



BUSINESS INTELLIGENCE I

DLMIWBI01

iu

INTERNATIONALE
HOCHSCHULE

BUSINESS INTELLIGENCE I

IMPRESSUM

Herausgeber:
IU Internationale Hochschule GmbH
IU International University of Applied Sciences
Juri-Gagarin-Ring 152
D-99084 Erfurt

Postanschrift:
Albert-Proeller-Straße 15-19
D-86675 Buchdorf
media@iu.org
www.iu.de

DLMIWBI01
Versionsnr.: 002-2023-0721
N. N.

© 2023 IU Internationale Hochschule GmbH
Dieses Lernskript ist urheberrechtlich geschützt. Alle Rechte vorbehalten.
Dieses Lernskript darf in jeglicher Form ohne vorherige schriftliche Genehmigung der
IU Internationale Hochschule GmbH (im Folgenden „IU“) nicht reproduziert und/oder
unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet wer-
den.

Die Autor:innen/Herausgeber:innen haben sich nach bestem Wissen und Gewissen
bemüht, die Urheber:innen und Quellen der verwendeten Abbildungen zu bestimmen.
Sollte es dennoch zu irrtümlichen Angaben gekommen sein, bitten wir um eine dement-
sprechende Nachricht.



WISSENSCHAFTLICHE KURSLEITUNG

PROF. DR. PETER POENSGEN

Seit Juli 2020 lehrt Herr Poensgen an der IU Internationale Hochschule im Fachbereich Business Intelligence. Seine Schwerpunkte liegen auf dem Gebiet der Datenanalyse und dem Datenbankmanagement.

Herr Poensgen begann seinen beruflichen Werdegang als Spezialist für Datenbanken und Entwicklung in Düsseldorf. Es folgte eine Anstellung in der Finanzbranche, wo er beginnend als Projektmanager seine Kenntnisse in datenbankbasierten Systemen weiter ausbauen konnte. In diversen leitenden Positionen von Banken und dem Energiehandel war er für die Entwicklung, den Support und die Projektleitung von Handels-, Backoffice- und Risikomanagementsystemen verantwortlich. Seit 2012 ist er IT-Koordinator eines Versicherungsunternehmens in Köln.

Parallel zu seiner beruflichen Tätigkeit unterrichtete er als Lehrbeauftragter IT-Management und Wirtschaftsmathematik und war externer wissenschaftlicher Mitarbeiter am Institut für Informationssysteme der Universität zu Lübeck, wo er im Forschungsbereich der Anfrageverarbeitung- und optimierung von Datenbanksystemen promovierte.

INHALTSVERZEICHNIS

BUSINESS INTELLIGENCE I

Wissenschaftliche Kursleitung	3
Einleitung	
Wegweiser durch das Studienskript	8
Literaturempfehlung	9
Pflichtliteratur	12
Übergeordnete Lernziele	13
Lektion 1	
Motivation und Begriffsbildung	15
1.1 Motivation und historische Entwicklung	16
1.2 BI als Rahmenwerk	19
Lektion 2	
Datenbereitstellung	23
2.1 Operative und dispositive Systeme	24
2.2 Das Data-Warehouse-Konzept	26
2.3 Architekturvarianten	29
Lektion 3	
Data Warehouse	37
3.1 ETL-Prozess	38
3.2 DWH und Data Mart	48
3.3 ODS und Metadaten	51
Lektion 4	
Modellierung multidimensionaler Datenräume	59
4.1 Datenmodellierung	60
4.2 OLAP-Würfel	62
4.3 Physische Speicherung	65
4.4 Star- und Snowflake-Schema	66
4.5 Historisierung	68

Lektion 5	
Analysesysteme	73
5.1 Freie Datenrecherche und OLAP	75
5.2 Berichtssysteme	77
5.3 Modellgestützte Analysesysteme	78
5.4 Konzeptorientierte Systeme	80
Lektion 6	
Distribution und Zugriff	83
6.1 Informationsdistribution	84
6.2 Informationszugriff	90
Verzeichnisse	
Literaturverzeichnis	96
Abbildungsverzeichnis	98

EINLEITUNG

HERZLICH WILLKOMMEN

WEGWEISER DURCH DAS STUDIENSKRIPT

Dieses Studienskript bildet die Grundlage Ihres Kurses. Ergänzend zum Studienskript stehen Ihnen weitere Medien aus unserer Online-Bibliothek sowie Videos zur Verfügung, mit deren Hilfe Sie sich Ihren individuellen Lern-Mix zusammenstellen können. Auf diese Weise können Sie sich den Stoff in Ihrem eigenen Tempo aneignen und dabei auf lernspezifische Anforderungen Rücksicht nehmen.

Die Inhalte sind nach didaktischen Kriterien in Lektionen aufgeteilt, wobei jede Lektion aus mehreren Lernzyklen besteht. Jeder Lernzyklus enthält jeweils nur einen neuen inhaltlichen Schwerpunkt. So können Sie neuen Lernstoff schnell und effektiv zu Ihrem bereits vorhandenen Wissen hinzufügen.

In der IU Learn App befinden sich am Ende eines jeden Lernzyklus die Interactive Quizzes. Mithilfe dieser Fragen können Sie eigenständig und ohne jeden Druck überprüfen, ob Sie die neuen Inhalte schon verinnerlicht haben.

Sobald Sie eine Lektion komplett bearbeitet haben, können Sie Ihr Wissen auf der Lernplattform unter Beweis stellen. Über automatisch auswertbare Fragen erhalten Sie ein direktes Feedback zu Ihren Lernfortschritten. Die Wissenskontrolle gilt als bestanden, wenn Sie mindestens 80 % der Fragen richtig beantwortet haben. Sollte das einmal nicht auf Anhieb klappen, können Sie die Tests beliebig oft wiederholen.

Wenn Sie die Wissenskontrolle für sämtliche Lektionen gemeistert haben, führen Sie bitte die abschließende Evaluierung des Kurses durch.

Die IU Internationale Hochschule ist bestrebt, in ihren Skripten eine gendersensible und inklusive Sprache zu verwenden. Wir möchten jedoch hervorheben, dass auch in den Skripten, in denen das generische Maskulinum verwendet wird, immer Frauen und Männer, Inter- und Trans-Personen gemeint sind sowie auch jene, die sich keinem Geschlecht zuordnen wollen oder können.

LITERATUREMPFEHLUNG

ALLGEMEIN

- Bachmann, R./Kemper, G. (2011): *Raus aus der BI-Falle. Wie Business Intelligence zum Erfolg wird*. 2. Auflage, mitp, Heidelberg.
- Bauer, A./Günzel, H. (2008): *Data Warehouse Systeme. Architektur, Entwicklung, Anwendung*. 3. Auflage, dpunkt Verlag, Heidelberg.
- Betz, R. (2015): *Werde Jäger des verlorenen Schatzes*. In: Immobilienwirtschaft, Heft 5, Freiburg, S. 60. (Datenbank: WISOnet).
- Engels, C. (2008): *Basiswissen Business Intelligence*. W3L, Herdecke, Witten.
- Gansor, T./Totok, A./Stock, S. (2010): *Von der Strategie zum Business Intelligence Competency Center (BICC). Konzeption - Betrieb - Praxis*. Carl Hanser Verlag, München.
- Grossmann, W./Rinderle-Ma, S. (2015): *Fundamentals of Business Intelligence*. Springer, Berlin Heidelberg.
- Kemper, H.-G./Baars, H./Mehanna, W. (2020): *Business Intelligence & Analytics – Grundlagen und praktische Anwendungen. Eine Einführung in die IT-basierte Managementunterstützung*. 4. überarb. und erw. Auflage, Springer Vieweg, Stuttgart.
- Sherman, R. (2014): *Business Intelligence Guidebook: From Data Integration to Analytics*. Morgan Kaufmann, Burlington, Massachusetts.
- Turban, E. et al. (2010): *Business Intelligence. A Managerial Approach*. 2. Auflage. Prentice Hall, Upper Saddle River, NJ.
- Turban, E./Sharda, R./Delen, D. (2014): *Business Intelligence and Analytics*. Pearson Education Canada, Toronto.

