced Scorecard soll letztendlich die betriebswirtschaftliche Wertschöpfungskette abgebildet werden. Daher sollte zwischen den Sichten eines Scorecard-Diagramms eine direkte Ursache-Wirkungs-Beziehung bestehen. Die folgenden Sichten werden von Kaplan und Norton für die Realisierung einer Balanced Scorecard vorgeschlagen [KaNo97]:

- In der *finanziellen Sicht* werden die klassischen Finanzkennzahlen betrachtet, wie z.B. der Return on Investment oder die Deckungsbeiträge. Die finanzielle Sicht nimmt die zentrale Stellung im Konzept ein, da der finanzielle Erfolg das generelle Ziel ist, das die anderen Sichten durch die Ursache-Wirkungs-Beziehungen anstreben müssen.
- Das Paradigma der Kundenorientierung findet seinen Niederschlag in der *Kundensicht*. Für Kunden und Marktsegmente müssen die erfolgsfaktor-basierten Kennzahlen definiert werden. Kennzahlen können das Verhältnis von Stammkunden zu Neukunden, die Größe des Kundenstamms oder Aussagen zur Kundenzufriedenheit sein.
- Die *Lern- und Innovationssicht* sollen die Zukunftsfähigkeit des betrachteten Bereichs beschreiben. Wie nimmt die Qualifikation meiner Mitarbeiter zu oder wie innovativ ist mein Produktprogramm (z.B. Anzahl neu ein-

geführter Produkte), können Fragestellungen sein, die in dieser Sicht eine Rolle spielen.

- Die *Geschäftsprozesssicht* spiegelt die internen Prozesse wider, die zum gewählten Betrachtungsbereich gehören.

Für jede Sicht sind Ziele zu definieren, die sich durch Kennzahlen quantifizieren lassen. Vorgaben und daraus abgeleitete Maßnahmen sollen die Zielerreichung sicherstellen. Mit Balanced Scorecards ist der Anspruch verbunden, eine Konzentration auf die wesentlichen Führungsinformationen zu erreichen, indem die für Führungsentscheidungen wichtigen Kennzahlen identifiziert werden und nur diese in Balanced Scorecards aufgenommen werden. Balanced Scorecards können dabei für jeden Informationsnachfrager individuell definiert werden.

Es wird vorgeschlagen, das Konzept der Balanced Scorecard sogar zum Vorbild für die inhaltliche Gestaltung von Data-Warehouse-Systemen für Management und Controlling zu machen, um sie dadurch als Grundlage für die strategische Unternehmensführung nutzen zu können [Gilm98]. Daten sollten aus der Lern- und Innovationssicht (z.B. Anzahl neu eingeführter Produkte), aus der Kundensicht und aus der internen geschäftsprozessorientierten Sicht (z.B. Bearbeitungszeit pro Teilprozess) integriert werden.

Neben der grafischen Visualisierung von Kennzahlen und deren Entwicklung einer Sicht gehören zu einer Balanced Scorecard aber auch textuelle Informationen wie Erläuterungen zu Zielen, Vorgaben und Maßnahmen.

Mit der Einführung des Balanced-Scorecard-Konzepts in einem Unternehmen sollte zeitgleich das konventionelle Berichtswesen modifiziert werden [MoSc98]. Balanced Scorecards sollten nicht einfach als zusätzliche Abschnitte in das bestehende Berichtswesen hinzugefügt werden. Vielmehr sollte das Berichtswesen um jetzt vielleicht überflüssige Abschnitte abgespeckt und um Balanced Scorecards sinnvoll erweitert werden, um nicht eine Informationsüberlastung der Berichtsempfänger zu verursachen [Acko67]. Beachtet werden sollte allerdings, dass eine Balanced Scorecard keine neue finanzielle Standardberichterstattung ist. Der Balanced-Scorecard-Ansatz dient außerdem nicht zur Entwicklung neuer Ziele oder Geschäftsstrategien [Klok99]. Vielmehr können mit seiner Hilfe der Erfolg von Geschäftsstrategien beurteilt und der Zielerreichungsgrad gemessen werden.

■ Kostenrechnung:

Entscheidungsobjekte der Kostenrechnung wie z.B. Kostenträger, einzelne Kostenstellen, Aufträge und Auftragsgrößen, Marktgebiete, Kunden oder Vertriebswege besitzen einen mehrdimensionalen Charakter. Entscheidungsobjekte der Kostenrechnung können daher gut in einem multidimensionalen Modell abgebildet und durch ein Data-Warehouse-System und OLAP-Werkzeug realisiert werden. Unter anderem bieten sich folgende Instrumente für eine Realisierung an [Oehl97]:

- Kostenstellenrechnung (Dimensionen: Kostenstelle, Szenario z.B. Plan oder Ist, Kostenart, Einflussgröße z.B. Maschinenstunden, Kostenspaltung)
- Auftragskalkulation (Dimensionen: Auftrag, Szenario, Kosten-/Erlösart, Einflussgröße, Kostenspaltung z.B. fix oder variabel, Kunde, Vertriebsweg)

Allerdings sind multidimensionale Informationssysteme nur beschränkt für die Abbildung von komplexen Werteflüssen geeignet. Data-Warehouse-System oder OLAP sind z.B. nur bedingt verwendbar, um den umfangreichen Wertefluss innerhalb einer Grenzplankostenrechnung zu beschreiben oder Buchungen mit Belastung und Entlastung einzelner Konten abzubilden. Hierfür sind eigene Kostenrechnungs- oder Finanzbuchhaltungssysteme prädestiniert, deren zweidimensionale Tabellenstruktur sich hervorragend durch relationale Systeme implementieren lässt. Daher ist es auch nicht sinnvoll, geschäftsvorfallorientierte Buchungen durch Controlling-Informationssysteme zu verwirklichen. Für den gelegentlichen Zugriff auf einzelne Geschäftsvorfälle bietet sich eher der fallweise Zugriff auf das operative System mittels SQL-Durchgriff an. Externe Rechnungswesendaten in multidimensionalen Informationssystemen basieren daher meistens auf den Monats- oder Jahresabschlussinformationen der Finanzbuchhaltung. Über eine Kontenrahmendimension kann monatlich der Kontenstand aus dem Rechnungswesensystem in das multidimensionale Informationssystem importiert werden. Hierarchien im Kontenrahmen können direkt abgebildet werden, sodass von verdichteten Bilanz- oder Gewinn- und Verlustrechnungspositionen bis auf die unterste Kontenebene navigiert werden kann. Oehler stellt fest [Oehl98], dass OLAP-Systeme nicht für eine geschäftsvorfallbasierte Abbildung von einzelnen Transaktionen, sondern stattdessen für eine periodische Gruppenbewertung geeignet sind. In einem Data-Warehouse-System sollten daher nicht einzelne Kostenbuchungen betrachtet, sondern nur statistische Nebenrechnungen durchgeführt werden.

