

## 2 *Data Warehouse* und OLAP \*

Fach- und Führungskräfte in den Unternehmungen sehen sich mit einer steigenden Datenflut bei einem gleichzeitigen Informationsdefizit konfrontiert. Zahlreiche Unternehmungen besitzen zwar große Mengen an Daten, ohne diese effektiv und effizient nutzen zu können. Vor diesem Hintergrund ist es Anliegen des folgenden zweiten Teils, die in der Praxis bewährten Ansätze zur Verbesserung der unternehmungsweiten Informationsversorgung zu diskutieren.

Zunächst werden dazu die Grundlagen von Data Warehouse- und OLAP-Systemen dargestellt, die heute als zentrale Technologien zur anforderungsgerechten Versorgung mit entscheidungsorientierten Informationen verstanden und genutzt werden:

- »Grundlagen«, S. 41

Anschließend erfolgt die Beschreibung der verfügbaren Methoden zur Modellierung multidimensionaler Datenstrukturen, die bei entscheidungs- und analyseorientierten Informationssystemen im Mittelpunkt stehen und daher in der gebührenden Breite und Tiefe zu behandeln sind:

- »Modellierung und Implementierung«, S. 73

Schließlich wird die im ersten Teil präsentierte Fallstudie aufgegriffen und eine exemplarische, phasenorientierte Systemgestaltung für das fiktive Unternehmungsszenario vorgestellt:

- »Fallstudie: TOPBIKE – BI«, S. 102

### 2.1 *Grundlagen* \*

Entscheidungs- und analyseorientierte Informationssysteme erweisen sich sowohl im Hinblick auf die verfolgte Zielrichtung als auch bezüglich der eingesetzten Bausteine und Komponenten als grundsätzlich verschieden zu den Vertretern der operativen Systemwelt. Nicht die zeitpunktgenaue und höchstaktuelle Erfassung von Daten, die das betriebliche Tagesgeschehen begleiten und dokumentieren, steht im Vordergrund, sondern die anforderungsgerechte Zurverfügungstellung eines aufbereiteten Informationsbestandes mit geeigneten Zugriffstechnologien. Die vorgestellten Kon-



zepte in den folgenden Ausführungen versprechen, eine umfassende Abdeckung dieser Anforderungen mit modernen Softwaretechnologien leisten zu können.



Zunächst soll dazu das Data Warehouse-Konzept vertiefend erörtert werden, welches das Ziel verfolgt, die Verfügbarkeit, Qualität und Integrität entscheidungsorientierter Informationsbestände herzustellen und langfristig zu sichern:

- »Einordnung und Komponenten des Data Warehouse-Konzeptes«, S. 43

Neben der allgemeinen Einordnung und Abgrenzung von Data Warehouse-Lösungen sind hier auch Architekturen und Komponenten zu diskutieren. Der nachfolgende Abschnitt widmet sich dem On-Line Analytical Processing (OLAP), welches die Aspekte hervorhebt, die für eine anforderungsgerechte Nutzung analyseorientierter Informationssysteme nötig sind:

- »OLAP – On-Line Analytical Processing«, S. 52

Mit Hilfe der OLAP-Technologie wird dem Management ein mächtiges Werkzeug geboten, um schnelle, interaktive und vielfältige Zugriffe auf relevante und konsistente Informationen in Verbindung mit dynamischen und multidimensionalen Analysen durchzuführen [ChGI06a, S. 18]. Danach wird ein Phasenmodell präsentiert, anhand dessen sich Data Warehouse- und OLAP-Systeme stufenweise und systematisch gestalten und betreiben lassen:

- »Vorgehensmodell zur Gestaltung multidimensionaler Informationssysteme«, S. 65

Das Modell bietet eine idealtypische Vorgehensweise, die sich durch vielfache Rücksprünge und Wiederholungen auszeichnet, und bildet die Basis für die an späterer Stelle durchgeführte Fallstudie (vgl. »Fallstudie: TOPBIKE – BI«, S. 102). Das letzte Kapitel ist den vielfältigen Einsatzbereichen von Data Warehouse- und OLAP-Systemen gewidmet:

- »Einsatzbereiche multidimensionaler Informationssysteme«, S. 69

## 2.1.1 Einordnung und Komponenten des *Data Warehouse*-Konzeptes \*

Das Thema *Data Warehouse* (DW) erfreut sich in Theorie und Praxis einer großen Beachtung. Schließlich verspricht es den betrieblichen Anwendern Lösungsvorschläge für den Aufbau einer abgestimmten Informationsinfrastruktur auf der Basis verfügbarer Hard- und Softwaretechnologien. Da ein spezifisches *Data Warehouse*-System für den praktischen Einsatz immer auf die Besonderheiten der jeweiligen Anwenderunternehmung ausgerichtet sein muss, kann das Konzept allerdings stets nur einen Orientierungsrahmen anbieten und kein Patentrezept.