LEKTION 1

- Chaudhuri, S./Dayal, U./Narasayya, V. R. (2011): *An overview of business intelligence technology*. In: Communications of the ACM, 54. Jg., Heft 8, New York, Seite 88-98. (Im Internet verfügbar).
- Jaeck, H.-F. (2013): *Wenn Potenzial brach liegt*. In: *acquisa*, 61. Jg., Heft 4, Freiburg, S. 46f. (Datenbank: WISOnet).

Krawatzek, R./Zimmer, M./Trahasch, S. (2013): *Agile Business Intelligence. Definition, Maßnahmen und Herausforderungen*. In: HMD - Praxis der Wirtschaftsinformatik, 50. Jg., Heft 290, Heidelberg, S. 56-63. (Datenbank: WISOnet).

Steinschaden, J. (2015): *Runtastic. Von der App-Schmiede zur Big-Data-Firma*. In: Horizont, Heft 14, Frankfurt/M., S. 14f. (Datenbank: WISOnet).

LEKTION 2

Ariyachandra, T./Watson, H. J. (2008): *Which Data Warehouse Architecture is Best?* In: Communications of the ACM, 51. Jg., Heft 10, New York, Seite 146-147. (Datenbank: EBSCO).

Clegg, D. (2015): *Evolving Data Warehouse and BI Architectures. The Big Data Challenge*. In: Business Intelligence Journal, 20. Jg., Renton, WA, Seite 19-24. (Datenbank: EBSCO).

is report (2013): *Business-Intelligence- und ERP-Systeme arbeiten Hand in Hand*. In: is report, Heft 11, München, S. 22-24. (Datenbank: WISOnet).

Ivan, M.-L. (2014): *Characteristics of In-Memory Business Intelligence*. In: Informatica Economica, 18. Jg., Heft 3, Bukarest, Seite 17-25. (Datenbank: EBSCO).

LEKTION 3

Ankorion, I. (2005): *Change Data Capture. Efficient ETL for Real-Time BI*. In: DM Review, 15. Jg., New York, Seite 36-43. (Datenbank: EBSCO).

Dill, M./Bukowski, A. (2014): *Qualitätskontrolle bei Big Data erfordert Umdenken*. In: is report, Heft 5, München, S. 16-19. (Datenbank: WISOnet).

Feld, T./Jost, W./Scheer, A.-W. (2014): *Die nächste Generation von Unternehmensanwendungen. Entwicklung des Phänomens Big Data*. In: ZFO - Zeitschrift Führung und Organisation, 83. Jg., Heft 6, Stuttgart, S. 364-371. (Datenbank: WISOnet).

Sherman, R. (2009): *Beyond ETL & Data Warehousing*. In: Information Management, Jan./Feb. Supplement, Overland Park, KS, Seite 2-7. (Datenbank: EBSCO).

LEKTION 4

Dill, M. (2014): *Data Warehousing. In-Memory braucht eine eigene Datenmodellierung*. In: is report, Heft 2, München, S. 22-27. (Datenbank: WISOnet).

Kimball, R. (2008): *Slowly Changing Dimensions, Types 2 and 3*. In: DM Review, 18. Jg., Heft 10, New York, Seite 19-38. (Datenbank: EBSCO).

Stiglich, P. (2014): *Data Modeling in the Age of Big Data*. In: Business Intelligence Journal, 19. Jg., Heft 4, Renton, WA, Seite 17-22. (Datenbank: EBSCO).

LEKTION 5

Allo, M. (2012): Strategic dashboards: designing and deploying them to improve implementation. *Strategy & Leadership*, 40. Jg., Heft 5, Seite 24-31. (Datenbank: Emerald).

Akbay, S. (2015): *How Big Data Applications Are Revolutionizing Decision Making*. In: *Business Intelligence Journal*, 20. Jg., Renton, WA, Seite 25-29. (Datenbank: EBSCO).

is report (2012): *Reporting-Projekte zeigen Wege aus der Datenflut*. In: *is report*, Heft 7-8, München, S. 14-17. (Datenbank: WISOnet).

Mödritscher, G. et al. (2014): *Die Steuerung von „Rennern“ und „Pennern“ im Handelssortiment*. In: *Controller Magazin*, Heft 1, Freiburg, S. 30-34. (Datenbank: WISOnet).

Panke, B. (2015): *Geringe Margen erfordern schnelle Entscheidungen. Das Management-Cockpit als Basis*. In: *Getränkfachgroßhandel*, Heft 2, Mindelheim, S. 11-13. (Datenbank: WISOnet).

LEKTION 6

Alpar, P./Schulz, M.(2016): *Self-Service Business Intelligence*. In: *Business & Information Systems Engineering*, 58. Jg., Seite 151-155. (Datenbank: Springer).

Berndt, T./Müller, L. (2015): *Extensible Business Reporting Language als Standard einer modernen Unternehmensberichterstattung*. In: *Betriebs-Berater*, 70. Jg., Heft 19, Frankfurt am Main, S. 1128-1132. (Datenbank: WISOnet).

Elm, K. v. (2014): *Mobiles Filialbüro*. In: *rt-retail technology*, Heft 1, Köln, S. 92-94. (Datenbank: WISOnet).

is report (2014): *Self-Service BI setzt die Anwender ans Steuer*. In: *is report*, Heft 8, München, S. 30-33. (Datenbank: WISOnet).

Lennerholt, C./van Laere, J./Söderström, E. (2018): *Implementation challenges of Self Service Business Intelligence: A literature review*. In: *Proceedings of the 51st Hawaii International Conference on System Sciences*. Bd. 51, S. 5055–5063. Hilton Waikoloa Village, Hawaii, USA. (Datenbank: Base).

Ordyniec, K. (2014): *IT im Autohaus. Im Labyrinth der Daten*. In: *Autohaus*, Heft 20, Berlin, S. 22-24. (Datenbank: WISOnet).

PFLICHTLITERATUR

LEKTION 1

Krawatzek, R./Zimmer, M./Trahasch, S. (2013): *Agile Business Intelligence. Definition, Maßnahmen und Herausforderungen*. In: HDM – Praxis Wirtschaftsinformatik, 50. Jg., Heft 290, Heidelberg, S. 56-63. (Datenbank: WISOnet).

LEKTION 2

Clegg, D. (2015): *Evolving Data Warehouse and BI architectures. The Big Data Challenge*. In: Business Intelligence Journal, 20. Jg., Renton, WA, S. 19-24. (Datenbank: EBSCO).

LEKTION 3

Shermann, R. (2009): *Beyond ETL & Data Warehousing*. In: Information Management, Jan./Feb. Supplement, Overland Park, KS, S. 2-7. (Datenbank: EBSCO).

LEKTION 4

Kimball, R. (2008): *Slowly Changing Dimensions, Types 2 and 3*. In: DM Review, 18. Jg., Heft 10, New York, S. 19-38. (Datenbank: EBSCO).

LEKTION 5

Akbay, S. (2015): *How Big Data Applications Are Revolutionizing Decision Making*. In: Business Intelligence Journal, 20. Jg., Renton, WA, S. 25-29. (Datenbank: EBSCO).

LEKTION 6

is report (2014): *Self-Service BI setzt die Anwender ans Steuer*. In: is report, Heft 8, München, S. 30-33. (Datenbank: WiSOnet).