Kampagnenbasierte Anwendungsfälle:

Kampagnenbasierte Anwendungsfälle kombinieren alle verfügbaren Werkzeuge und Auswertungsmöglichkeiten in einem Projektkontext. Soll beispielsweise ein neues Produkt erstellt, eingeführt und vertrieben werden, so kann mithilfe des Data-Warehouse-Systems das gesamte Projekt als Kampagne unterstützt werden – von der Marktanalyse und der Auswertung der Bestandskundengruppen bis zur Optimierung der Absatzwege und der Erfolgsüberwachung nach Markteinführung.

Die betrieblichen Anforderungen innerhalb eines Unternehmens sind meist so vielfältig, dass der dadurch induzierte Informationsbedarf häufig nicht einem einzelnen Zweck der Datenbereitstellung wie Informationsversorgung, Auswertungs- oder Planungsunterstützung zugeordnet werden kann. Vielfach entstehen Situationen, in denen verschiedene Anwendungsmöglichkeiten von Data-Ware-

house-Systemen kombiniert eingesetzt werden oder nur für bestimmte Aktionen verfügbar sein müssen. Kampagnen sind im Kontext der Datenanalyse Anwendungen, die zur Unterstützung strategischer Ziele nicht ständig, sondern nur zu besonderen Anlässen und mit wechselnden Vorgehensweisen eingesetzt werden. Die Kampagnenunterstützung ist eine funktional noch komplexere Aufgabenstellung als die Befriedigung des einfachen Informationsbedarfs im täglichen Geschäftsablauf oder die Auswertung von Datenbeständen. Die Durchführung liegt in der Hand eines kleinen Anwenderkreises von Spezialisten, die idealerweise fachliches Know-how hinsichtlich des Auswertungsziels und methodisches Wissen hinsichtlich der eingesetzten Auswertungsverfahren kombinieren. Die Integration der anderen Anwendungsgebiete ist dabei umfassend, da sowohl informations-, planungs- als auch auswertungsorientierte Unterstützung beim Kampagnenmanagement gebraucht wird.

Kampagnen stellen besondere Anforderungen an die Planung, Durchführung und Ergebniskontrolle sowie an die Datengrundlage, die eingesetzt wird. Als Gründe sind hier zu nennen:

- *Fehlende Standardisierung:*
Nicht die regelmäßige Bereitstellung von Information in Berichten oder die ständige Möglichkeit der Onlineanalyse ist Ziel von Kampagnen, sondern die sporadische Durchführung von Maßnahmen, die eine bestimmte strategische Aufgabe erfüllen.
- *Langfristiger Zeithorizont:*
Die Initiierung, Durchführung und Auswertung einer Kampagne erfolgt nur selten in einem Zug oder gar in einer Analysesitzung, sondern wird über Tage, Wochen oder gar Monate hinweg vorbereitet und implementiert.
- *Breite Datengrundlage:*
Die Datenquellen, aus denen Informationen für die Planung und Durchführung von Kampagnen gewonnen werden, sind in der Regel nicht auf die unternehmensinternen Anwendungssysteme beschränkt.
- *Umfassende Datenanalyse:*
Durch die Komplexität der Aufgabenstellung sollte eine möglichst breite Basis an Möglichkeiten zur Datenanalyse verfügbar sein.
- *Komplette Prozessunterstützung:*
Nicht nur die Auswertung, sondern auch alle vor- und nachgelagerten Schritte im Prozess einer Informationsgewinnung aus Datenbanken müssen ausreichend unterstützt werden.
- *Feedback-Schleife:*
Die Auswertung und Dokumentation der Kampagnenergebnisse sind besonders wichtig, da die Schaffung eines Erfahrungsschatzes bedeutende Vorteile bieten kann.

Kampagnen basieren immer auf dem dreistufigen Vorgehen Kampagnenplanung, -durchführung und Ergebnisevaluation. Ein Management dieser Phasen erfordert aufgrund der im nachfolgenden Abschnitt skizzierten Eigenschaften von Kampagnen den Einsatz besonderer Werkzeuge und Methoden.

Standardwerkzeuge sind in diesem Bereich nur teilweise vorhanden. Durch den projekthaften Charakter von Kampagnen werden eher verschiedene, teils eigenentwickelte Instrumente kombiniert. An besonderen Methoden bei weitergehenden Auswertungen wie z.B. *Abwanderungsanalysen*, *Response-Vorhersagen* bei *Marketingkampagnen* oder Ähnlichem sollten so vor allem ein ausgeprägtes statistisches Methodenrepertoire und Data-Mining-Verfahren verfügbar sein. Gerade die gering ausgeprägte Standardisierung der verschiedenen eingesetzten Methoden und Anwendungen führt häufig zu »Dateninseln«, die nicht zu den üblicherweise integrierten Datenquellen eines Data-Warehouse-Systems zählen.

Die besondere Komplexität des Kampagnenmanagements bedeutet jedoch auch, dass nicht alle Probleme durch den Einsatz eines Data-Warehouse-Systems gelöst werden können: Durch die stark schwankenden Anforderungen kann es häufig vorkommen, dass auch weitere, nicht im Data-Warehouse-System verfügbare Daten benötigt werden. Die Integration von ständig neuen unternehmensinternen oder externen Daten ist hier an der Tagesordnung. Die vorherige Planung des Data-Warehouse-Datenmodells ist im Kampagnenbereich besonders schwierig. Sind schon Analysemodelle ständigen Veränderungen ausgesetzt, so sind Kampagnenergebnisse noch weniger voraussagbar. Neben der Bereitstellung von eleganten Mechanismen zur Befriedigung der wechselnden Anforderungen stellt die Entscheidung, ob Daten dauerhaft integriert werden sollen oder nur für den Einzelfall gebraucht werden, eine besondere Herausforderung dar. Die Vorteile des Einsatzes eines Data-Warehouse-Systems überwiegen jedoch deutlich vor allem durch den verminderten Zeitbedarf bei der Kampagnendurchführung aufgrund von wegfallenden Schnittstellen- und Abstimmungsproblemen.