In der einschlägigen Literatur besteht weitgehende Einigkeit über den Zweck und das Wesen eines *Data Warehouse*, wie die folgenden Beispiele belegen:

Der Begriff *Data Warehouse* im engeren Sinne bezeichnet eine von den operationalen DV-Systemen isolierte, unternehmungsweite Datenbasis, die anhand einer konsequenten Themenausrichtung unternehmensrelevanter Sachverhalte (z. B. Absatzkanäle, Kunden- und Produktkriterien) speziell für Endbenutzer aufgebaut ist [ChGl06a].

DW-Definition  
Mucksch

Ein *Data Warehouse* ist eine physische Datenbank, die eine integrierte Sicht auf beliebige Daten zu Analysezwecken ermöglicht [BaGü04, S. 7].

DW-Definition  
Bauer/Günzel

Für die weiteren Überlegungen soll folgendem Begriffsverständnis gefolgt werden, das vor allem auch die inhaltliche Trennung zu den operativen Datenbasen betont:

Unter einem *Data Warehouse* ist eine Systemlösung zu verstehen, die die unternehmungsweite Versorgung der Front-End-Systeme zur Managementunterstützung mit den benötigten Informationen gewährleistet. Zweckmäßigerweise wird das *Data Warehouse* getrennt von den operativen Vordatenbanken aufgebaut und betrieben. Nur so lässt sich eine konsistente unternehmungsweite Datenbasis etablieren, in die selektierte und verdichtete Informationen anwendungsgerecht aufbereitet einfließen und auf die interaktiv und intuitiv zugegriffen werden kann.

DW-Definition

Für die gespeicherten Dateninhalte ist deren thematische Ausrichtung sowie Vereinheitlichung, Dauerhaftigkeit und Zeitorientierung charakteristisch.

DW-Schwerpunkt

Somit sind es beim Data Warehousing primär die technischen Implikationen auf der Back-End-Seite (Hintergrundstruktur), auf denen der Fokus liegt. Von den Oberflächenwerkzeugen wird dagegen zunächst weitgehend abstrahiert, wenngleich das *Data Warehouse* sicherlich effiziente Zugriffsformen ermöglichen und so ausgerichtet sein soll, dass die Analysierbarkeit und Auswertbarkeit der Daten auch durch Nicht-Computerexperten gewährleistet ist.

DW-Ziel

Schließlich gilt es als explizites *Ziel* des Ansatzes, den Datenbestand derart vorzuhalten, dass der Endbenutzer (bzw. das entsprechende Endbenutzersystem) wie in einem realen Warenhaus auf die für ihn relevanten Informationen zugreifen und diese nutzen kann. Beim Aufbau eines *Data Warehouse*-Konzeptes sind sowohl betriebswirtschaftlich-organisatorische als auch technische Gestaltungsaspekte sorgfältig zu durchdenken.

Betriebswirtschaftlich-organisatorische Gestaltungsziele

Aus betriebswirtschaftlich-organisatorischer Sicht ist zu überlegen, welche Informationen auf welchen Verdichtungsstufen im Datenspeicher abgelegt werden müssen und welchen Mitarbeitern diese zugänglich gemacht werden sollen. Dabei eignet sich eine direkte und ausschließliche Verwendung der in den operativen Systemen abgelegten Rohdaten nur sehr eingeschränkt für eine angemessene Unterstützung betrieblicher Entscheidungsprozesse. Vielmehr müssen die Daten aufbereitet werden, um als entscheidungsgerechte Informationen dem Endbenutzer wertvolle Dienste leisten zu können. Zudem ist zu klären, was konkret unter einzelnen Begriffen zu verstehen ist bzw. woraus sich die einzelnen Größen zusammensetzen, was sie repräsentieren und wie sie ermittelt werden. Die Definition betriebswirtschaftlicher Größen erweist sich in größeren Unternehmen als durchaus ernstzunehmendes und schwieriges Unterfangen, insbesondere wenn in den einzelnen unternehmerischen Organisationseinheiten unterschiedliche Begriffsverständnisse existieren.

Daneben muss ebenfalls ein tragfähiges technisches Realisationskonzept erarbeitet werden. Wie bereits ausgeführt, sind bei der Realisierung eines *Data Warehouse* Anforderungen zu erfüllen, die in ähnlicher Weise auch für die Datenhaltungssysteme der operativen Anwendungsebene Gültigkeit aufweisen. Daneben allerdings sind weitere Aspekte zu beachten, die eine reine Adaption der im operativen Bereich erfolgreichen Konzepte als untauglich erscheinen lässt.

