



Leseprobe

Datenintegration

Integrationsansätze, Beispielszenarien, Problemlösungen, Talend Open Studio

Herausgegeben von Ines Rossak

ISBN (Buch): 978-3-446-43221-5

ISBN (E-Book): 978-3-446-43491-2

Weitere Informationen oder Bestellungen unter

<http://www.hanser-fachbuch.de/978-3-446-43221-5>

sowie im Buchhandel.

2

Grundlagen der Datenintegration

In diesem Kapitel wird nach einer kurzen Einführung in das Thema am Beispiel typischer Anwendungsfelder verdeutlicht, dass Datenintegration in einer Vielzahl von Anwendungsfeldern zum Einsatz kommt. Es werden die Herausforderungen beschrieben, die sich ergeben, wenn Datenbestände isoliert voneinander verwaltet, aber gemeinsam ausgewertet werden sollen. Ausgehend vom klassischen Schichtenmodell wird dargestellt, auf welchen Ebenen Integration in Angriff genommen werden kann, welche Architekturen sich dafür bewährt haben und welche Aufgaben zu lösen sind.

■ 2.1 Einleitung

Die meisten Unternehmen sind heute durch die Risiken und den Wettbewerbsdruck globalisierter Märkte sowie den Zwängen durch gesetzliche Bestimmungen usw. unter enormen Druck, zeitnah über Ablauf und Ergebnisse aller geschäftlichen Prozesse Bescheid zu wissen. Sie müssen insbesondere Risiken wie z. B. drohenden Zahlungsausfall oder Störungen und Verzögerungen in der Lieferkette durch unzuverlässige oder nicht mehr verfügbarer Zulieferer schnell erkennen können. Ebenso zügig müssen Unternehmen auf geänderte Situationen reagieren können wie z. B. eine plötzliche Nachfrage nach einem bestimmten Produkt, Abwanderung von Kunden, Preisschwankungen auf den Rohstoffmärkten usw.

Dies setzt voraus, dass vorhandene, oft über verschiedene Fachabteilungen und Standorte verteilte Datenbestände übergreifend ausgewertet werden können, um eine Gesamtsicht auf die gesamte Wertschöpfungskette zu erhalten.

Jedes Unternehmen, jede Institution, jede Organisation verfügt über enorme Datenmengen. In den meisten Unternehmen haben Umstrukturierungen, Übernahmen, Unternehmensfusionen oder -teilungen zu einer über die Jahre hinweg historisch gewachsenen fragmentierten, heterogenen Systemlandschaft geführt, die auch eine Vielzahl von isolierten Dateninseln umfasst.

In einem durchschnittlichen Unternehmen ist heute eine kaum noch überschaubare Menge unterschiedlichster Anwendungssysteme im Einsatz, die für bestimmte Geschäftsprozesse oder Aufgabenbereiche optimiert sind. Am häufigsten finden sich naturgemäß Systeme, die Kunden-, Produkt-, oder Lieferanteninformationen enthalten. Diese werden ergänzt durch Lagerhaltungs-, Logistik-, Finanzsysteme und viele andere mehr. Laut einer

IBM-Studie von 2007 [Cor07] sind in einem mittleren Unternehmen etwa 50 unterschiedliche Systeme im Einsatz, eine andere Studie von Accenture verzeichnet bis zu 200 unterschiedliche ERP-Systeme in großen, global agierenden Unternehmen. Informatica spricht sogar von einer regelrechten *Anwendungsschwemme* [Inf11].

Hinter diesen Systemen liegen in der Regel für den jeweiligen Zweck entworfene und optimierte Datenbanken. Die darin enthaltenen Daten sind häufig *redundant*, oft sogar aus anderen Systemen *repliziert*, denn die meisten Geschäftsprozesse benötigen ähnliche oder sogar gleiche Daten. So benötigt das Marketing für Mailingaktionen die Kundenadressen, ebenso die Kundenbetreuung, die Stammkunden mit einer Sonderaktion belohnen möchte. Auch die Logistik, die Waren an Kunden zustellt, muss die Kundenadresse kennen oder die Buchhaltung, die eine Mahnung schicken muss. Eine von Omikron 2010 beauftragte Studie [OMI12], in der 339 Unternehmen zum Umgang mit Kundendaten befragt wurden, kommt zum Ergebnis, dass bei 60 % der befragten Unternehmen zwischen drei und zehn separate Kundendatenbanken bestehen. Die Hoheit über diese Daten liegt bei 52 % der Befragten im jeweiligen Geschäftsbereich.