Am Beispiel eines *Customer-Relationship-Management-Systems* kann dies deutlich werden. Das Customer-Relationship-Management (CRM) fokussiert das Management der Kundenbeziehungen eines Unternehmens auf verschiedenen Ebenen und gehört somit zu den betriebswirtschaftlichen Bereichen Vertrieb und Marketing, in denen traditionell verstärkt Kampagnen eingesetzt werden. Besonders dieser Bereich hat in den letzten Jahren einen deutlichen Zuwachs erlebt, da die analytische Verarbeitung von Kundendaten im Gegensatz zur operativen, transaktionalen Verwendung stärkeres Interesse hervorgerufen hat. Auch hier kommen die Grundlagen des Data Warehousing zur Anwendung.

1.4.2 Wissenschaftliche Anwendungsbereiche

Schon seit den 1970er Jahren existiert der Bereich der Statistical and Scientific Databases (SSDB, [Mich91]), der ebenfalls die Integration, Verarbeitung und Auswertung großer Rohdatenmengen zum Ziel hat. Aus diesem Bereich stammen auch die technischen Wurzeln der Data-Warehouse-Technologie, vor allem im Hinblick auf die datenbanktechnische Speicherung und Verarbeitung.

Bei wissenschaftlichen, empirischen Untersuchungen fallen oft große Mengen an Daten, vor allem Messwerte, an. Beim Projekt *Earth Observing System* (EOS, [Mich91]) aus dem Bereich der Klima- und Umweltforschung werden sehr große Mengen an meteorologischen Daten von Bodenstationen und Satelliten gesammelt. Täglich fällt dabei ein Datenvolumen im Terabyte-Bereich an. Aus der unüberschaubaren Menge von Einzelwerten gilt es, die gewünschten Schlussfolgerungen zu ziehen, indem die Daten aufbereitet und analysierbar gemacht werden. Mithilfe von statistischen Untersuchungsmethoden oder Data-Mining-Techniken [FaPS96a] können schließlich Informationen aus den Daten extrahiert werden, die zur Gewinnung von neuen Erkenntnissen beitragen.

Ein weiteres Beispiel sind die Tsunami-Informationssysteme, die verstärkt nach der Tsunami-Katastrophe im Indischen Ozean im Jahre 2004 aufgebaut wurden. Diese Systeme sammeln große Mengen an seismologischen Daten und bereiten diese im Sinne eines Data-Warehouse-Systems auf. Dabei werden Warnsysteme genutzt, um in Echtzeit Muster zu erkennen und Tsunami-Warnungen aussprechen zu können. Außerdem werden die gesammelten Daten historisiert, um den Ablauf von Tsunami-verursachenden Seebeben im Nachhinein analysieren zu können und so neue Suchmuster für die Warnsysteme abzuleiten [NaKu10].

1.4.3 Technische Anwendungsbereiche

Auch in technisch orientierten Anwendungsfeldern gibt es viele mögliche Verwendungsfälle. Es seien hier einige wenige angedeutet. Die dem Data Warehousing zugrunde liegenden Mechanismen und Prinzipien gelten entsprechend.

Im öffentlichen Sektor ist die Sammlung von Messwerten aus Wasseranalysen notwendig, bei der die chemische Zusammensetzung des Wassers aus verschiedenen Quellen über die Zeit gesammelt wird. Dies bildet die Grundlage, die Wasserqualität und deren Veränderung zu beobachten und Einflussfaktoren darauf zu ermitteln. Ein Unternehmen des produzierenden Gewerbes kann beispielsweise eine Stoff- oder Materialdatenbank aufbauen, um die in Produkten befindlichen Inhaltsstoffe oder verwendeten Materialien zu dokumentieren. Damit ist es möglich, bei Rückrufaktionen oder Gewährleistungsfällen die betroffenen Chargen zu ermitteln oder eventuell verantwortliche Lieferanten ausfindig zu machen sowie Mängelquoten oder Ausfallwahrscheinlichkeiten zu analysieren.

1.4.4 Betriebswirtschaftliche Anwendungsbereiche

Es gibt keinen Bereich eines Unternehmens, in dem auf Daten und Informationen für eine erfolgreiche Abwicklung der Geschäftsprozesse verzichtet werden kann. Informationssysteme für die Abwicklung operativer Geschäftsprozesse, also Systeme für das Enterprise Resource Planning (ERP), besitzen oftmals funktional orientierte Module, wie z.B. für die Auftragsabwicklung. Diese Systeme bieten zwar meist funktionsbezogene Auswertungssichten, sie ermöglichen jedoch nur selten echte übergreifende Sichten für die Steuerung bzw. das Controlling gesamter Geschäftsprozesse. Data-Warehouse-Systeme besitzen demgegenüber eine effiziente Infrastruktur zur Informationsintegration und -bereitstellung und weisen deutliche Vorteile gegenüber herkömmlichen Methoden der verteilten Informationssammlung und -aufbereitung auf.

Im Sinne einer vertikalen Integration der Auswertungs- und Berichtsprozesse dient die integrierte Datensammlung des Data-Warehouse-Systems dabei zur Unterstützung von operativen, taktischen und strategischen Entscheidungen (Abb. 1–2). Auch hinsichtlich der Funktionsbereiche entlang der Wertschöpfungskette einer Organisation hat das Data-Warehouse-System integrativen und querschnittsorientierten Charakter.