Folgende Themen werden detaillierter behandelt:

- »Data Warehouse-Architekturen und -Komponenten«, S. 45
- »Prozesse zum Extrahieren, Transformieren und Laden von Daten«, S. 49



### 2.1.1.1 *Data Warehouse*-Architekturen und -Komponenten \*

***Data Warehouse*-Lösungen lassen sich heute auf der Basis unterschiedlicher Architekturvarianten realisieren. Unterscheidungskriterien stellen insbesondere die Anzahl der Ebenen und Art der eingesetzten Datenhaltungseinrichtungen (Daten-Layer), die Möglichkeiten des Datenzugriffs einschließlich der zugehörigen, weiterführenden Analyseoptionen sowie die Techniken und Werkzeuge zur Umwandlung und Aufbereitung der Daten beim Austausch zwischen den einzelnen Hierarchieebenen dar.**

Ein erster, fast naiver Ansatz kann in einem Durchgriff auf die Datenspeicher der operativen Systeme mit lokaler Ablage der Extrakte gesehen werden. In der Regel erfolgt hierbei über feste Verknüpfungen ein periodischer Datenimport auf die lokale Festplatte, um auf diesen Daten benötigte Auswertungen mit den gängigen Desktop-Softwarewerkzeugen (z. B. MS Excel) durchzuführen. Auch können die Datenextrakte durch individuelle Softwarelösungen genutzt werden, die etwa mit den verbreiteten GUI-Entwicklungswerkzeugen (Delphi oder Visual Basic) erstellt worden sind.

Die Funktionsfähigkeit derartiger Lösungen wird in hohem Maße durch die Mächtigkeit der Extraktionswerkzeuge und durch die Flexibilität der verwendeten Schnittstelle bestimmt. Einschränkungen ergeben sich häufig durch die begrenzte lokale Speicherkapazität sowie im Falle von Struktur- oder Bedarfsänderungen durch den hohen Anpassungsaufwand bei vielen angeschlossenen Endbenutzergeräten. Dennoch kann ein derartiges Vorgehen zu akzeptablen Lösungen führen, insbesondere wenn die Endbenutzerrechner auch im mobilen Einsatz benötigt werden (etwa im Außendienst).

#### Virtuelles Data Warehousing

Ganz ohne separate Datenhaltung kommen Architekturkonzepte aus, die einen direkten Ad-Hoc-Durchgriff auf die operativen Datenbestände ermöglichen und dadurch die Bildung zusätzlicher multidimensionaler Datenpools in einer Unternehmung vermeiden (*Virtuelles Data Warehousing*). Leistungsstarke und grafisch-orientierte Abfragewerkzeuge, die bis auf die operativen Datenbasen durchgreifen können, versprechen hier besonders leichten Zugriff auf relevante Informationen. Eine spezielle Middleware sorgt für die korrekte Abwicklung von Benutzeranfragen (z. B. EDA/SQL).

Dem Vorteil dieser Lösung, dass direkt und ohne vielfältige konzeptionelle Vorüberlegungen sowie ohne redundante Datenhaltung auf die benötigten Daten zugegriffen werden kann, stehen allerdings gravierende Nachteile gegenüber. Zunächst ergeben sich große semantische Probleme. Die Bedeutung der oftmals unverständlichen Feldbezeichnungen operativer Datenbanksysteme nämlich kann nur über ein ausreichend dokumentiertes und vollständiges *Data Dictionary* erschlossen werden, das allerdings wohl eher selten in dieser Form vorliegt und erst mühsam erarbeitet werden muss. Die bei dieser Architekturform häufig erforderliche Nutzung von SQL als Datenbankabfragesprache durch den Endbenutzer kann zudem nur dann erfolgreich sein, wenn dieser sich auszeichnet mit den Sprachmechanismen auskennt, da ansonsten Datenbankabfragen leicht zu Ergebnissen führen, die fehlinterpretiert werden können.

Des Weiteren werden die operativen Systeme, deren originäre Aufgabe in der Abwicklung des Tagesgeschäftes zu sehen ist, durch zeitaufwändige und rechenintensive Abfragen

stark belastet, so dass diese Lösung nur in Ausnahmefällen akzeptabel sein dürfte (z. B. bei Zugriffen durch spezielle Anwendungen, bei denen es auf größte Aktualität ankommt). Da entsprechende Abfragen allerdings häufig im Rahmen eines Standard-Berichtswesens bereits durch die operativen Systeme abgedeckt werden, entfällt die Notwendigkeit zur Installation zusätzlicher Durchgriffe. Wenn überhaupt, dann lässt sich diese Architekturform darüber hinaus lediglich als Ad-Hoc-Informationsquelle für technisch geschulte Anwender sinnvoll einsetzen. Eine Nutzung des Datenmaterials im Sinne flexibler Navigation in mehrdimensionalen und hierarchisch organisierten Datenbeständen lässt sich auf diese Art allerdings kaum realisieren.

Bedingt durch die Probleme, die aus einem direkten Zugriff auf die operativen Datenbestände erwachsen, wird heute oftmals eine Architekturform bevorzugt, die auf separaten Datenbanksystemen basiert. Losgelöst von den operativen Datenbeständen verwalten diese Datenbanksysteme ein auf die Belange der betrieblichen Entscheidungsträger zugeschnittenes Informationsangebot. Häufig werden derartige Datenbanklösungen heute als *Enterprise Data Warehouse* oder auch als *Core Data Warehouse* bezeichnet. Da der Betrieb dieser *Data Warehouse*-Datenbasis von den operativen Systemen separiert erfolgt, entfällt das oben beschriebene Problem der Beeinträchtigung operativer Anwendungen. Im Zuge des Datenimports aus den angeschlossenen Vordatenbanken in diese Datenbank sind die Daten von Inkonsistenzen zu bereinigen und unter Umständen zu verdichten.