ÜBERGEORDNETE

LERNZIELE

Business Intelligence (BI) dient der Gewinnung von Informationen aus Unternehmensdaten, die sowohl für eine gezielte Unternehmenssteuerung als auch für die Optimierung von Geschäftsaktivitäten relevant sind. Im Rahmen dieses Kurses werden Techniken, Vorgehensweisen und Modelle zur Datenbereitstellung, Informationsgenerierung und -analyse sowie der Verteilung der gewonnenen Informationen vorgestellt und diskutiert. Sie werden danach in der Lage sein, die verschiedenen Themengebiete des Data Warehousing zu erläutern und Methoden bzw. Techniken für konkrete Anforderungen selbständig auszuwählen. Zum Abschluss des Kurses können Sie selbstständig auf der Grundlage konkreter Anforderungen Business Intelligence-Anwendungen konzipieren und prototypisch umsetzen.

LEKTION 1

MOTIVATION UND BEGRIFFSBILDUNG

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- welche Bedeutung der Begriff Business Intelligence (BI) hat.
- wie der Begriff Business Intelligence entstanden ist.
- welche Merkmale ein Data Warehouse hat.
- welche BI-Verständnisse in der Praxis verbreitet sind.

1. MOTIVATION UND BEGRIFFSBILDUNG

Einführung

Seit mehreren Jahren ist ein Trend zur Globalisierung und Dynamisierung der Märkte zu beobachten. Veranlasst durch diesen härter werdenden Wettbewerb versuchen viele Unternehmen Informationsvorsprünge und somit Wettbewerbsvorteile zu erlangen. Dadurch werden Informationen zur unternehmerischen Ressource und sind als Wettbewerbs- und Produktionsfaktoren von strategischer Bedeutung. Eine funktionierende Informationsversorgung ist dabei die Voraussetzung dafür, unternehmerische Entscheidungsprozesse qualitativ verbessern zu können.

Business Intelligence (BI) umfasst die Integration von Strategien, Prozessen und Technologien, um im Umfeld der entscheidungsunterstützenden Systeme aus fragmentierten, inhomogenen Unternehmens-, Markt- und Wettbewerberdaten erfolgskritisches Wissen über Status, Potenziale und Perspektiven zu generieren und dies für Analyse-, Planungs- und Steuerungszwecke geeignet darzustellen.

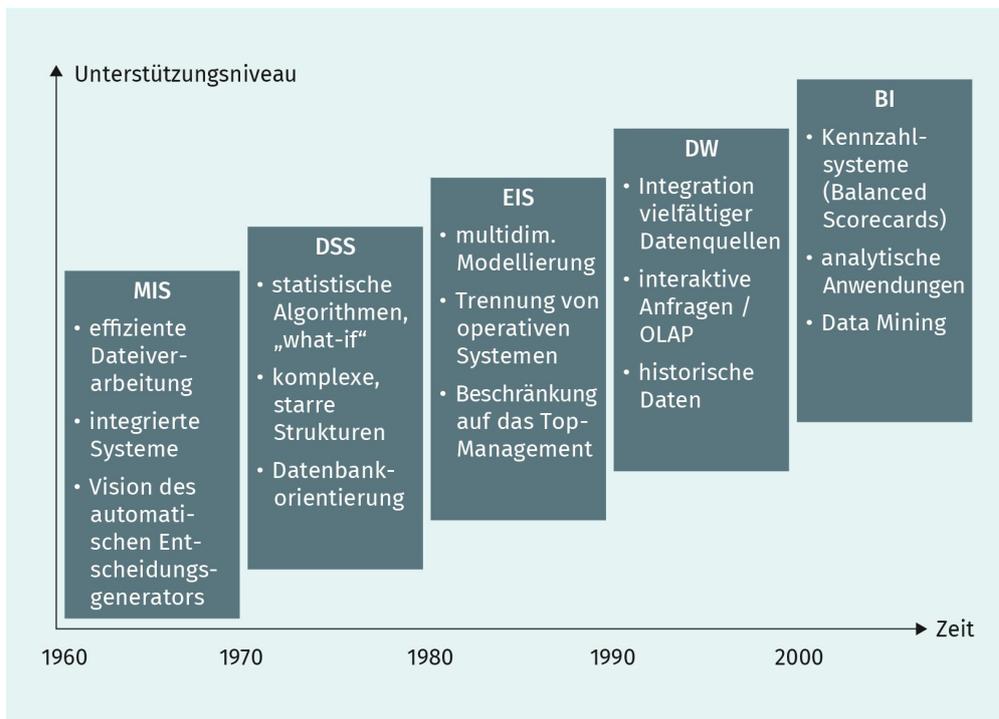
Neben dem übergeordneten Konzept Business Intelligence wird der Begriff Data Warehouse (DWH, DW) aufgrund der verschiedenen Definitionen und Deutungen oftmals unterschiedlich verstanden. Nachfolgend stellen wir zunächst die historische Entwicklung von Business Intelligence dar. Anschließend wird der Begriff Data Warehouse formal beschrieben und dann im übergeordneten Konzept BI eingeordnet.

1.1 Motivation und historische Entwicklung

Die historische Entwicklung von Business Intelligence reicht bis in die 1960er Jahre zurück. Wie die folgende Abbildung zeigt, existierten schon damals Systeme zur Entscheidungsunterstützung für das Management.

Das Data Warehouse wurde zunächst mit DW abgekürzt. Heute erfolgt die Abkürzung vermehrt mit DWH. Daher finden Sie, insbesondere in Abbildungen aus Standardwerken, unterschiedliche Abkürzungen.

Abbildung 1: Historische Entwicklung



Quelle: Humm/Wietek 2005, S. 4.

Management Information System (MIS)

Ende der 1960er Jahre wurden die ersten Informationssysteme mit dem Begriff Management Information System (MIS) eingeführt. Nach Grothe bestand das Ziel darin, „Managern in ihren Unternehmen die Informationen anzubieten, die sie für ihre Entscheidungen benötigen. Als Nebenbedingungen waren Zeit, Inhalt und Art der Informationsdarbietung zu optimieren“ (Grothe 2000, S. 65). Die Ziele konnten jedoch nach Gluchowski/Gabriel/Dittmar (2008) nur begrenzt eingehalten werden. Grund waren zumeist die damaligen technischen Voraussetzungen.

Decision Support Systems (DSS)

Mitte der 70er Jahre wurde der Begriff MIS durch Decision Support System (DSS) abgelöst. Mit den interaktiv-gestützten EDV-Systemen wurden zusätzlich Modelle, Methoden und Szenarien angeboten, die individuelle Analysen von Informationen ermöglichten (Gluchowski/Gabriel/Dittmar 2008). Durch den Fortschritt der Hardware konnten Informationen effizient verarbeitet werden. Hier wurde die Basis für datenbasierte Entscheidungsunterstützung gelegt (Grothe 2000).

Allerdings erfüllten auch die DSS die hohen Erwartungen nicht. Dank des technischen Fortschritts konnten strukturierte Daten analysiert werden. Die Analyse der Daten war jedoch nur für Teilbereiche des Unternehmens möglich. Außerdem konnten diese nur mit operativen Daten durchgeführt werden (Gluchowski/Gabriel/Dittmar 2008; Grothe 2000).

Nach Hannig bestand ein weiteres Problem darin, dass Manager die Systeme nicht akzeptierten, da sie einem Computer nicht zutrauten, kreative Entscheidungsprozesse zu unterstützen (Hannig 2002).

Executive Information Systems (EIS)

Mitte der 80er Jahre etablierte sich mit dem Einzug leistungsfähiger Personal Computer (PC) in die Unternehmen der Begriff Executive Information Systems (EIS) (Gluchowski/Gabriel/Dittmar 2008; Hannig 2002).