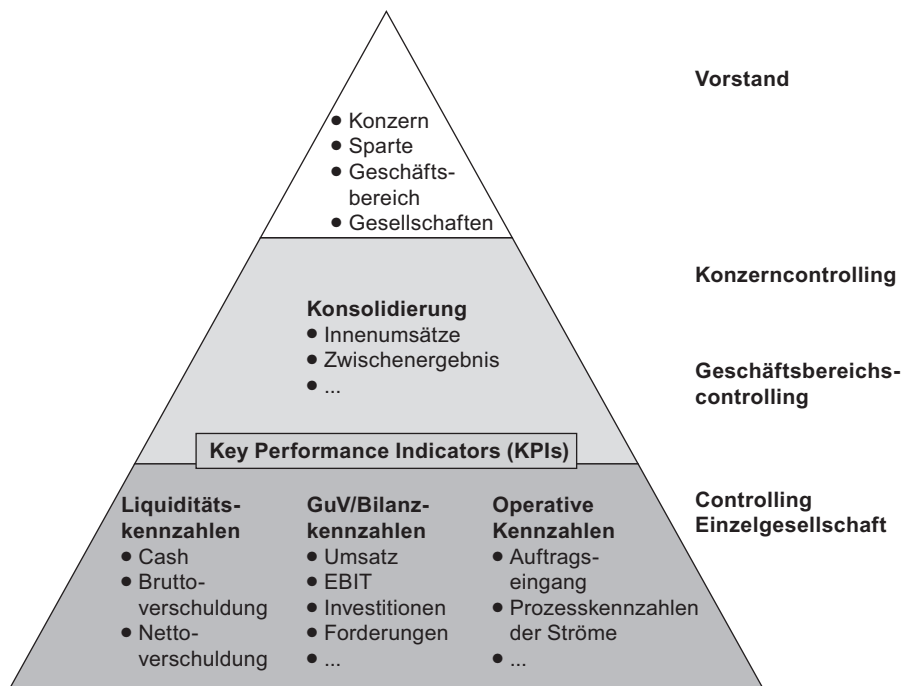


Abb. 1–2 Vertikale Integration der Auswertungs- und Berichtsprozesse in einem Data-Warehouse-System

Übergreifende Anwendungsgebiete

Schwerpunkt und Kernfunktion von Data-Warehouse-Systemen war schon immer die Informationslieferung, beginnend mit der Abbildung des klassischen Berichtswesens bis hin zu komplexen Auswertungen und Simulationen. Während historisch bedingt funktionsbereichsübergreifende Fragestellungen aus dem Controlling und der Logistik sowieso Auswertungen der Warenwirtschaftsströme dominierten, finden sich heute auf allen Ebenen der Organisation und in allen Schritten der Wertschöpfungskette Fragestellungen, die mithilfe eines Data-Warehouse-Systems und geeigneten Auswertungen zu beantworten sind. Zwar bringen Systeme zur Abwicklung operativer Geschäftsprozesse (z.B. ERP-System) mittlerweile eine Vielzahl von Auswertungsmodulen mit, die u.U. Auswertungsmöglichkeiten bereitstellen, die mit der Bezeichnung Business Intelligence versehen werden können. Allerdings fehlt hier in den meisten Fällen die funktionsbereichsübergreifende und integrierte Sicht auf die Unternehmensdaten. Diese wird im Data-Warehouse-System realisiert.

Managementprozesse

Anwendungen im Bereich der Managementprozesse zeichnen sich im Allgemeinen dadurch aus, dass unternehmensübergreifende Auswertungen auf hohem Aggregationsniveau durchgeführt und meist in Form von unternehmensweiten Kennzahlen aufbereitet werden. Die Entwicklung und Anpassung der Unternehmensstrategie oder Teilstrategien wie beispielsweise einer Produktstrategie werden u.a. auf der Basis von Auswertungen vorgenommen, die auf Kennzahlensystemen aufbauen. Bei den Kennzahlen kann es sich um (Key) Performance Indicators (KPI, z.B. »Anzahl der Kundenbeschwerden pro Produktgruppe«) oder (Key) Result Indicators (KRI, z.B. »Eigenkapitalrendite«) handeln. Die Auswertung der Kennzahlen kann über gezielte Auswertungen als Vorbereitung der Strategieentwicklung vorgenommen werden oder im Rahmen einer Überprüfung der Strategie unterjährig durchgeführt werden. Werden die Kennzahlen ständig überprüft, wird dies meist im Rahmen einer sog. »Balanced Scorecard« implementiert, indem aktuelle Werte für definierte Kennzahlen in einem integrierten System dargestellt werden. Die Werte für die vereinbarten Kennzahlen werden im Data-Warehouse-System errechnet und vorgehalten, das wiederum aus den Daten der eigenen operativen Systeme und externen Datenquellen (z.B. Marktforschungsdaten, demografische Datenbanken) gespeist wird. Auf diese Art werden beispielsweise Informationen zu Absatz, Umsatz, Kosten und Lagerbeständen aus den jeweiligen operativen Systemen extrahiert und im Data-Warehouse-System verwaltet. Gleichzeitig werden Daten zu Schulungsaktivitäten der Mitarbeiter, Mitarbeiterfluktuation oder Kundentreue gesammelt und mit den anderen Kennzahlen aus dem Data-Warehouse-System heraus in einer Balanced Scorecard dargestellt. Durch den Vergleich der tatsächlichen Werte mit den Zielwerten kann

so der Wirkungsgrad der aktuellen Unternehmensstrategie bewertet werden und es können ggf. Gegenmaßnahmen ergriffen werden.

Die Unternehmensfunktionen Planung, Steuerung und Controlling bauen auf Informationen aus dem Data-Warehouse-System auf, die zum größten Teil finanzieller Natur sind und im Wesentlichen aus den Systemen der internen und externen Rechnungslegung beliefert werden. Dabei werden etwa spezifische Umsatzzahlen pro Produktgruppe mithilfe des Data-Warehouse-Systems ausgewertet und Trends analysiert. Erweitert wird diese Sicht im Rahmen des Corporate Performance Management (CPM) – hier werden teilweise sehr detailliert Finanzdaten mit Daten aus dem Prozessmanagement zu einem Prozesscontrolling vereint, um detaillierte Leistungsbewertungen von unternehmerischen Prozessen zu ermöglichen. Die Grenzen zu oben beschriebenen strategischen Anwendungsgebieten sind fließend.

Die Anforderungen an Komplexität und Detailtiefe im Risikomanagement sind von Branche zu Branche verschieden. Die gesetzlichen Auflagen an das Risikomanagement von Banken (vgl. u.a. Basel II) und Versicherungen (vgl. u.a. Solvency II) beinhalten Prozesse und Risikoberechnungen, die in ihrer Komplexität ohne spezielles Data-Warehouse-System inkl. Auswertungen kaum zu erfüllen wären. So müssen beispielsweise Versicherungskonzerne die Risikoprofile aller Versicherungsnehmer einer Versicherungssparte übergreifend auswerten, um Risikohäufungen zu identifizieren. Sollten bei einer solchen Auswertung Ballungen von gleichartigen Risiken gefunden werden, so schlägt sich dies in der Berechnung des von der Versicherung vorzuhaltenden Risikokapitals nieder. Aufsichtsbehörden prüfen regelmäßig die Prozesse der Risikoberechnung und prüfen in diesem Zusammenhang auch die Qualität der Datenbeschaffung und -auswertung. In anderen Branchen reichen jedoch u.U. einfachere Datenbanken, die z.B. in der Automobilindustrie Daten über Produktlinien und Auslieferungen für den Fall einer Rückrufaktion vorhalten.