Diese Art der *isolierten Datenhaltung* führt nicht nur zu redundanten, sondern naturgemäß auch widersprüchlichen Daten, da ein Datenabgleich oder -austausch nicht kontinuierlich oder gar nicht stattfindet. Selbst ursprünglich identische Datenbestände beginnen im Laufe der Zeit auseinanderzudriften, wenn sie ohne unternehmensweit geltende Standards autonom verwaltet werden. So entstehen Informationsinseln mit eigenen Strukturen, Kodierungen, Formaten, Standardwerten, semantischen Bedeutungen usw. (siehe auch Abschnitt 2.3).

Dazu kommt die ständig weiter wachsende Datenflut durch die zunehmende Zahl von Kanälen, über die Daten generiert werden sowie immer mehr proprietäre Formate. Schätzungen gehen davon aus, dass die Datenbestände um mindestens 50 % pro Jahr wachsen [IDC11]. Andere Schätzungen sprechen sogar von 64 % pro Jahr [Inf11]. Dies macht es oft unerlässlich, einen Teil der Altdaten auszulagern. So werden historische Daten häufig getrennt von operativen Daten für das Tagesgeschäft gespeichert und müssen für Analysen, die größere Zeiträume umfassen sollen, oft mühsam wiederhergestellt werden.

Diese Ausgangslage erschwert es, Prozesse über Bereichs- oder Standortgrenzen oder größere Zeiträume hinweg verlässlich zu analysieren, da Daten einbezogen werden, die sich in ihrer Aktualität, Detailliertheit, Genauigkeit, Zuverlässigkeit usw. unterscheiden können. Daraus abgeleitete geschäftliche Entscheidungen sind oft wenig transparent und mitunter nicht nachvollziehbar. Das Risiko falscher Entscheidungen wächst, finanzielle, juristische und Imageprobleme können die Folge sein.

So wird z. B. in einer Postwurfsendung ein bestimmter Artikel beworben, der nur kurze Zeit im Angebot ist. Es gab im Vorfeld keine Analyse des Einzugsgebietes und der Kundenstruktur jeder Filiale, daher bekommen alle Filialen die gleiche Anzahl des Artikels. Zum Verkaufsstart tritt die bekannte Situation auf, dass in einigen Filialen der Artikel nach wenigen Minuten ausverkauft ist, in anderen gibt es noch nicht einmal eine Nachfrage. Da es keine zentrale Warenwirtschaft gibt, bleiben in der einen Filiale verärgerte Kunden zurück und in der anderen eine Menge unverkaufter Artikel. Eine Empfehlung für die Kunden, die andere Filiale aufzusuchen oder den Artikel von dort zu beschaffen, gibt es aus Unkenntnis der Situation nicht. Zum entgangenen aktuellen Umsatz kommen möglicherweise zukünftige Umsatzverluste, weil die verärgerten Kunden künftig woanders kaufen. Ähnlich ärgerlich bzw. nachteilig sind fehlgeleitete oder doppelte Sendungen oder nicht zustellba-

re Rechnungen, weil es widersprüchliche Adressdaten gibt, Lieferung falscher Artikel, weil Produktdaten unterschiedlich abgelegt sind, mehrfache Beschaffung von Teilen, weil sie im Lager nicht mehr auffindbar sind usw.

Ein wichtiger Schlüssel für die Lösung dieser Probleme ist die *Konsolidierung und Standardisierung* der gesamten IT-Infrastruktur, insbesondere der Datenhaltung. Für große Unternehmen hat das höchste Priorität [Mad09].

Das Kernstück aller Konsolidierungsvorhaben ist die *Datenintegration*, denn keine Anwendung kommt ohne Daten aus. Die Anwendungsfelder für Datenintegration sind außerordentlich vielfältig und umfassen deutlich mehr Bereiche als nur Data Warehousing. In den folgenden Abschnitten werden daher einzelne Aspekte der Datenintegration näher betrachtet.

Der Abschnitt 2.2 betrachtet zwei typische Anwendungsbereiche von Integration – Konsolidierung für operative und Konsolidierung für analytische Zwecke.

Im Abschnitt 2.3 werden die Herausforderungen beschrieben, die sich aus der Verteilung und Autonomie der Datenbestände ergeben.