Zielgruppen der Programme waren insbesondere das Topmanagement und Controlling. Durch EIS sollten individuelle Systeme bereitgestellt werden, welche dem Management entscheidungsrelevante, multidimensionale Daten aktueller und besser präsentierten (Gluchowski/Gabriel/Dittmar 2008). Im Gegensatz zu den Vorläufern (MIS und DSS) war EIS aufgrund der Verbreitung von PCs in Unternehmen leichter umzusetzen. Bei MIS und DSS mussten zentrale Rechner eingesetzt werden. Der Nachteil von individuellen Systemen war jedoch, dass diese nur innerhalb der Abteilung bzw. des Unternehmens eingesetzt werden konnten, da sie hierfür individuell entwickelt wurden.

Der Durchbruch von EIS wurde verhindert, da auch EIS keine Akzeptanz bei Endanwendern fand (Hannig 2002); ebenso waren nachträgliche Änderungen – aufgrund der individuellen Entwicklung – aufwendig (Grothe 2000).

DWH (Date Warehouse)

Der Durchbruch kam insbesondere durch die fortschreitende Globalisierung Anfang der 90er Jahre. Die Manager waren in den Jahren zuvor eher skeptisch gewesen. Nun waren Manager auf diese Informationen angewiesen. Die Entscheidungsfindung hatte sich aufgrund der Dezentralisierung grundlegend geändert. Entscheidungen wurden nicht mehr in der Zentrale getroffen, die am anderen Ende der Welt liegen konnte, sondern zeitnah, vor Ort und mit aktuellen Informationen. Ein weiterer Grund für die verstärkte Nachfrage war die Datenflut, die auch aus der Internationalisierung und den damit verbundenen Unternehmensstandorten in aller Welt resultierte. Bisherige Systeme (MIS, DSS, EIS) waren nicht in der Lage, diese Anforderungen umzusetzen.

Das Hauptproblem war, dass in Unternehmen mehrere inkonsistente Datenquellen bestanden. Aufgrund dessen wurde eine Richtung für das neue System vorgegeben. Es wurde eine vollständige, einheitliche und konsistente Datenbasis benötigt (Hannig 2002). Man entwickelte eine zentrale Datenbasis, die Daten aus den verschiedenen Systemen eines Unternehmens zusammenführte: daher der Begriff Data Warehouse (DWH) (Grothe 2000; Hannig 2002).

In den 90er Jahren entstanden im Rahmen der Erstellung von Data Warehouse Analysewerkzeuge, die häufig mit dem Begriff Business Intelligence (BI) beschrieben wurden. Der Begriff BI wird heute zumeist als Oberbegriff verwendet (Grothe 2000).

1.2 BI als Rahmenwerk

Viele Unternehmen erkennen die Situation stetig steigender Datenflut bei gleichzeitigem Informationsdefizit. Dabei stellt die Informationsversorgung des Managements einen wichtigen Wettbewerbsfaktor dar. Häufig fehlt jedoch die richtige Information in der richtigen Menge am richtigen Ort zur richtigen Zeit. Mithilfe des Data Warehouse kann die betriebliche Informationslogistik verbessert werden.

Merkmale eines DWH

Ein DWH ist immer ein Konzept, welches die Datenproblematik des Managements behandelt. Der Vater des Data Warehousing (W. H. Inmon) hat die folgende Definition geprägt:

**DATA WAREHOUSE**

A data warehouse is a subject-oriented, integrated, nonvolatile, time-variant collection of data in support of management's decisions (Inmon 2005, S. 31).

Die elementaren vier Merkmale werden im Folgenden beschrieben:

1. **Subject-oriented** (Themenorientierung) bedeutet, dass der Datenbestand eines DWH nach fachlichen bzw. betriebswirtschaftlichen Kriterien ausgewählt und organisiert wird.
2. **Integrated** (Integration) steht für die Integration von Daten aus heterogenen Quellsystemen. Dabei müssen die Daten hinsichtlich Struktur und Format vereinheitlicht werden.
3. **Nonvolatile** (Persistenz) steht für die beständige Speicherung von Daten im DWH. Gespeicherte Daten werden nicht geändert oder gelöscht.
4. **Time-variant** (Historisierung) bedeutet, dass im DWH Zeitreihenanalysen (Vergleich von Daten über die Zeit) möglich sind. Dadurch können Veränderungen und Entwicklungen analysiert werden.

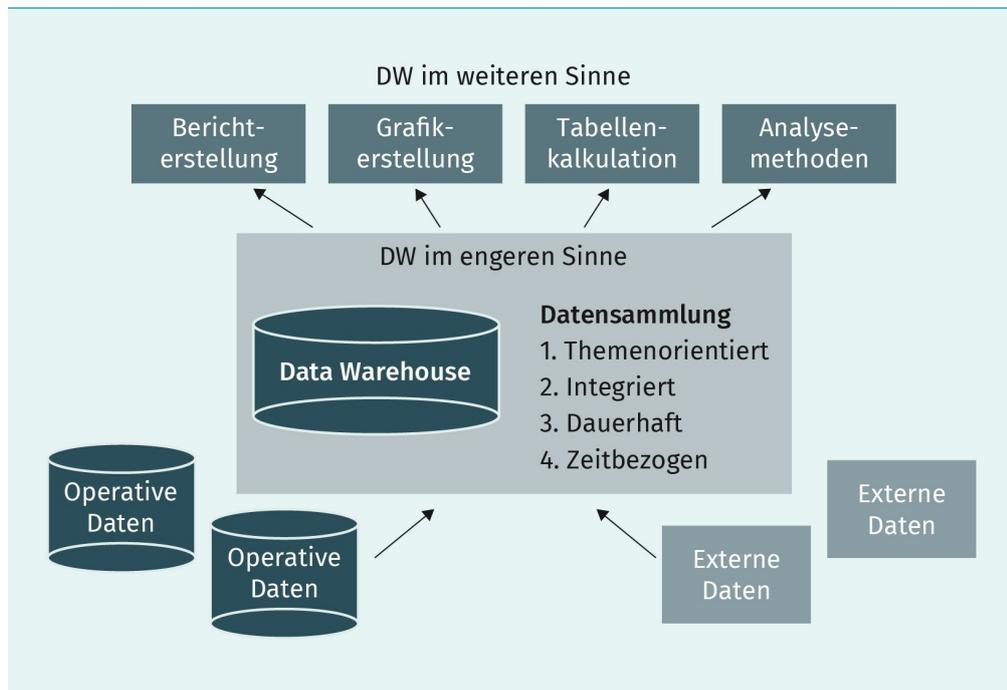
Definitionsvielfalt

In der Literatur existieren eine Vielzahl von DWH- und BI-Definitionen. Im Folgenden wird das Verständnis von DWH bzw. BI zusammenfassend dargestellt.

DWH-Verständnis

Die oben genannte Definition von Inmon wurde von vielen Autoren erweitert, die neben der Datensammlung und Verwaltung auch Aufgaben wie Anbindung, Extraktion und Transformation von Fremddaten nennen. Weiterhin erfolgt nach Schinzer/Bange/Mertens (1999) eine Ausdehnung in Richtung Analyse und Präsentation mithilfe entsprechender Werkzeuge.

Abbildung 2: Abgrenzungen des DWH-Begriffs



Quelle: erstellt im Auftrag der IU, 2016.

Data Warehouse (DWH) im engeren Sinne
Es umfasst die reine Datensammlung.

In der Abbildung ist zu erkennen, dass das **Data Warehouse (DWH) im engeren Sinne** die reine Datensammlung umfasst. DWH im weiteren Sinne schließt die Berichterstellung, Grafikerstellung, Tabellenkalkulation und Analysemethoden mit ein.