Die Anwendungsfälle im Bereich der Prüfungen (Audits, interne Revision, externe Wirtschaftsprüfung) sind im Wesentlichen auf das Erstellen standardisierter Berichte zu den Finanzkennzahlen des Unternehmens beschränkt. Einen Sonderfall stellt eine forensische Untersuchung durch externe Prüfer im Falle eines Verdachts auf kriminelle Machenschaften dar: In diesem Szenario wäre es denkbar, Data-Mining-Techniken auf spezielle Datenkreise durchzuführen, um Betrugsmuster aufzufindig zu machen.

Komplexer sind die Anwendungsfälle im Bereich Compliance: Nahezu alle Branchen sehen sich mittlerweile mit wachsenden externen regulatorischen Anforderungen konfrontiert. So sind beispielsweise Finanzinstitute bereits seit Jahren dazu verpflichtet, Finanzströme zu überwachen und auf Muster auszuwerten (Data Mining) bzw. mit Alarmmechanismen zu versehen, sollten zuvor definierte Muster gefunden werden oder definierte Limits über- oder unterschritten werden. Ziel der Echtzeit-Compliance-Überwachung ist beispielsweise die

Geldwäscheprävention. Data-Warehouse-Systeme liefern hier die nötige Datenbasis. Aber auch andere Branchen sehen sich Auflagen ausgesetzt, deren Erfüllung das Sammeln und Auswerten großer Datenmengen erfordert. Als Beispiel sei hier die Überwachung von Schadstoffemissionen der vielen Flugzeuge durch Fluggesellschaften genannt.

Kernprozesse

Im Bereich der Kernprozesse dominierten lange Zeit die operativen Insellösungen der einzelnen Bereiche. In einigen Fällen wurden diesen Systemen mittlerweile Auswertungsmodule hinzugefügt, die u.U. Business-Intelligence-ähnliche Auswertungen zulassen. Die Sicht auf die hier vorliegenden Informationen bleibt aber silobasiert – erst durch Data-Warehouse-Systeme werden Informationen strukturell und inhaltlich integriert nutzbar.

In der Produktentwicklung ist eine Vielzahl von Szenarien denkbar, in denen die Auswertung von großen Datenmengen Design und Verbesserung von Produkten unterstützen kann. Dabei kann es sich um wissenschaftliche Anwendungsfälle in der Forschung handeln wie auch um statistische Szenarien als analytische Basis für das zielgruppenorientierte Produktdesign.

Im Rahmen der Teilwertschöpfungskette Eingangslogistik – Produktion – Ausgangslogistik spielen Auswertungen auf der Basis von Data-Warehouse-Daten insbesondere eine Rolle bei der Optimierung von Warenflüssen, Bestellverhalten und Lagerung. Während sich der Einkäufer über ein Cockpit den aktuellen Warenbestand anzeigen lassen kann, sind OLAP- oder andere Ad-hoc-Analysen geeignet, um geografische Häufungen und Absatzmuster zu identifizieren und so beispielsweise frühzeitig erkannte Trends aktiv zu unterstützen und die nötigen Anpassungen in Fertigung und Logistik zu planen. Auch im Sinne des oben bereits erwähnten Corporate-Performance-Managements sind hier Auswertungen von Prozessperformancedaten beliebte Einsatzszenarien.

Marketing und Vertrieb waren historisch neben den Steuerungsfunktionen die ersten Nutzer von Data-Warehouse-Daten und deren Auswertungen. Neben auswertungs- und informationsorientierten Anwendungsfällen sind besonders hier kampagnenorientierte Anwendungsfälle zu beobachten. An dieser Stelle soll nur eine Auswahl aus der Vielzahl an Beispielen genannt werden: Kundensegmentierung nach Alter, Geografie und Kaufverhalten ermöglicht zielgruppengerechte Werbung und Kundenansprache, während Auswertungen im Vertriebscontrolling erlauben, die Außendienstmitarbeiter gezielt zu steuern. Die konsequente Fortführung der detaillierten Auswertungen der Zielgruppe ist der Kundenservice mittels Data-Warehouse-Systemen. Kundenverwaltungssysteme (Customer-Relationship-Management – CRM) sind hier ein Beispiel für eine spezialisierte Lösung für diesen Bereich.

Ein Unternehmen möchte über sein Marktumfeld wie auch über die Mitbewerber gut informiert sein. Mithilfe der unstrukturierten Datenauswertung kön-

nen relevante Daten aus Internetpräsenzen, Shopsystemen oder Presseberichten extrahiert und ausgewertet werden. Jedes verfügbare digitalisierbare Schriftstück mit Themenbezug kann als Quelle verwendet werden. Werden die dabei gewonnenen Informationen zusammengefasst, können sie einen recht genauen Überblick geben. Die folgende Auflistung gibt einen Einblick in die Themengebiete:

- Produkt und Preisevaluation der Mitbewerber
- Test/Produktbewertungen
- Medienpräsenz und Ranking
- Bekanntheit und Bewertungen in Blogs, Foren oder sozialen Netzwerken

Ein Szenario kann dabei sein, Kundenstimmen in Foren oder Blogs auszuwerten und typische Probleme oder positive Merkmale automatisch zu identifizieren. Ein weiteres Szenario ist, Nachrichten in Kurznachrichtendiensten wie z.B. Twitter über die eigenen Produkte auszuwerten und so z.B. auf Vorkommnisse rasch reagieren zu können.

Die gewonnenen Informationen, wie z.B. Menge der monierten Beschwerden bzw. Produkte, können in strukturierter Form in einem Data-Warehouse-System zur weiteren Auswertung bereitstehen. Durch diese Quellenauswertung entstehen komplett neue Marketingsichten, da beispielsweise eine Häufung von negativen Beiträgen in Foren, Blogs oder Kurznachrichtendiensten direkt nach Einführung eines Produkts eine schnelle Reaktion ermöglicht.