*Enterprise Data Warehouse*

Hierzu werden oftmals spezielle Zusatzwerkzeuge (ETL-Werkzeuge) eingesetzt, deren Funktionalität vielfältige Möglichkeiten zur Umwandlung und Aufbereitung von Daten umfasst. Die Datenablage erfolgt dann anwendungs- und auswertungsorientiert, d. h. losgelöst von den operativen Geschäftsabläufen sowie hinsichtlich relevanter Themen organisiert. Das in diesem zentralen *Data Warehouse* abgelegte Datenmaterial umfasst beliebige Verdichtungsstufen, die von aktuellen (z. B. Vortagsdaten) und detaillierten (z. B. Einzelartikel- und -kundendaten) Informationseinheiten bis zu stark verdichteten Kennzahlen reichen.

ETL-Werkzeuge

*Data Marts*

Allerdings erweist sich auch die Datenablage in einem zentralen, relationalen *Data Warehouse* als zu wenig performant, wenn es darum geht, umfangreiche Datenbestände rasch und flexibel zu analysieren. Häufig werden deshalb Datenextrakte zur weiteren Verarbeitung gebildet, die sich als personen-, anwendungs-, funktionsbereichs- oder problemspezifische Segmente des zentralen *Data Warehouse*-Datenbestandes verstehen lassen und als *Data Mart* bezeichnet werden (*Hub and Spoke*-Architektur). Auf diesen dezentralen Teildatenbeständen können dann anforderungsgerechte Untersuchungen erfolgen, z. B. indem mit Statistikpaketen Regressionsanalysen und Trendberechnungen durchgeführt oder aufwändige Verfahren zur Datenmustererkennung (*Data Mining*) auf der Basis statistischer Methoden oder neuronaler Netze angewendet werden.

Aufgrund ihres eingeschränkten Datenvolumens erweisen sich dezentrale *Data Marts* per se als geeigneter für die Durchführung anspruchsvoller Operationen auf den Datenbeständen. Dennoch kann auch hier eine Ad-Hoc-Befriedigung des variierenden Informationsbedarfs nur durch den Einsatz spezieller Komponenten gewährleistet werden. Welche Komponenten dabei zum Einsatz gelangen, hängt von der gewählten Datenbanktechnologie ab. Einerseits ist es möglich, als Speichertechnik auch für die *Data Marts* auf die verbreiteten relationalen Datenbanksysteme zurückzugreifen, andererseits werden alternativ in verstärktem Maße multidimensionale Datenbanksysteme eingesetzt.

Zusammenfassend lassen sich die unterschiedlichen Strukturkomponenten und Prozesse dann in einem Schaubild visualisieren (Abb. 2.1-1), wobei die Richtungen der verbindenden Pfeile die jeweiligen Datenflüsse darstellen. Zu beachten ist, dass die vorgestellten Architekturbestandteile eine grobe Einteilung der im Umfeld multidimensionaler Informationssysteme vorzufindenden Komponenten vornehmen. Auch lassen sich konkrete Softwarewerkzeuge nicht immer trennscharf einzelnen Komponentenblöcken zuordnen, sondern übernehmen zum Teil die Funktionen mehrerer Blöcke.



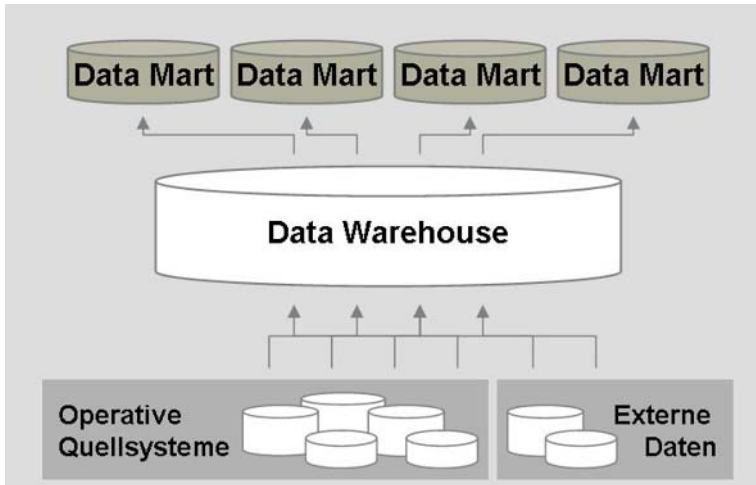


Abb. 2.1-1: *Hub and Spoke*-Architektur für die Anordnung von *Data Warehouse*- und *Data Mart*-Komponenten.