BI-Verständnis

Die alleinige Archivierung von Daten bringt jedoch keine Wettbewerbsvorteile. Diese werden erst durch eine kreative und intelligente Verwendung realisiert (Muksch/Behme 1998). Die Nutzung von unternehmensweit verfügbarem Wissen wird daher als BI bezeichnet. BI stellt damit eine nochmalige Ausdehnung des erweiterten DWH-Begriffs dar, sozusagen als Frontend des DWH. Geprägt wurde der Begriff von der Gartner Group und ist wie folgt definiert:



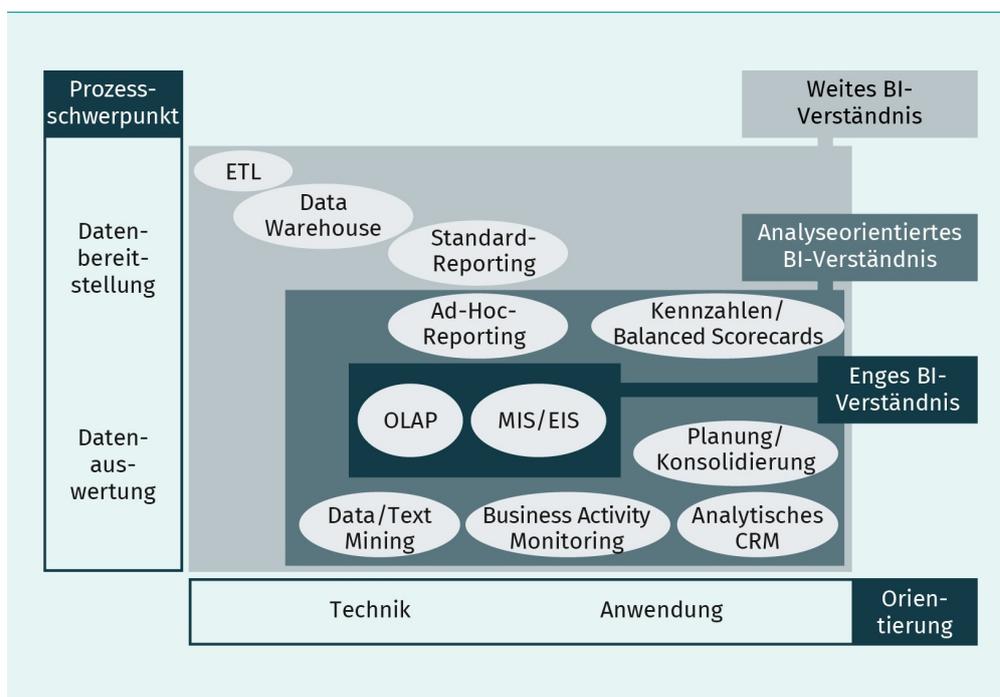
BUSINESS INTELLIGENCE

Business Intelligence is the process of transforming data into information and, through discovery, into knowledge (Muksch/Behme 1998, S. 37).

Nach Gluchowski/Gabriel/Dittmar (2008) handelt es sich bei BI um Techniken und Anwendungen, die entscheidungsunterstützenden Charakter aufweisen und zu einem besseren Verständnis für die Mechanismen relevanter Wirkungsketten führen sollen.

Die folgende Abbildung zeigt die Einordnung von BI in ein enges, analyseorientiertes und ein weites BI-Verständnis.

Abbildung 3: Einordnung von BI



Quelle: Gluchowski/Gabriel/Dittmar 2008, S. 92.

BI im engeren Sinn beschreibt Kernapplikationen, die eine Entscheidungsfindung ohne große Methoden- und Modellkonzepte unterstützen. Diese sind z. B. Online Analytical Processing (OLAP), MIS und EIS.

Bei analyseorientierter BI werden Anwendungen berücksichtigt, bei denen Entscheider mithilfe einer Benutzeroberfläche direkt auf dem System unter Verwendung von Methoden und Modellen vorhandene Daten analysieren können. Dies sind z. B. OLAP, MIS und EIS, Text Mining, Data Mining und Ad-Hoc-Reporting.

Business Intelligence (BI) im weiteren Sinne

Dies umfasst alle BI-Anwendungen, die direkt oder indirekt angewandt werden.

Der Begriff Business Intelligence **(BI) im weiteren Sinne** umfasst alle Anwendungen, die direkt oder indirekt für die Entscheidungsfindung angewandt werden. Dies umfasst Auswertungs- und Präsentationsfunktionen wie auch Datenaufbereitung und Speicherung (Kemper/Baars/Mehanna 2010; Gluchowski/Gabriel/Dittmar 2008).



ZUSAMMENFASSUNG

Mithilfe von Business Intelligence versuchen Unternehmen Informationsvorsprünge und somit Wettbewerbsvorteile zu erlangen. Informationen werden dadurch zur unternehmerischen Ressource und sind als Wettbewerbs- und Produktionsfaktoren von strategischer Bedeutung.

Die historische Entwicklung von Business Intelligence reicht bis in die 1960er Jahre zurück (BI, DWH, EIS, DSS, MIS).

Die elementaren Merkmale von DWH sind Themenorientierung, Integration, Historisierung und Persistenz. In der Literatur existieren eine Vielzahl von DWH- und BI-Definitionen. Hierbei wird von DWH- bzw. BI-Verständnis im jeweils engeren und weiteren Sinne gesprochen.

BI im weiteren Sinne umfasst alle Anwendungen, welche den Entscheider bei der Entscheidungsfindung direkt (z. B. OLAP) oder indirekt (z. B. Datenextraktion) unterstützen.

LEKTION 2

DATENBEREITSTELLUNG

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- wie sich operative und dispositive Systeme voneinander unterscheiden.
- wie sich eine typische BI-Referenzarchitektur darstellt.
- welche grundlegenden BI-Komponenten existieren.
- welche Architekturvarianten möglich sind.

2. DATENBEREITSTELLUNG

Einführung

Der Begriff Business Intelligence bezeichnet Verfahren und Prozesse zur systematischen Analyse von Daten in elektronischer Form. Durch die Gewinnung von Erkenntnissen wird es Unternehmen ermöglicht, bessere operative oder strategische Entscheidungen zu treffen. Die Grundvoraussetzung für den Einsatz leistungsfähiger BI-Werkzeuge ist die Aufbereitung und Speicherung von konsistenten Daten, welche auf die betriebswirtschaftlichen Bedürfnisse der Manager ausgerichtet sind.

Technisch wird ein DWH auf Basis eines Datenbanksystems realisiert. Die oftmals speziellen Systeme sind auf die Bedürfnisse komplexer Anfragen konfiguriert, da Ladeoperationen mit hohen Datenvolumen zusätzliche Anforderungen darstellen. Die Speicherung der Daten in Tabellen auf Basis relationaler Datenbanksysteme ist weitverbreitet.

2.1 Operative und dispositive Systeme

Die Klassifizierung von operativen und dispositiven Arbeitsleistungen wurde bereits von Erich Gutenberg vorgenommen. Demnach werden Arbeiten als operativ bezeichnet, wenn diese unmittelbar mit der Leistungserstellung, Leistungsverwertung und mit finanziellen Aufgaben in Zusammenhang stehen, ohne dispositiver Natur zu sein. Dispositive Tätigkeiten liegen vor, wenn diese mit der Leitung und Lenkung von betrieblichen Vorgängen in Zusammenhang stehen.

In Anlehnung an die Unterteilung von Erich Gutenberg (1983, S. 6, 10 und 103ff.) können Anwendungssysteme nach Art der unterstützenden Arbeiten in operative und dispositive Systeme unterschieden werden. Die zugehörigen Daten können ebenfalls in oben genannter Form klassifiziert werden.

OLTP und OLAP

Operative Systeme
Diese Systeme (z. B. ERP, CRM) verwalten aktuelle Informationen.