Im Abschnitt 2.4 wird Integration ausgehend vom klassischen Schichtenmodell betrachtet, d. h. die Betrachtung von Integration auf Daten-, Anwendungs-, bzw. Präsentationsebene.

Im Abschnitt 2.5 werden die daraus resultierenden verschiedenen Integrationsarchitekturen beschrieben und gegeneinander abgegrenzt.

Im Abschnitt 2.6 werden typische Aufgaben, die bei einer Integration zu lösen sind und eine bewährte Schrittfolge zur Abarbeitung beschrieben.

■ 2.2 Typische Anwendungsfelder

Datenintegration wird häufig gleichgesetzt mit *Data Warehouse*, spielt aber auch in anderen Bereichen des Datenmanagements wie z. B. der *Migration* von Daten aus Altsystemen, der *Replikation* und *Synchronisation* von Daten in redundanten Systemen, dem *Datenqualitätsmanagement* oder dem *Master Data Management* eine wichtige Rolle [BAR09]. Zentrale Aufgabe in all diesen Bereichen ist die *Konsolidierung* der Daten mit dem Ziel, möglichst vollständige, einheitliche und korrekte Datenbestände zur Verfügung zu stellen.

Stellvertretend für die vielfältigen Einsatzmöglichkeiten sollen hier einige Anwendungsfelder aus zwei großen Kernbereichen des Datenmanagements herausgegriffen werden – die *Konsolidierung im operativen Bereich* und die *Konsolidierung im analytischen Bereich*.

Im operativen Bereich wird Datenintegration überall dort benötigt, wo operative Daten, die aus verschiedenen Gründen verteilt vorliegen, aus praktischen Gründen in einer Datenbank vereinigt werden sollen.

Im analytischen Bereich steht die Auswertung von Daten im Vordergrund mit dem Ziel, einen hohen Grad an Wissen aus den verfügbaren Informationen zu generieren und dieses so aufzubereiten und darzustellen, dass es schnellstmöglich erfasst und für operative und strategische Entscheidungen genutzt werden kann.

2.2.1 Konsolidierung im operativen Bereich

Es gibt verschiedene Ursachen, die eine Konsolidierung von Daten im operativen Bereich notwendig machen. An dieser Stelle sollen drei typische Szenarien kurz dargestellt werden.

Fusionierungen und Umstrukturierungen

Die Fusion zweier Unternehmen oder Umstrukturierungen innerhalb des Unternehmens sind relativ häufig anzutreffende Szenarien, wenn beispielsweise mehrere Standorte oder Abteilungen existieren, die bisher unabhängig voneinander agierten und aus wirtschaftlichen Gründen zusammengeführt werden sollen oder andere Unternehmen mit den gleichen Kerngeschäften übernommen bzw. strategische Partnerschaften eingegangen werden. Die zu fusionierenden Unternehmen oder Unternehmensteile haben in der Regel historisch gewachsene Datenbanken auf eigenen IT-Systemen, manchmal mit völlig verschiedenen, oft aber mit vergleichbaren Strukturen und Inhalten.

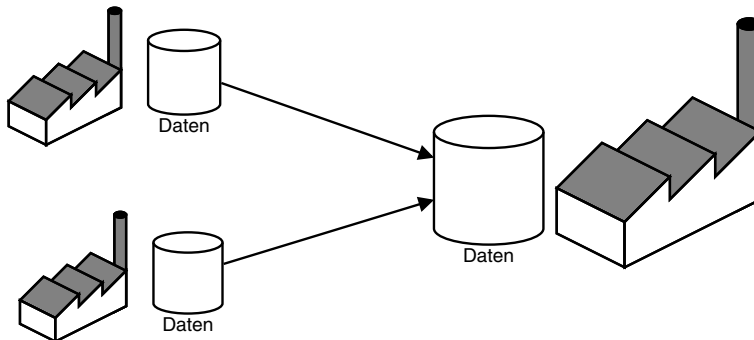


BILD 2.1 Fusionierung

Eigentlich zusammengehörnde Daten liegen oft verteilt und in uneinheitlichen Formaten vor (siehe Abschnitt 2.3). Auch *Flatfiles* wie Comma-Separated Values (CSV)-Dateien oder Excel-Tabellen, in denen zusätzliche Informationen hinterlegt werden, sind keine Seltenheit. Ziel sollte eine Kompletierung des Datenbestandes sein, so dass alle zusammengehörigen Daten vollständig und einheitlich vorliegen [Sch05], [KR05].