Ein Unternehmen kann mit seiner Umwelt (z.B. Kunden, Zulieferer, externe Mitarbeiter) über eine breite Bandbreite an Möglichkeiten kommunizieren. Wird dabei die Übermittlung formal definierter Daten, z.B. Systemschnittstellen, nicht in Betracht gezogen, erfolgen diese Kommunikation und Interaktion zumeist über unstrukturierte Daten. Darunter fallen z.B. E-Mails, Briefe oder Chat-Nachrichten, aber auch indirekte Kommunikationsformen wie beispielsweise Freitextfelder in einem Shopsystem. Die Auswertung dieser Daten kann über Natural Language Processing erfolgen. Extrahierbare Informationen sind beispielsweise:

- Adressinformationen, z.B. E-Mail, Anschrift
- Kategorisierung, z.B. Beschwerde, Angebotsanfrage
- Profilerstellung, z.B. passende Angebote für einen Kunden
- Semantik einer Korrespondenz wie z.B. Bedeutung und Inhalt einer Mail. Die semantische Analyse geht dabei über eine reine Kategorisierung hinaus; mit ihrer Hilfe wird versucht, den Inhalt eines Dokuments zu erschließen. Als Beispiele können hier die Erkennung einer Beschwerde (Kategorisierung) und gleichzeitig die Erkennung des Beschwerdegrunds (z.B. defektes Gerät oder Beschwerde über einen Mitarbeiter) angeführt werden.

Diese aus den unstrukturierten Daten extrahierten Informationen können in einem Data-Warehouse-System zusammen mit den bereits vorhandenen strukturierten Daten verarbeitet werden.

Die Auswertung von unstrukturierten Daten in sozialen Netzwerken liefert ein breites Spektrum an Informationen. So können Blogs oder Benutzereinträge benutzt werden, um Produkte oder Unternehmen zu evaluieren. In sozialen Netzwerken können Benutzervernetzungen z.B. durch die Analyse von Trackbacks ausgewertet werden. Die Anzahl der Trackbacks könnte beispielsweise einen Hinweis auf die Relevanz eines Benutzers geben. Ein weiteres Szenario ist die Auswertung der »Follower« eines Twitter-Accounts als Maßstab für Bekanntheit und seine Einflussmöglichkeit auf andere Benutzer. Wenn dieser Benutzer beispielsweise einen Eintrag zu einem Produkt eines Unternehmens erstellt, kann durch die Anzahl der Follower zugleich die Relevanz des Twitter-Eintrags bewertet werden. Eine hohe Anzahl an Followern kann zugleich eine hohe Bedeutung des Eintrags implizieren. Weiter können automatisch Benutzerprofile und Postings in einem sozialen Netzwerk analysiert werden, um Trends zu erkennen.

Unterstützungsprozesse

In der Informationstechnologie selbst werden Informationen eines Data-Warehouse-Systems zunehmend durch die Ausbreitung von IT Service Management (ITSM) relevant. ITSM betrachtet IT auf der Basis definierter Dienstleistungen, die in einem definierten Rahmen und einer definierten Qualität zu erbringen sind. In vielen Fällen erleichtert diese Betrachtungsweise ein Outsourcing einzelner IT-Prozesse. Unabhängig davon, ob der den Prozess betreibende Dienstleister nun intern oder extern angesiedelt ist – IT bedeutet heute oft das Überwachen und Managen von Prozessen und deren Servicequalität wie Antwortzeiten, Rechenperformance etc. Auch hier entstehen in operativen Systemen komplexe Datenmengen, die mithilfe der Auswertungswerkzeuge analysiert werden: Durch Standardberichte, Cockpits und Warnsysteme und im Falle von besonderen Vorkommnissen kann die Historie der erbrachten Servicequalität durch Ad-hoc-Auswertungen oder OLAP untersucht werden.

In der Personalwirtschaft (engl. Human Resources) fassen Data-Warehouse-Systeme und ihre Auswertungsmöglichkeiten auch mehr und mehr Fuß. Das prominenteste Beispiel ist die US-Firma Google, die der hohen Fluktuation in der eigenen Belegschaft ein Data-Warehouse-System mit Verhaltensdaten der Mitarbeiter (u.a. Krankheit, Urlaub, Bewertungen) entgegengestellt hat. Ein intelligenter Algorithmus kann durch Auswertungen die Wahrscheinlichkeit berechnen, dass ein individueller Mitarbeiter kurz davorsteht zu kündigen, sodass der Arbeitgeber proaktiv gegensteuern kann. Das – mit deutschem Arbeitsrecht kaum vereinbare – Beispiel soll illustrieren, welche Möglichkeiten die Auswertung von Data-Warehouse-Daten im Personalbereich bieten.

Bei großen Unternehmen spielen Beschaffungs- und Liegenschaftsmanagement u.U. eine wichtige Rolle, da die entstehenden Kosten nicht unerheblich sein können. Je nach Größe der verwaltenden Abteilungen bzw. des verwalteten Portfolios sind hier unterschiedliche Anwendungsfälle sinnvoll.

1.5 Überblick über das Buch

Die bisher adressierten Themen werden in den folgenden Kapiteln des Buches vertieft. Dazu wird aus der Fülle der Anwendungsgebiete ein typisches Beispiel für eine genauere Beschreibung herausgegriffen, um die im Buch diskutierten Konzepte zu veranschaulichen. Sowohl die einzelnen Aspekte zum Aufbau eines Data-Warehouse-Systems als auch die Theorie der Datenbankmodellierung können an diesem Szenario verdeutlicht werden.

Im Folgenden wird zuerst das Beispiel *Star*Kauf* erläutert und danach die Übersicht der Kapitel gegeben.

1.5.1 *Star*Kauf*

Die fiktive Kaufhauskette *Star*Kauf* möchte ihre innerbetrieblichen Abläufe, das Kaufverhalten der Kunden und die Zusammensetzung des Sortiments überprüfen. Dazu werden vom Management u.a. folgende Anforderungen gestellt:

- Überprüfung des Sortiments zur Identifizierung von Ladenhütern (engl. poor sellers) oder Verkaufsschlagnern (engl. good sellers)
- Durchführung einer Warenkorbanalyse mithilfe der Information aus den Kassensbons
- Treffen von Entscheidungen mittels Kundenwünschen und -strukturen zur Durchführung von gezielten Werbemaßnahmen oder Nutzung zur systematischen Regalbestückung
- Messen der Kundenzufriedenheit durch die Sammlung von Informationen über Reklamationen – auch speziell bzgl. bestimmter Produkte
- Messen des Zusammenhangs der Einführung einer Kundenkarte auf die Verkäufe
- Untersuchung der Wirksamkeit von verkaufsfördernden Maßnahmen
- Durchführung von Standortanalysen zur Messung der Rentabilität von neuen und bestehenden Kaufhäusern

Weiterhin existieren bereits Vorgaben, welche Auswertungen möglich sein sollen und in welcher Form diese aufzubereiten sind. Ein Beispiel eines Berichts, der bisher nur in Papierform verfügbar war, ist in Tabelle 1–4 zu sehen. Darin werden Umsätze nach Jahren und Bundesländern auf der vertikalen Achse sowie nach einer Produktklassifikation auf der horizontalen aufgegliedert. Diese Darstellungsform ist relativ statisch, d. h., auf Änderungswünsche der Informationsempfänger kann nur mit hohem Erstellungsaufwand reagiert werden.