Die **operativen Systeme** dienen dazu, alltägliche Informationen für ein Unternehmen abzuspeichern und zu verwalten. Dies kann eine Kundendatenbank oder ein Mitarbeiterverzeichnis sein. Informationen in diesen Systemen werden regelmäßig geändert und häufig abgefragt. Ausschließlich aktuelle Datensätze sind von Interesse. Vergangene Adressdaten eines Kunden sind beispielsweise von geringem Wert und können gelöscht oder überschrieben werden. Die Datenmodelle sind für eine hohe Transaktionszahl optimiert. Derartige Systeme werden auch als Online Transactional Processing (OLTP) bezeichnet.

Dispositive Systeme
Mit ihrer Hilfe werden Informationen aus operativen Daten gewonnen.

Das Data Warehouse wird den **dispositiven Systemen** zugeordnet. Sie werden dazu eingesetzt, Informationen aus operativen Daten zu gewinnen. So kann beispielsweise festgestellt werden, dass eine bedeutende Anzahl Kunden ihren Standort in den letzten sechs

Monaten in die Hauptstadt verlegt hat. Mithilfe der Information kann die eigene Vertriebsstruktur angepasst und optimiert werden. Durch Einsatz eines DWH werden die operativen Systeme von analytischen Anfragen entlastet.

Aus den verschiedenen Zielsetzungen ergeben sich unterschiedliche Anfrageprofile. Operative Systeme werden von relativ vielen Benutzern angewendet. In den Geschäftszeiten finden zahlreiche Leseanfragen auf einzelne Datensätze statt.

Analytische Systeme werden von einzelnen Experten zu komplexen Sachverhalten befragt. Dabei werden anspruchsvolle Anfragen, die eine große Anzahl von Datensätzen auswerten, gestellt. Die Systeme werden entsprechend ihrem jeweiligen Anwendungszweck optimiert, um eine verbesserte Verarbeitungsgeschwindigkeit zu erreichen. Das zugehörige Datenhaltungskonzept heißt Online Analytical Processing (OLAP).

Aufgrund der Heterogenität von transaktional orientierten, operativen Systemen und dem analytisch ausgerichteten Data Warehouse werden beide Systeme physikalisch getrennt voneinander bereitgestellt.

Operative und dispositive Daten

Beim DWH ergibt sich die Frage nach der technischen Notwendigkeit. Schließlich ist es, salopp gesagt, nur eine Replizierung von Daten, die ein Unternehmen in seinen Datenverarbeitungssystemen generiert und speichert. Grundsätzlich ergeben sich zwei unterschiedliche Sichtweisen auf die Daten: operativ und dispositiv.

Die operativen Daten haben einen direkten Bezug zu den Tätigkeiten der Leistungserbringung des Unternehmens. Dispositive Daten hingegen haben einen analytischen Charakter und dienen der Leitung und Steuerung des Unternehmens (Kemper/Baars/Mehanna 2010).

Im Folgenden werden die wichtigsten Unterschiede der beiden Sichtweisen dargestellt.

Tabelle 1: Charakteristika operativer und dispositiver Daten

	Charakteristika operativer Daten	Charakteristika dispositiver Daten
Ziel	Abwicklung der Geschäftsprozesse	Informationen für das Management; Entscheidungsunterstützung
Ausrichtung	Detaillierte, granulare Geschäftsvorfälle	Meist verdichtete, transformierte Daten; umfassendes Metadatenangebot
Zeitbezug	Aktuell; zeitpunktbezogen; auf die Transaktion ausgerichtet	Unterschiedliche, aufgabenabhängige Aktualität; Historienbetrachtung
Modellierung	Altbestände oft nicht modelliert (funktionsorientiert)	Sachgebiets- oder themenbezogen, standardisiert und endbenutzertauglich
Zustand	Häufig redundant; inkonsistent	Konsistent modelliert; kontrollierte Redundanz

	Charakteristika operativer Daten	Charakteristika dispositiver Daten
Update	Laufend und konkurrierend	Ergänzend; Fortschreibung abgeleiteter, aggregierter Daten
Queries	Strukturiert; meist statisch im Programmcode	Ad-hoc für komplexe, ständig wechselnde Fragestellungen und vorgefertigte Standardauswertungen

Quelle: Kemper/Baars/Mehanna 2010, S. 16.

Am Beispiel eines Versicherungsunternehmens lassen sich die Unterschiede verdeutlichen. Operative Daten in diesem Zusammenhang wären beispielsweise:

- detaillierte Informationen zu den einzelnen Versicherungsverträgen,
- Datenänderungen rund um die Uhr durch das Onlineportal,
- Speicherung der Verträge in zwei unterschiedlichen Systemen, jeweils ein eigenes für Kraftfahrzeug- und Lebensversicherungen.

Als Beispiele für dispositive Daten ließen sich nennen:

- Summierung der Umsätze und Profite pro Kundengruppe,
- Darstellung der zeitlichen Veränderungen zum Vorjahr,
- Gegenüberstellung der Produktparten Kraftfahrzeug- und Lebensversicherung.

Eine direkt auf operativen Daten beruhende Auswertung erfüllt somit nicht die dispositiven Anforderungen. Besonders die heterogene Systemlandschaft erschwert das Gegenüberstellen von Informationen. Zusätzlich verdeutlichen die Aspekte „Abfragen“ und „Transaktionen“, dass direkte Analysen auf operative Daten gefährlich sind: Ressourcenintensive Abfragen mit langen Laufzeiten können das gesamte operative System blockieren und das Tagesgeschäft beeinträchtigen (Kemper/Baars/Mehanna 2010).

2.2 Das Data-Warehouse-Konzept

Das Data-Warehouse-Konzept kann in der Praxis je nach Anforderung unterschiedliche Prozessphasen, Architekturen und BI-Komponenten umfassen. Im Folgenden werden daher grundlegende Konzepte bzw. Referenzarchitekturen erläutert, die dann im konkreten Projekt spezifiziert werden müssen.

Bitte beachten Sie, dass die Begriffe Business Intelligence und Data Warehouse (im weiteren Sinne) in der Literatur und Praxis weitestgehend synonym verwendet werden.

Prozessphasen und Referenzarchitektur

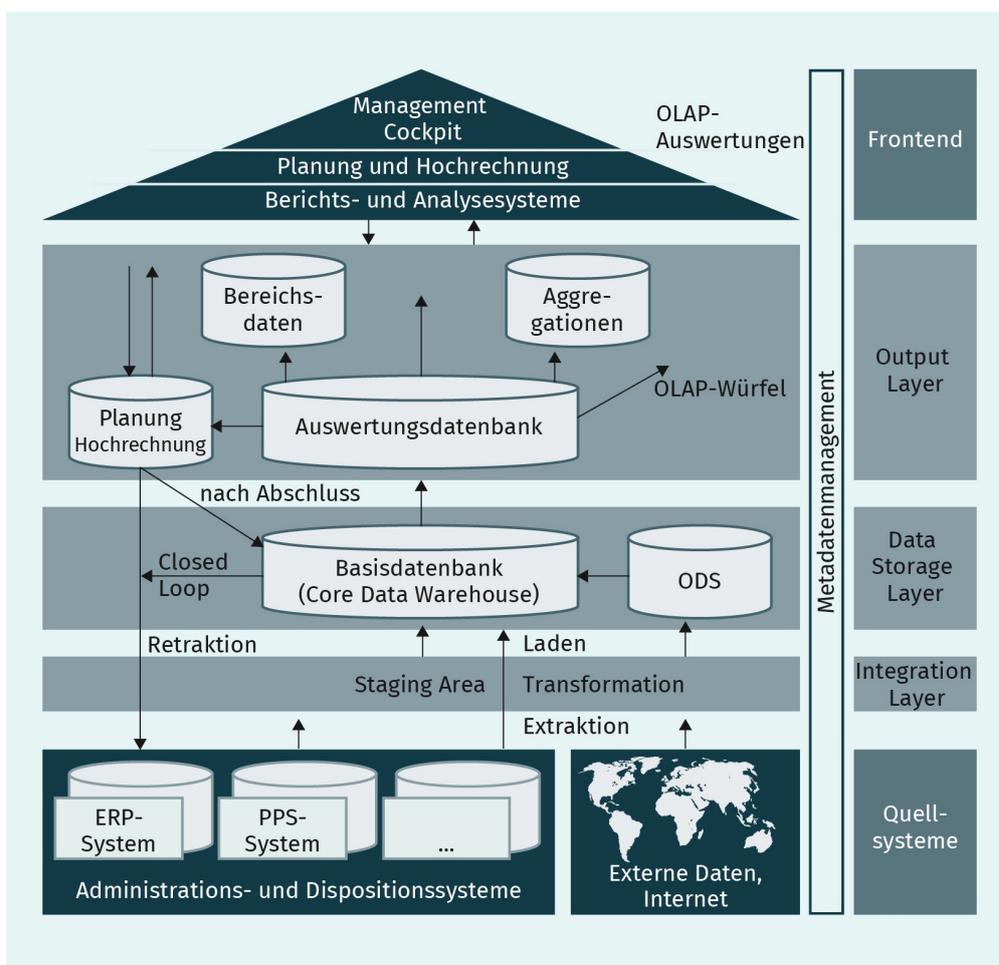
Grundsätzlich existiert in der Literatur eine Vielzahl von Prozessphasen und Referenzarchitekturen. Nach Kemper/Baars/Mehanna (2010) werden die folgenden Prozessphasen unterschieden:

- Datenbereitstellung,
- Informationsgenerierung, Speicherung und Distribution,
- Informationszugriff.

Im ersten Schritt werden Daten und Informationen aus heterogenen Quellen zusammengeführt. Die Daten können dabei aus dem Supply Chain Management (SCM), dem Enterprise Resource Planning (ERP), dem Customer Relationship Management (CRM) oder externen Systemen stammen und werden dann im DWH zusammengeführt. Im zweiten Schritt werden Daten mithilfe von OLAP und Data Mining analysiert. Neben umfangreichen Abfragemöglichkeiten können diese Systeme auch ereignisgesteuerte Warnmeldungen erzeugen. Im dritten Schritt werden die resultierenden Erkenntnisse aus dem zweiten Schritt dem Unternehmen in Form von Aktionen mitgeteilt.

In folgender Abbildung nach Gansor/Totok/Stock (2010) werden die verschiedenen Komponenten einer BI-Referenzarchitektur dargestellt, die als Hilfestellung und Visualisierung bei deren nachfolgender Beschreibung dienen soll.

Abbildung 4: BI-Referenzarchitektur



Quelle: Gansor/Totok/Stock 2010, S. 56.

BI-Komponenten

Quellsysteme

In der klassischen BI-Referenzarchitektur werden Daten aus heterogenen Quellen dargestellt, deren Struktur, Inhalt und Zugriffsschnittstellen unterschiedlich sind. In der Regel werden Systeme des Online Transaction Processing (OLTP) erfasst. Grundsätzlich sind jedoch beliebige Quellsysteme denkbar. Hierzu zählen semistrukturierte Webseiten oder unstrukturierte Textdateien. Hierbei kann es sich sowohl um unternehmensinterne (z. B. ERP-, PPS-System) als auch um extern beschaffte Daten (z. B. Börsenkurse, aktuelle Rohstoffpreise) handeln.

Staging Area

Bei der Staging Area handelt es sich um einen Arbeitsbereich, in welchem Daten temporär zwischengespeichert werden. Die Staging Area wird benötigt, um nachgelagerte Systeme bei der Verarbeitung großer Datenmengen zu entlasten (Inmon 2005).

Operational Data Store (ODS)

Der ODS (Operational Data Store) besitzt im Gegensatz zum klassischen DWH-Ansatz keine aggregierten Daten und längere Historienbetrachtungen. Er wird zudem häufig als Vorstufe zur Datenversorgung herkömmlicher DWH-Ansätze verwendet (Kemper/Baars/Mehanna 2010).

Basisdatenbank (Core Data Warehouse)

Das Core Data Warehouse (Basisdatenbank) stellt die zentrale Datenbank innerhalb des DWH dar. Hierbei werden Daten nach einem ersten Transformationsprozess für unterschiedliche Auswertungszwecke bzw. nachgelagerte Systeme bereitgestellt.

Auswertungsdatenbank (Data Mart)

Die Auswertungsdatenbank bildet die Grundlage für nachgelagerte Analysewerkzeuge. Die Daten werden mithilfe eines multidimensionalen Modells vorgehalten. Aus technischer Sicht liegen Auswertungsdatenbanken zumeist relationale Datenbanken zugrunde. Häufig werden mehrere Auswertungsdatenbanken eingesetzt, wobei Daten nach Analyseanforderungen oder Organisationseinheiten aufgeteilt werden (Bauer/Günzel 2013, S. 68).

ETL-Prozess

Der ETL-Prozess integriert die Daten der Quellsysteme in das DWH. Mithilfe sogenannter ETL-Werkzeuge werden die Verarbeitungsschritte des Extrahierens, Transformierens und Ladens (ETL) vorgenommen.

Aus den Datenquellen werden die Daten über den Datenextraktionsschritt in die Staging Area übertragen. Nach Extraktion der Daten aus den Datenquellen und Laden in den Arbeitsbereich müssen die Daten entsprechend der Anforderungen umgewandelt werden.

Transformationen betreffen die Struktur wie auch den Inhalt der Daten. Daten, die aus unterschiedlichen Quellen stammen, müssen in ein einheitliches Format gebracht werden. Dabei kann anhand von Plausibilitätsprüfungen die Datenqualität verbessert werden. Die Daten werden mithilfe des Ladeschritts in die Basisdatenbank überführt, sobald diese nach der Transformation in einem bereinigten Zustand vorliegen. Da die Basisdatenbank bereits integrierte und bereinigte Daten enthält, müssen Daten vor dem Laden in die Auswertungsdatenbank nur in das Zielschema transformiert und eventuell angereichert bzw. aggregiert werden (Bauer/Günzel 2013, S. 158).

Aggregationen

Die Aggregation von Daten erfolgt, wenn diese in einer geringeren Granularität als in den Quellsystemen benötigt werden. Dadurch wird die Datenmenge erheblich reduziert. Daten werden aus Gründen der Performance in der Regel auf die minimal notwendige Granularität aggregiert. Ein mögliches Beispiel ist die Aggregation von Tages- auf Monatsumsätze.

Frontend

Das Frontend einer BI-Architektur bilden die Analysewerkzeuge. Frontend-Werkzeuge können je nach Anwendungsfall mehr oder weniger aufwendig sein. Mithilfe von Werkzeugen für Data Mining und OLAP werden Auswertungen des Datenbestandes durchgeführt und Informationen aus der Masse der Daten gewonnen. Speziell durch Data Mining können bisher unbekannte Zusammenhänge aufgedeckt werden. Hierbei werden beispielsweise Techniken der Klassifikation und Clusterbildung eingesetzt. OLAP-Werkzeuge machen den Datenbestand interaktiv zugänglich. Auswahl und Verdichtungsgrad der angezeigten Daten können durch den Benutzer beeinflusst werden. Beim Informationszugriff kommen in der Regel Portalsysteme zum Einsatz (Kemper/Baars/Mehanna 2010).

2.3 Architekturvarianten

In der Praxis existiert eine Vielzahl von Architekturvarianten, welche teilweise auch aus historisch gewachsenen BI-Landschaften entstanden sind. Im Folgenden werden bewusste Architekturvarianten genannt und anschließend beschrieben, die eine wirksame Managementunterstützung des individuellen Unternehmens sicherstellen:

- unabhängige Data Marts,
- Data Marts mit abgestimmten Datenmodellen,
- zentrales Core Data Warehouse (C-DWH) (Verzicht auf Data Marts),
- mehrere C-DWH,
- C-DWH und abhängige Data Marts,
- DWH-Architektur-Mix.