Um die sich aus der Fusionierung ergebenden Synergien effektiv zu nutzen, die IT-Landschaft zu verschlanken und/oder zu homogenisieren, wird fast immer entschieden, die ursprünglichen Systeme nicht parallel weiter zu betreiben. Vielmehr wird entweder eines der Systeme zum führenden System erklärt und die Daten aus dem anderen System werden in dieses überführt oder es wird ein ganz neues Zielsystem definiert, in das die Daten der Altsysteme integriert werden. Ein praxisnahes Anwendungsbeispiel für den ersten Fall wurde für dieses Buch konzipiert und soll später ausführlich behandelt werden.

Anwendungsmodernisierung

Ein weiteres typisches Szenario ist die parallele Existenz und Nutzung von Neu- und Legacy-Anwendungen (dt. Altlast, Erbe). Im Laufe der Jahre wächst die IT-Landschaft in vielen Unternehmen enorm. Es werden neue Systeme eingeführt, jedoch erscheint es nicht

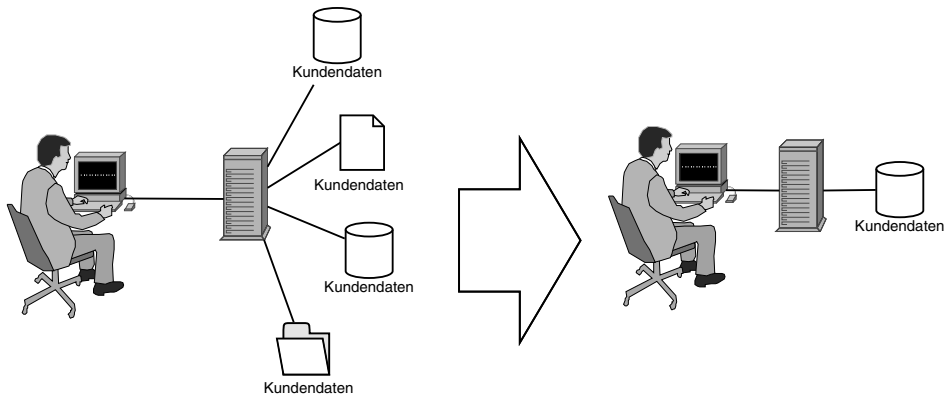


BILD 2.2 Verteilte Daten

immer sofort als erforderlich, die Daten bisheriger Programme in die neuen Anwendungssysteme zu integrieren. Da aber in bestimmten Fällen neben aktuellen auch Altdaten benötigt werden, werden Legacy-Anwendungen nicht abgeschaltet, sondern parallel weiter betrieben. Mitarbeiter, die lange im Unternehmen beschäftigt sind, tendieren erfahrungsgemäß ohnehin dazu, bestehende Arbeitsabläufe (und damit oft auch gewohnte Anwendungen) beizubehalten. Dadurch kommt es zu einer Parallelbenutzung von verschiedenen Systemen mit der Gefahr, dass benötigte Daten an verschiedenen Stellen bearbeitet, aber nicht über alle Systeme hinweg konsistent gehalten werden. Fehlende oder widersprüchliche Informationen können die Folge sein. Zudem erhöht sich durch den Parallelbetrieb auch der Wartungs- und Supportaufwand (und damit die Kosten) für die IT-Abteilung.

Bei der Einführung von neuen Anwendungssystemen muss deshalb frühzeitig geklärt werden, welche Alt-Systeme abgelöst werden sollen und wie mit den damit verbundenen Datenbeständen vefahren werden soll. Diese Probleme können entweder auf Anwendungsebene gelöst werden, wie z. B. durch die Möglichkeiten wie sie Enterprise Application Integration (EAI) bietet (vgl. dazu auch Abschnitt 2.5.2) oder auf Datenebene selbst, indem die verwendeten Datenbanken fusioniert werden. Die Überführung der Altdaten muss strategisch geplant und zeitnah umgesetzt werden. Sobald sichergestellt ist, dass alle Daten vollständig und korrekt in das neue System integriert wurden, sollten Legacy-Anwendungen aus dem täglichen Betrieb entfernt werden [Han09, S.2], [Tie06, S.10].