Umsatz		Unterhaltungselektronik				Bekleidung	
		Video	Audio	TV	SUMME	...	Summe
2010	Bayern	12	31	15	58
	Hessen	22	51
	SUMME	34
2011	Bayern	48	67	55	170
	Hessen	50
	SUMME	98
2012	Bayern	58	66	51	175
	Hessen	67
	SUMME	125
SUMME		257	904

Tab. 1–4 Beispiel eines zu erstellenden Berichts

Einer der Kernpunkte der durchgeführten Anforderungsanalyse ist eine flexible und schnelle Erstellung sowie Änderung von Berichten. Weiterhin soll mit entsprechender Software eine dynamische Auswertung der Daten nach frei definierbaren Kriterien möglich sein. Als Kriterien bzw. Betrachtungsperspektiven, die in die Auswertungen einfließen sollen, sind Produkte, Kunden, Filialen, Zeit und Bondaten zu nennen. Relevante Kennzahlen sind beispielsweise Ein- und Verkäufe, Lagerbestände in Stück, Umsatz und Preis.

Aufgrund des Zusammenschlusses mit einigen teilweise ausländischen Unternehmen kam es zu einer heterogenen IT-Landschaft mit vielen kleinen Datenbanken. Auf diesen autonomen Datenbeständen arbeiten unabhängige Anwendungen. Nach den Wünschen der IT-Abteilung wird über WAN oder Internet eine zentrale Datenbank zur Integration im Mutterhaus geschaffen, auf der operative Anwendungen, wie beispielsweise die unternehmensweite Kostenrechnung, aufsetzen sollen (Abb. 1–3). Um die oben aufgeführten Anforderungen erfüllen zu können, möchte man ein Data-Warehouse-System aufsetzen, das seine Daten ebenfalls aus der zentralen Datenbank beziehen soll.

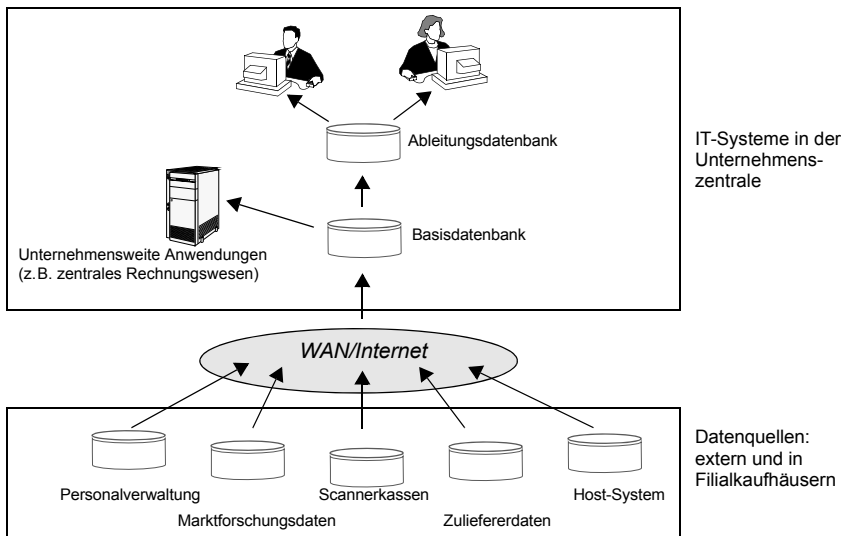


Abb. 1–3 Architekturskizze für das geplante System

1.5.2 Kapitelübersicht

Nachfolgend soll ein Überblick über Aufbau und Zielsetzung des Buches gegeben werden. Die Vorteile des Buches liegen in der Vielfalt der Themen und ihrer integrierten Darstellung. Es wurde bewusst darauf geachtet, keinen Forschungsbericht zu schreiben, sondern durch die Erklärung von allgemeinen Konzepten, deren Umsetzung sowie Praxisbeispiele möglichst einen großen Bereich des Themenkomplexes abzudecken. Auf eine genaue Erklärung von Produkten wird wegen der Schnelllebigkeit der Soft- und Hardware verzichtet.

Das Buch ist in die drei Teile Architektur (I), Entwicklung (II) und Anwendung (III) aufgeteilt (Abb. 1–4). Sowohl »Architektur« als auch »Anwendung« betrachten das Thema »Data-Warehouse-System« ganzheitlich. Dagegen fokussiert sich der Teil »Entwicklung« auf die datenbankspezifischen Themen. Inhaltlich wurde das Buch so gestaltet, dass im Teil I (*Architektur*) ein Architekturvorschlag eines Data-Warehouse-Systems gegeben wird. Aus der geläufigen Überlegung heraus, dass eine Architektur, also die Struktur eines Systems, sowohl statisch als auch dynamisch beschrieben werden muss, um sie vollständig zu diskutieren, werden die Grundlagen einerseits anhand der benötigten statischen Komponenten (Referenzarchitektur) und andererseits im dynamischen Ablauf (Phasen) erläutert. Die physische Architektur greift die konzeptionellen Ansätze der Referenzarchitektur unter Implementierungsgesichtspunkten und weiterführenden Ansätzen wieder auf. Teil II beschreibt die *Entwicklung* eines Data-Warehouse-Systems mit der Konzeption, Modellierung und Realisierung und fokussiert