### 2.1.1.2 Prozesse zum Extrahieren, Transformieren und Laden von Daten \*\*

Das Befüllen eines *Data Warehouse* stellt ein schwieriges und komplexes Unterfangen dar, da die Datenbestände u. U. aus einer Reihe unterschiedlicher und häufig sehr heterogener Vorgesysteme stammen können. Als Datenlieferanten kommen neben den für die Abwicklung des Tagesgeschäftes zuständigen operativen Anwendungen auch Speziallösungen für Planung und Budgetierung sowie kommerzielle Informationsanbieter in Betracht. Bei den operativen Systemen ist weiterhin zwischen betriebswirtschaftlichen Standardlösungen und individuellen Anwendungen zu unterscheiden, da sich diese Systemkategorien in der Regel erheblich hinsichtlich ihrer Offenheit und damit Zugänglichkeit unterscheiden.

Grundsätzlich sind mit dem initialen Befüllen vor Produktivsetzung sowie dem periodischen Aktualisieren im Betrieb zwei unterschiedliche Anwendungsszenarien für die Datenübergabe an ein *Data Warehouse*-System gegeben.

- Initiales Befüllen** Bevor sich das System durch den betrieblichen Anwender sinnvoll nutzen lässt, sind die benötigten Datenstrukturen in der Zielumgebung anzulegen und erstmalig mit Problem-  
daten zu befüllen. Da sich das initiale Laden (*initial load*) als einmaliger, häufig nicht zeitkritischer Prozess erweist, können hier Vorgehensweisen mit geringem Automatisierungsgrad und weitgehend manueller Koordination gewählt werden. Wichtig ist dagegen, dass der Anwender bereits bei der ersten Nutzung des Systems auf einen breiten Informationspool zurückgreifen kann. Da eine wesentliche Analyseart in der Untersuchung längerer Zeitreihen besteht, müssen die Inhalte gegebenenfalls sogar aus Langzeitarchivsystemen extrahiert werden.
- Periodische Aktualisierung** Als vergleichsweise anspruchsvoll gegenüber dem erstmaligen Laden erweist es sich, einen Übernahmeprozess zu konzipieren und zu implementieren, der eine dauerhafte periodische Aktualisierung der *Data Warehouse*-Datenbasis ermöglicht. Da jede Übernahme sowohl auf der Seite der Vorgesysteme als auch in der multidimensionalen Zielumgebung Zeit kostet und Ressourcen verbraucht, sind in Abstimmung mit dem Anwender anforderungsgerechte und wirtschaftliche Übernahmezeitpunkte und -datenvolumina zu bestimmen.
- Kompletter Neuaufbau** Prinzipiell kann auch bei diesen periodischen Ladevorgängen der komplette Datenbestand einschließlich der zugehörigen Datenstrukturen jeweils neu aufgebaut werden. Diese Strategie erweist sich allerdings lediglich dann als sinnvoll, wenn das gesamte zu übernehmende Datenvolumen relativ gering ist und/oder weitreichende Strukturänderungen in den Daten der Vorgesysteme auch im Falle einer partiellen Datenübernahme zu erheblichen Reorganisationsmaßnahmen im multidimensionalen Datenspeicher führen würden. Einen besonderen Problemkreis bildet bei dieser Vorgehensweise die Behandlung der Informationsobjekte, die zwischenzeitlich in den Vorgesystemen gelöscht oder archiviert worden sind, bei einer Übernahme also folglich keine Berücksichtigung finden. Da multidimensionale Informationssysteme häufig eine Betrachtung langer Zeiträume gewährleisten sollen, müssen für diese Objekte gesonderte Verfahren gefunden werden, um den zugehörigen Informationsgehalt nicht zu verlieren.

Häufiger wird dagegen bei der periodischen Datenübernahme eine inkrementelle Vorgehensweise (*incremental load*) einzuschlagen sein, die nur jene Problemdaten berücksichtigt, die sich seit der letzten Datenübernahme geändert haben bzw. neu hinzugekommen sind.

Inkrementelle  
Aktualisierung

Bevor ein Transformationskonzept implementiert werden kann, sind zusammen mit den Endanwendern die zu übernehmenden Dateninhalte sowie die Häufigkeit und die Aktualität des Datentransfers zu diskutieren. Vor allem die Granularität und Breite der Dateninhalte hat erheblichen Einfluss sowohl auf Datenvolumen in der Zielumgebung als auch auf die Komplexität und Ressourcenintensität des Transformationsprozesses.

Granularität

Die Häufigkeit und Aktualität der Datenübernahme werden primär durch die Granularität der Zeitdimension im multidimensionalen System bestimmt. Häufig finden sich hier Tages-, Wochen- oder Monatsgrößen. Tagesdaten werden oftmals nachts überspielt, vor allem um die operativen Verarbeitungsaufgaben der datenliefernden Systeme nicht zu beeinträchtigen. Eine Übernahme von Monatsdaten kann gegebenenfalls auch am Wochenende vorgenommen werden. In jedem Fall ist dafür Sorge zu tragen, dass das verfügbare Zeitfenster für die Datenübernahme ausreicht, zumal hier auch andere Aufgaben erfolgen müssen, wie beispielsweise Backup- oder Replikations-Läufe, die nicht gestört werden dürfen.