Master Data Management – MDM

Bei global agierenden Unternehmen, aber auch bei Unternehmen mit heterogenen Unternehmens- und IT-Strukturen werden Daten zu gleichen oder ähnlichen Sachverhalten in den Fachabteilungen an verschiedenen Standorten in einer Vielzahl von Anwendungssystemen bereitgestellt. So muss z. B. in vielen Unternehmensbereichen (Sales, Marketing, Finance, Logistic usw.) mit Kundendaten gearbeitet werden. Diese Daten werden jedoch im Allgemeinen mit speziellen Anwendungen verwaltet (CRM-Systeme, ERP-Systeme usw.), die den Kunden ganz unterschiedlich abbilden. Üblicherweise unterscheiden sich diese Daten auch in ihrer Beschaffenheit, Qualität, Aktualität und Genauigkeit, da sie in der Regel für bestimmte Zwecke erfasst und strukturiert wurden. Für standort- und/oder abteilungs-

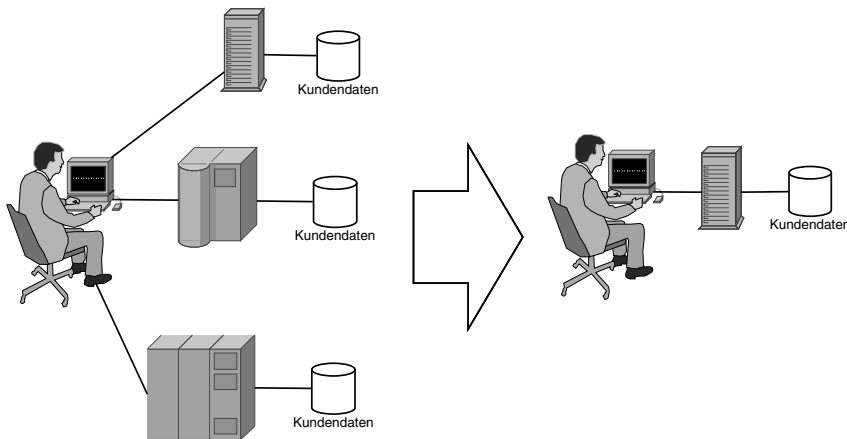


BILD 2.3 Legacy Anwendungen

übergreifende Geschäftsprozesse müssen diese Daten aus den IT-Systemen der jeweiligen Geschäftsbereiche aufwendig extrahiert und homogenisiert werden.

Daten, die standort-, abteilungs-, oder auch prozessübergreifend zur Verfügung stehen müssen, werden im Allgemeinen als Stamm- oder Masterdaten bezeichnet. Das Management solcher Masterdaten rückt in den letzten Jahren immer mehr in den Fokus der Unternehmen. Bei Umfragen haben über 90% der Unternehmen erklärt, bereits Master Data Management (MDM)-Lösungen einzusetzen oder für die nächste Zeit zu planen. Ziel des MDM ist die zentrale Verwaltung von Stammdaten bzw. Referenzdaten, um system- und anwendungsübergreifende Konsistenz sicherzustellen, ohne immer neue Integrationsprojekte zu starten.

Für ein vernünftiges MDM müssen zunächst die relevanten Systeme identifiziert werden, die Stammdaten liefern oder empfangen. Aus den gelieferten Datenmengen sollte eine Schnittmenge gebildet werden, um die für das MDM-System sinnvollen Daten zu finden. Diese Masterdaten müssen aus den relevanten Systemen extrahiert und nach vorab zentral definierten Regeln und Standards in einer Master-Datenbank integriert und konsolidiert werden. Herausforderungen, die in MDM-Projekten zu meistern sind, sind jedoch nicht in erster Linie technischer Art (die Integrationsmethoden sind vergleichbar mit denen von Data Warehouse Projekten), sondern eher organisatorischer Art. Diese und weitere Herausforderungen werden u. a. bei [Los08] und [HOW11] vertiefend behandelt und durch zahlreiche Fallstudien und Anwendungsszenarien ergänzt.