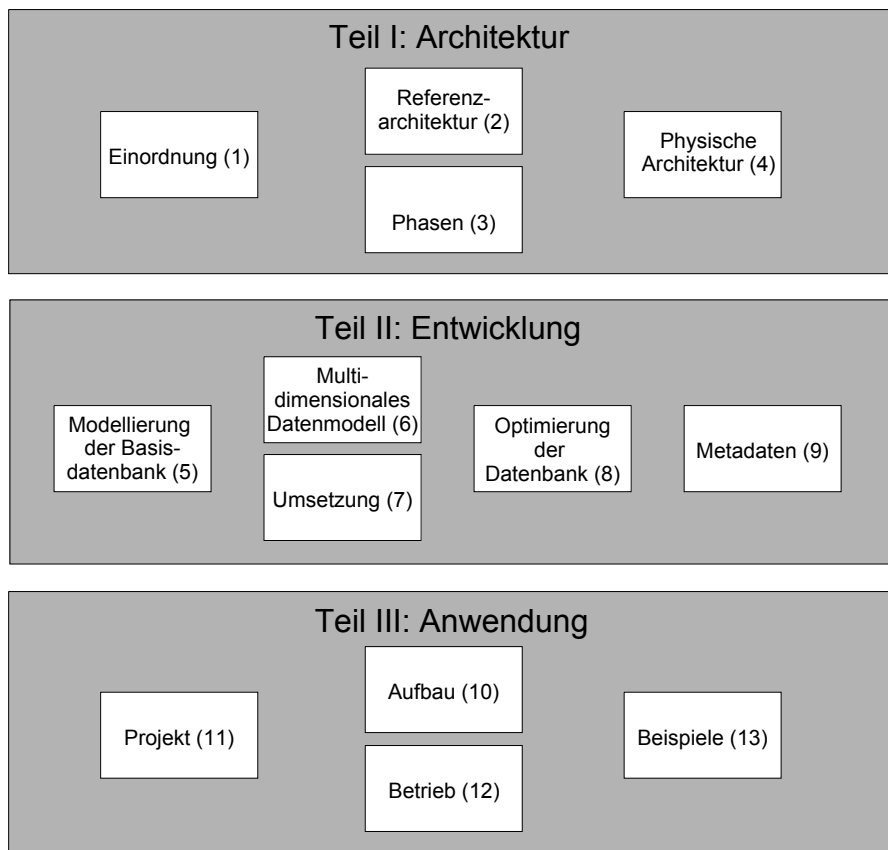


Abb. 1–4 Zusammenhänge der Teile und Kapitel des Buches

datenbanktechnische Aspekte. Metadaten stellen in diesem Kontext einen weiteren wesentlichen Beitrag dar. Im Teil III (*Anwendung*) werden die Konzepte aus den Teilen I und II aus Praxissicht, d.h. die methodische Vorgehensweise des Aufbaus, das Data-Warehouse-Projekt und die Betriebsphase eines Data-Warehouse-Systems, beschrieben sowie mit konkreten Anwendungsbeispielen illustriert.

Die behandelten Themenbereiche ermöglichen es, ein breites Leserspektrum anzusprechen. Das Buch ist für Studenten der Informatik, Betriebswirtschaft oder Wirtschaftsinformatik als Grundlagenwerk mit vielen Informationen und Literaturreferenzen gedacht. Es eignet sich auch zur gezielten Prüfungsvorbereitung. Professoren finden eine geeignete Basis, um direkt eine Vorlesung vorzubereiten. Anwender und Entwickler besitzen ein Nachschlagewerk sowohl für die Theorie als auch für die tägliche Praxis.

Überblick über die Kapitel

Teil I: Architektur

- *Kapitel 1 – Abgrenzung und Einordnung:*
Grundlegende Begriffe und die Vielfalt des Anwendungsgebietes werden beschrieben.
- *Kapitel 2 – Referenzarchitektur:*
Die Referenzarchitektur dient als idealtypische Grundlage. Durch die Vielzahl der benötigten Komponenten wird deutlich, dass ein Data-Warehouse-System mehr als nur eine Datenbank ist.
- *Kapitel 3 – Phasen des Data Warehousing:*
Neben den statischen Komponenten der Referenzarchitektur existiert auch der dynamische Aspekt, der den Vorgang des Data Warehousing aus der Prozesssicht beleuchtet.
- *Kapitel 4 – Physische Architektur:*
Es gibt eine Vielzahl von Realisierungsmöglichkeiten und zu realisierende Teilbereiche.

Teil II: Entwicklung

- *Kapitel 5 – Modellierung der Basisdatenbank:*
Die Basisdatenbank als Integrationsplattform der Datenquellen steht im Mittelpunkt der Betrachtung.
- *Kapitel 6 – Das multidimensionale Datenmodell:*
Die Datenbank und deren Modell stellen das Herzstück eines Data-Warehouse-Systems dar.
- *Kapitel 7 – Umsetzung des multidimensionalen Datenmodells:*
Die im vorherigen Kapitel eingeführten Konzepte werden sowohl in einem relationalen als auch multidimensionalen Datenbanksystem umgesetzt.
- *Kapitel 8 – Optimierung der Datenbank:*
Aufgrund der großen zu verarbeitenden Datenmengen ist eine Optimierung der Datenbank notwendig. Es werden Strategien und Ansätze aufgezeigt, um die Anwenderanfragen in adäquater Zeit zu erfüllen.
- *Kapitel 9 – Metadaten:*
Metadaten stellen den Schlüssel zum Aufbau und zur Wartung des Data-Warehouse-Systems dar, da nur durch sie die Integrationsleistung erzielt werden kann. Im Mittelpunkt stehen die Klassifikation der Metadaten und der Zusammenhang zum Data-Warehouse-System.

Teil III: Anwendung

- *Kapitel 10 – Vorgehensweise beim Aufbau eines Data-Warehouse-Systems:*
Die im Teil I motivierte Architektur ist eng mit der Methodik zum Aufbau verbunden. Es werden neben der Vorgehensweise auch die Einbettung in den unternehmensweiten Kontext und die Strategie diskutiert.
- *Kapitel 11 – Das Data-Warehouse-Projekt:*
Der Aufbau eines Data-Warehouse-Systems aus der Sichtweise des Projektmanagements wird mit all seinen Fallstricken und Besonderheiten erläutert.
- *Kapitel 12 – Betrieb und Weiterentwicklung eines Data-Warehouse-Systems:*
Der Aufbau eines Data-Warehouse-Systems ist nur ein Teil der zu erbringenden Leistung bei der Umsetzung eines Data-Warehouse-Vorhabens. Genauso wichtig ist der Wartungs- und Betriebsaspekt, der Gegenstand dieses Kapitels ist.
- *Kapitel 13 – Praxisbeispiele:*
Praxisbeispiele aus verschiedenen durchgeführten oder in Durchführung befindlichen Projekten runden das Buch ab und geben einen Einblick in die Komplexität, aber auch in die Verwendbarkeit der Data-Warehouse-Idee.