Jede Datenübernahme erweist sich als Sammlung unterschiedlicher Bearbeitungs- und Transportschritte, die teilweise aufeinander aufbauen. Nur wenn alle Schritte durchlaufen und erfolgreich abgeschlossen werden, liegen die Daten anschließend im multidimensionalen Informationsspeicher in der gewünschten Form und Qualität vor. Insofern erscheint es angebracht, von einem Prozess der Datenübernahme zu sprechen. Teilweise wird die Aktionskette auch als Transformationsprozess bezeichnet, um herauszustellen, dass hierbei neben dem reinen Austausch von Daten zwischen verschiedenen Speicherkomponenten auch vielfältige Umwandlungsoperationen durchzuführen sind. In idealisierter Form läuft der Prozess ab, wie in der Abb. 2.1-2 dargestellt.

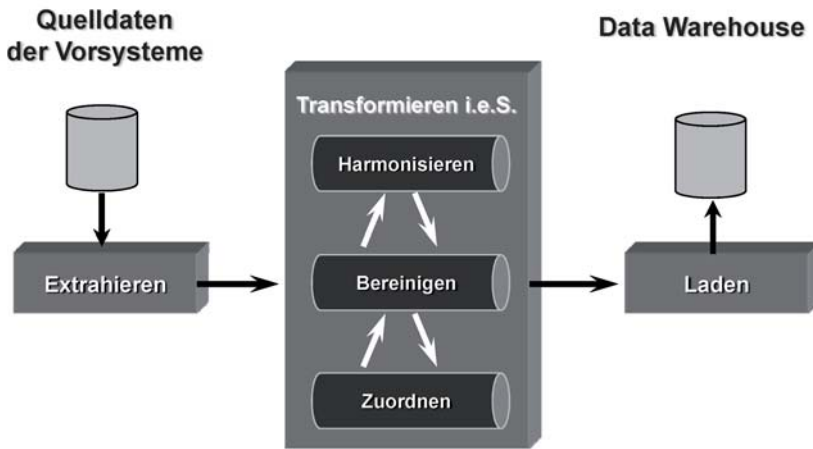


Abb. 2.1-2: ETL-Prozess zur Befüllung analyseorientierter Datenspeicher.

## 2.1.2 OLAP – *On-Line Analytical Processing* \*

Charakterisierung OLAP

Mit dem Konzept des *OLAP* scheint ein Leitbild gegeben zu sein, das teils aus fachlicher teils auch aus systemtechnischer Perspektive die Aspekte hervorhebt, die für eine anforderungsgerechte Nutzung entsprechender Systeme unabdingbar sind. Demgemäß repräsentiert *On-Line Analytical Processing* eine Softwaretechnik, die Managern wie auch qualifizierten Mitarbeitern aus den Fachabteilungen schnelle, interaktive und vielfältige Zugriffe auf relevante und konsistente Informationen ermöglichen soll. Im Vordergrund stehen dabei dynamische und multidimensionale Analysen auf historischen, konsolidierten Datenbeständen. Durch die gewählte Begrifflichkeit werden OLAP-Systeme bewußt von OLTP-Systemen abgegrenzt, die transaktionsorientiert die Abwicklung der operativen Geschäftstätigkeit unterstützen.

Multidimensionalität

Als zentrales Charakteristikum von OLAP wird zumeist die Multidimensionalität gewertet. Diese zielt auf eine Anordnung betriebswirtschaftlicher Variablen bzw. Kennzahlen (wie z. B. Umsatz- oder Kostengrößen) entlang unterschiedlicher Dimensionen (wie z. B. Kunden, Artikel, Niederlassungen oder Regionen), wie sie spätestens seit der Etablierung

von EIS (*Executive Information Systemen*) als geeignete Sichtweise für das Management auf betriebswirtschaftliches Zahlenmaterial akzeptiert ist. Versinnbildlicht erscheinen die quantitativen Größen dann als Sammlung von Würfeln, wobei die einzelnen Dimensionen durch entsprechend textindizierte Würfelkanten verkörpert werden.

Als besondere Anekdote der DV-Geschichte ist der Umstand zu verstehen, dass es mit E.F. Codd einer der geistigen Urväter der relationalen Datenbanken war, der dem Begriff des *On-Line Analytical Processing* Leben einhauchte. Unbestritten erweisen sich relationale Technologien als stabile, sichere und schnelle Alternative bei der Datenspeicherung in ungezählten transaktionsorientierten Anwendungssystemen. Allerdings offenbaren relationale Datenbanken Schwächen bei der flexiblen und benutzeradäquaten Zurverfügungstellung entscheidungsrelevanter Informationen für das Management. Schließlich, so argumentiert Codd, sind relationale Systeme auch nicht darauf ausgelegt, multidimensionale Analysen mit der geforderten Funktionalität und in der gewünschten Schnelligkeit zu gewährleisten. Als Konsequenz definieren Codd et al. zwölf *Evaluationsregeln*, die bei Erfüllung die OLAP-Fähigkeit der Informationssysteme garantieren sollen.