2.2.2 Konsolidierung im analytischen Bereich

Das wichtigste Ziel im analytischen Bereich ist die Transformation von Informationen in Wissen. Je umfangreicher und je zuverlässiger Informationen verknüpft werden können, desto fundierteres Wissen lässt sich daraus generieren, um operative und strategische Entscheidungen ziel- und bedarfsgerecht durchzuführen. Um dieses Ziel zu erreichen, müssen verfügbare Daten aus verschiedenen Quellen integriert, angereichert, aufbereitet und dargestellt werden. Das große Gebiet, unter dem sich diese Aufgaben zusammenfassen lassen,

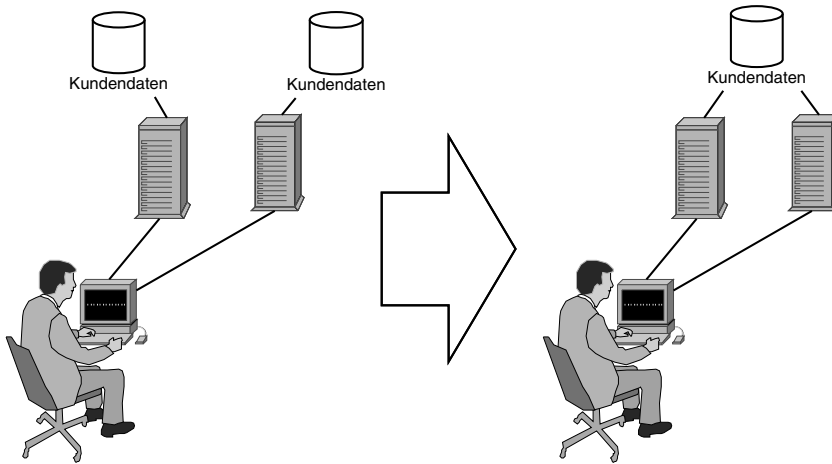


BILD 2.4 Gemeinsame Datenbasis

wird im Allgemeinen als *Business Intelligence* (BI) bezeichnet. Wie für viele Anwendungen im operativen Bereich bildet Datenintegration auch für BI-Anwendungen das Rückgrat. Aussagekräftige und verlässliche Analysen, Vorhersagen, Berichte usw. sind nur auf der Basis konsolidierter Datenbestände möglich.

KONZEPT	METHODE	BEISPIEL
Reporting	Dashboards Score Cards	Übersicht potentielle Kündiger
OLAP	Data Cubes	Kündiger pro Quartal pro Region
Mining	Klassifikation Regression Clusterbildung Assoziation	Charakteristik Kündiger-Nichtkündiger Vorhersage Kündigungs- wahrscheinlichkeit

BILD 2.5 Business Intelligence

OLAP und Data Mining

Für Analysen findet in der Regel ein Data Warehouse (DWH) Verwendung, da nicht nur aktuelle, sondern auch historische Daten einen hohen Informationsgehalt besitzen. Daten aus verschiedenen internen und externen Quellen werden mittels verschiedener Techniken (siehe dazu auch Abschnitt 2.5.1) in ein DWH integriert und dort langfristig gespeichert. In Abhängigkeit von geplantem Verwendungszweck und -ziel wird dieses in festgelegten Zeitintervallen durch aktuelle Daten ergänzt und ggf. mit weiteren externen Daten angereichert.

Die in einem DWH gesammelten Daten können dann mit unterschiedlichen Analysemethoden genauer untersucht werden. Typischerweise werden Operationen und Verfahren angewendet, die sogenannte *gerichtete (OLAP)* bzw. *ungerichtete (Data Mining)* Analysen ermöglichen. Durch die mittels *OLAP* und *Data Mining* gewonnenen Erkenntnisse können Unternehmen ihre Wertschöpfungskette verbessern, indem Kosten gesenkt, Risiken minimiert und Beziehungen zu Kunden und Lieferanten verbessert werden.

Online Analytical Processing (OLAP) ist eine der Schlüsselkomponenten von BI. Auf der Basis speziell aufbereiteter und strukturierter Datenbestände (Basis sind *multidimensionale Datenmodelle*) können Operationen eingesetzt werden, die es erlauben, durch sehr große, hierarchisch aufgebaute Datenbestände zu navigieren. Typische Operationen für *gerichtete Analysen* sind *Drill Down*, *Drill Across*, *Roll Up*, *Slice* und *Dice*. Sie erlauben es z. B. den Gesamtumsatz aus verschiedenen Blickwinkeln zu betrachten und genauer aufzugliedern. So kann z. B. festgestellt werden, wie sich der Umsatz eines Produktes einer bestimmten Kategorie in einem bestimmten Verkaufsgebiet im Laufe der Zeit entwickelt. Verkauft sich ein Produkt besonders gut, ein anderes eher verhalten, könnte entschieden werden, in einer Aktion gezielt beide Produkte kombiniert anzubieten (*Cross Selling*).