12  
Evaluierungs-  
regeln

Die Evaluationsregeln werden im Folgenden aufgegriffen und detailliert vorgestellt:



- »Die zwölf OLAP-Evaluierungsregeln«, S. 54

Die multidimensionale Anordnung von betriebswirtschaftlichem Datenmaterial in Datenwürfeln wird einer eingehenden Betrachtung unterzogen:

- »Multidimensionalität durch die Verwendung von Datenwürfeln«, S. 57

Bei der Implementierung von OLAP-Systemen sind alternative Realisierungsformen möglich:

- »Speicherkonzepte für OLAP-Lösungen«, S. 59

Für die Endanwender stehen die Funktionalitäten zur Navigation im Vordergrund, die letztlich durch die eingesetzten Front-End-Werkzeuge bestimmt werden:

- »Navigation in multidimensionalen Datenstrukturen«, S. 60
- »Frontend-Techniken und -Funktionen«, S. 61

### 2.1.2.1 Die zwölf OLAP-Evaluierungsregeln \*

Es gibt zwölf Evaluierungsregeln, die eine Liste idealtypischer Anforderungen an OLAP-Systeme darstellen.

#### 1. Mehrdimensionale konzeptionelle Sichten

Die konzeptionelle Sicht der OLAP-Modelle muss, bezogen auf die Problemsicht des Benutzers, mehrdimensional sein. Bezugsgrößen, wie etwa Kunden, Artikel oder Regionen, bilden die Dimensionen, entlang derer sich die betriebswirtschaftlichen Kenngrößen (z. B. Umsatz) im zeitlichen Verlauf untersuchen lassen. Die Abfragetechnik von OLAP-Systemen besteht aus Manipulationen des mehrdimensionalen Datenraums. Durch die Auswahl eines bestimmten Abfrageergebnisses wird die Extraktion beliebiger Aggregate aus dem originären Datenbestand veranlasst. Der Benutzer muss sich dabei frei in dem **Datenwürfel** bewegen und an beliebiger Stelle Schnitte durch den Würfel ziehen bzw. den Datenwürfel drehen können, um aus verschiedenen Perspektiven auf den Datenbestand blicken zu können. Dies ermöglicht dem Entscheider, Informationen zu vergleichen und selbständig Berichte erstellen zu können.

#### 2. Transparenz

OLAP-Werkzeuge sollen sich problemlos in die offene Architektur der Arbeitsumgebung einbetten lassen. Das Ziel ist eine möglichst homogene Benutzungsoberfläche mit allen notwendigen Funktionalitäten. Dadurch soll erreicht werden, dass der Benutzer keinen formalen Unterschied zwischen Informationseinheiten aus unterschiedlichen Datenquellen ausmachen kann.

#### 3. Zugriffsmöglichkeit

Der Zugriff auf möglichst viele heterogene unternehmensinterne und externe Datenquellen und -formate soll durch eine offene Architektur der OLAP-Systeme ermöglicht werden. Diese Daten bilden die Basis eines umfassenden analy-

tischen Datenmodells und müssen daher sorgfältig konvertiert werden, damit eine einheitliche, konsistente Datensicht für den Benutzer gewährleistet ist.

#### 4. Gleichbleibende Berichtsperformance

Stabile Antwortzeiten und gleichbleibende Berichtsleistung bei Datenabfragen sind zwei wichtige Punkte für die Nutzung des OLAP-Systems. Auch bei einer überproportionalen Zunahme der Dimensionen bzw. des Datenvolumens sollte sich die Antwortzeit nicht signifikant ändern.

#### 5. Client-Server-Architektur

OLAP-Systeme sind derart zu gestalten, dass sie den Einsatz von **Client-Server-Architekturen** unterstützen. Dabei sollte sowohl eine verteilte Programmausführung als auch eine verteilte Datenhaltung möglich sein. Die verteilten Datenquellen sollten beliebig integriert und aggregiert werden können. Das *Data Warehouse*-Konzept bietet hierbei eine gute Problemlösung.

#### 6. Gleichgestellte Dimensionen

Alle Dimensionen sind als gleichwertig anzusehen. Es soll ein einheitlicher Befehlsumfang zum Aufbauen, Strukturieren, Bearbeiten, Pflegen und Auswerten der Dimensionen vorhanden sein.

#### 7. Dynamische Verwaltung dünn besetzter Matrizen

Multidimensionale Datenmodelle enthalten dünn besetzte Matrizen. Das begründet sich daraus, dass nicht alle denkbaren Kombinationen der Dimensionselemente werttragende Verbindungen eingehen. Durch eine optimale Datenspeicherung müssen diese Lücken in den Datenwürfeln effizient gehandhabt werden, ohne die multidimensionale Datenmanipulation zu beeinträchtigen.

#### 8. Mehrbenutzerfähigkeit

Der Mehrbenutzerbetrieb ist eine grundlegende Anforderung an ein OLAP-System. Das heißt, es müssen mehrere Benutzer gleichzeitig auf den Datenwürfel zugreifen können,

um ihre Analysen durchzuführen. Dazu gehört aber auch ein Sicherheitskonzept, das den Datenzugriff und die Datenverfügbarkeit der verschiedenen Benutzer regelt.

## 9. Unbeschränkte kreuzdimensionale Operationen

Zur Datenanalyse (z. B. Kennzahlenberechnungen oder Konsolidierungen) werden Operationen über die Dimensionen hinweg benötigt. Dafür sind eine vollständige, integrierte **Datenmanipulationssprache** sowie eine unbeschränkte Abfragemöglichkeit notwendig.

## 10. Intuitive Datenmanipulation

Eine einfache und ergonomische Benutzerführung und Benutzungsoberfläche erleichtern das intuitive Arbeiten an der Datenbasis. Exemplarisch soll hier die verständliche Adressierung der Daten im multidimensionalen Raum und ein einfacher **Drill-Down** in weitere Detaillierungsebenen genannt werden. Dem Benutzer müssen somit Zugriffe auf die Basiselemente der Dimensionen und Werkzeuge zur beliebigen Zusammenstellung neuer Konsolidierungsgruppen zur Verfügung gestellt werden.

## 11. Flexibles Berichtswesen

Aus dem multidimensionalen Modell müssen leicht und flexibel aussagekräftige Berichte hergeleitet werden können. Diese sollen, je nach den Anforderungen des Benutzers, als tabellarische (**Ad-hoc**)-Berichte oder dynamische Grafiken zur Verfügung stehen. Das OLAP-Werkzeug soll darüber hinaus dem Benutzer derart Unterstützung bieten, dass Daten flexibel in beliebiger Form bearbeitet, analysiert und betrachtet werden können.

## 12. Unbegrenzte Dimensions- und Aggregationsstufen

Ein OLAP-System soll eine unbegrenzte Anzahl an Dimensionen, Relationen und Variablen innerhalb der einheitlichen Datenbank ermöglichen. Weiterhin soll keine Limitie-



zung der Anzahl und Art der Aggregationen der Daten bestehen.<sup>1</sup>

Die zwölf aufgestellten Anforderungen an OLAP-Systeme sind z. T. sehr heftig kritisiert worden. Grundsätzlicher Angriffspunkt ist die unscharfe Trennung zwischen fachlich-konzeptionellen Anforderungen und technischen Realisierungsaspekten. So bleibt etwa unklar, ob die konzeptionelle mehrdimensionalen Datensichten auch die zwingende Nutzung spezieller Speicher- und Datenverwaltungstechniken impliziert, wie sie durch die mehrdimensionalen Datenbanken abgedeckt werden. Zudem wurden von unterschiedlichen Produkthanbietern Sinnhaftigkeit und Notwendigkeit einzelner Forderungen bestritten, nicht zuletzt, weil deren Produkte eine abweichende Funktionalität aufweisen. So wird insbesondere die Regel 6 angegriffen, die die Dimensionen eines mehrdimensionalen Modells gleichstellt. Bestimmte Dimensionen jedoch – wie z. B. die Zeitdimension mit ihrer inhärenten Zeitlogik, für die gar eine kompilierte Zeitintelligenz gefordert wird – unterscheiden sich erheblich von den übrigen Dimensionen.

Kritik

### 2.1.2.2 Multidimensionalität durch die Verwendung von Datenwürfeln \*

**Auf der Speicherebene kann die Struktur eines multidimensionalen Datenmodells als Sammlung von Datenwürfeln verstanden werden, wobei die einzelnen Elemente der Dimensionen eines Würfels oftmals hierarchische Verknüpfungen aufweisen.**

Die Organisation multidimensional gestalteter Inhalte erfolgt in Form so genannter Datenwürfel (*Cubes*). Ein Würfel besteht aus der logischen Anordnung quantitativer Größen bzw. betriebswirtschaftlicher Variablen (z. B. Kennzahlen) entlang sachlich zusammengehöriger Beschreibungsobjekte (z. B. Produkte, Zeit oder Region). Gruppierungen dieser Objekte, die dann als Dimensionen bezeichnet werden, finden sich als Kanten des Würfels wieder [GaGl98]. In der betrieblichen Umsetzung sind derartige Würfel nicht auf drei Dimensionen beschränkt, sondern weisen häufig sechs, acht

Datenwürfel

<sup>1</sup>Unter DV-technischen Gesichtspunkten ist eine unbegrenzte Anzahl an Elementen natürlich nicht realisierbar. Besser wäre vielleicht der Begriff »ausreichend«.