Data Mining ist eine weitere wichtige Komponente von BI. Verfahren des Data Mining werden eingesetzt, um bisher unbekannte Muster, Abweichungen und Auffälligkeiten zu identifizieren. Sie werden daher oft als *ungerichtete Analysen* bezeichnet. Die verwendeten Techniken stammen meist aus anderen Bereichen wie z. B. Statistik und maschinelles Lernen. Bekanntestes Einsatzgebiet für derartige Verfahren sind Empfehlungsmaschinen, bei denen Algorithmen die Transaktionen (Warenkörbe) vieler Kunden analysieren und auf dieser Basis Empfehlungen geben. Data Mining-Techniken werden aber auch zur Früherkennung potentieller Betrüger oder bei der Analyse der Risikofaktoren für den Zahlungsausfall bei Krediten eingesetzt.

Ausführlichere Darstellungen, Erklärungen und Anwendungsbeispiele für BI, OLAP und Data Mining finden sich z. B. in [KBM10] und [CG10].

Forecasting

Auf den für analytische Zwecke aufbereiteten Datenbeständen lassen sich neben IST-Analysen auch Prognosen (Forecasting) erstellen. Auf der Grundlage historischer und aktueller und oft unter Hinzuziehung externer Daten können mittels einer Vielzahl von Verfahren Vorhersagen erstellt werden, die mit einer bestimmten Wahrscheinlichkeit Aussagen über künftige Entwicklungen machen und frühzeitige Entscheidungen ermöglichen. Damit können geeignete Strategien für die Zukunft erstellt werden, um die eigene Wettbewerbsfähigkeit zu erhalten bzw. um sich einen Wettbewerbsvorteil gegenüber den Mitbewerbern zu schaffen. Typische Anwendungsbereiche sind z. B. die Planung und Einführung

neuer Produkte, Prognosen zu potentiellen Absatzmöglichkeiten und Marktanteilen oder für Lagerhaltung und Logistik wichtige Vorhersagen zu Absatztrends.

Realtime-Analysen

Die Bereitstellung von Analysen, Kennzahlen und Berichten im Rahmen klassischer BI-Lösungen wird zunehmend ergänzt und teilweise abgelöst durch sogenanntes *Realtime-BI*. Fachanwender und Analysten wollen nicht mehr tagelang auf Auswertungen warten, sondern diese möglichst in Sekundenschnelle in die Hand bekommen oder sogar selbst in die Hand nehmen. Transaktions- und Analysedaten aus fast allen Datenquellen werden dafür nahezu in Echtzeit integriert, ausgewertet und aufbereitet. So lässt sich beispielsweise das aktuelle Nachfrageverhalten im Onlineshop direkt erkennen und der Warenfluss in den Lieferketten kann in Echtzeit gesteuert werden. Für die schnelle Ermittlung von auftretenden Qualitätsabweichungen und Fehlern in der Produktion sind Realtime-Auswertungen ebenfalls von großem Vorteil. In vielen Branchen (Lebensmittelindustrie, Pharmaindustrie, Wasserversorgungsunternehmen usw.) werden z. B. regelmäßig Produktproben im Labor analysiert. Mit der Auswertung der Daten in Echtzeit kann schnell in den betreffenden Produktionsprozess eingegriffen und damit vermieden werden, dass größere Mengen fehlerhafter Produkte erzeugt oder sogar in Umlauf gebracht werden. Teure und imageschädliche Rückrufaktionen können so verhindert werden.



Kontrollfragen/Aufgaben:

1. *Wodurch ist das Datenmanagement in vielen Unternehmen gekennzeichnet?*
2. *Welche Probleme können durch die Mehrfachhaltung von Daten im operativen und analytischen Bereich entstehen?*
3. *Welche Vorteile bietet Master Data Management?*
4. *Welche Anwendungsmöglichkeiten für Realtime-Analysen könnten ebenfalls interessant sein?*

■ 2.3 Herausforderungen

Im folgenden werden die verschiedenen Probleme, welche die Integration von Daten zu einer herausfordernden Aufgabe macht, dargestellt und näher beschrieben. Dabei wird näher auf die *Verteilung* von Daten, die *autonome Verwaltung* von Datenbeständen und die daraus entstehende *Heterogenität* eingegangen. *Verteilung*, *Autonomie* und *Heterogenität* werden auch als *orthogonale Dimensionen der Informationsintegration* bezeichnet [LN07, S.50].