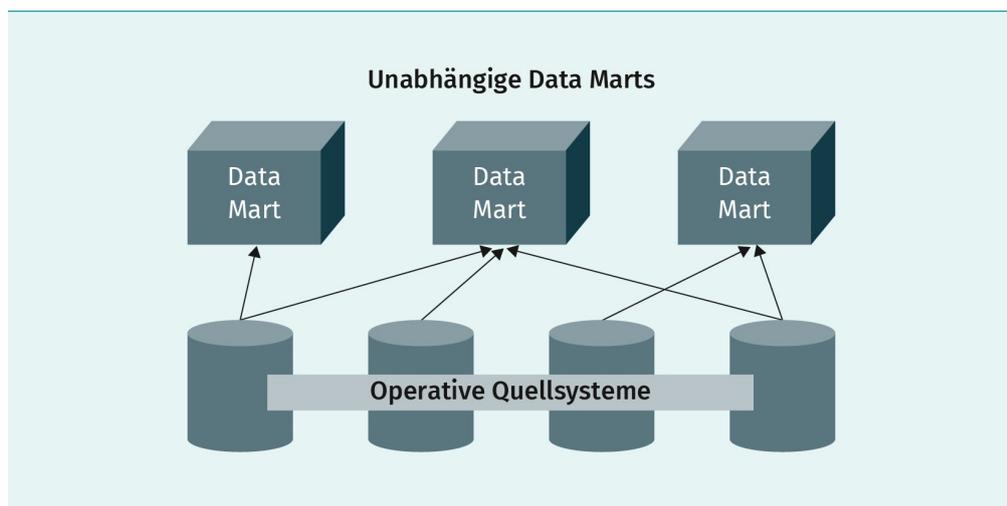
Unabhängige Data Marts

Unabhängige Data Marts

Bei diesen entstehen in den einzelnen Abteilungen voneinander unabhängige DWHs.

Die Architekturform der **unabhängigen Data Marts** entsteht in der Praxis relativ oft, indem einzelne Abteilungen unabhängig voneinander eigene DWHs aufbauen, siehe folgende Abbildung:

Abbildung 5: Unabhängige Data Marts



Quelle: Kemper/Baars/Mehanna 2010, S. 22.

Bei der Verwendung unabhängiger Data Marts wird auf eine zentrale Basisdatenbank (Core Data Warehouse) verzichtet. Dieser Ansatz führt zu einer reduzierten Komplexität des gesamten DWH und macht es damit einfacher und überschaubarer. Hierdurch können in relativ kurzer Zeit nutzbare Ergebnisse für die Abteilungen erzielt werden. Der Aufbau eines unternehmensweiten Data Warehouse wird jedoch aufgrund der entstehenden Insellösungen deutlich schwieriger (Kemper/Baars/Mehanna 2010).

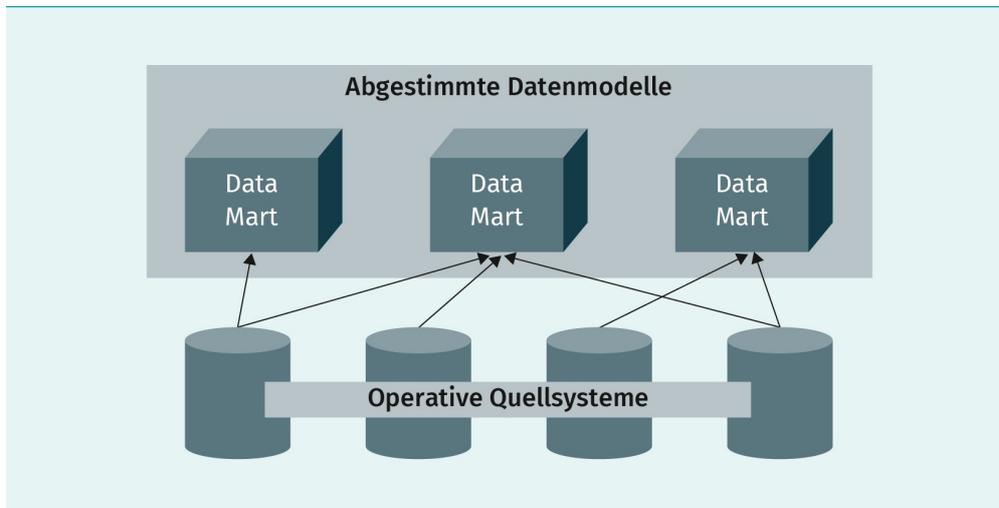
Data Marts mit abgestimmten Datenmodellen

Data Marts mit abgestimmten Datenmodellen

Dabei existieren mehrere Data Marts mit einem gemeinsamen Datenmodell.

Wie in der vorigen Variante werden hier Quelldaten mehrfach für unterschiedliche Datenhaltungssysteme aufbereitet. Allerdings werden die einzelnen **Data Marts** hinsichtlich eines gemeinsamen **Datenmodells** aufeinander **abgestimmt**, siehe folgende Abbildung:

Abbildung 6: Data Marts mit abgestimmten Datenmodellen



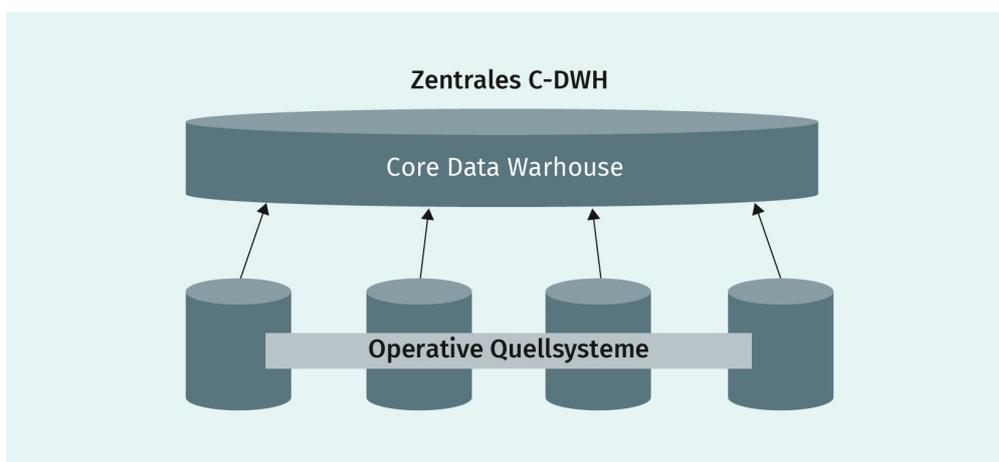
Quelle: Kemper/Baars/Mehanna 2010, S. 22.

Durch den Einsatz von Data Marts mit konzeptionell abgestimmten Datenmodellen wird die Konsistenz und Integrität des dispositiven Datenmodells sichergestellt. Der Aufbau eines unternehmensweiten Data Warehouse wird gegenüber der ersten Variante mit weniger Aufwand verbunden sein (Kemper/Baars/Mehanna 2010).

Zentrales C-DWH (Verzicht auf Data Marts)

Für kleinere BI-Lösungen kann der Verzicht auf Data Marts sinnvoll sein, z. B. bei einer geringen Anzahl von Endbenutzern und Datenvolumen. Hier empfiehlt sich eher die Variante: **Zentrales C-DWH**:

Abbildung 7: Zentrales C-DWH



Quelle: Kemper/Baars/Mehanna 2010, S. 22.

Zentrales C-DWH
Es stellt die Auswertungsfunktion des C-DWH in den Vordergrund.

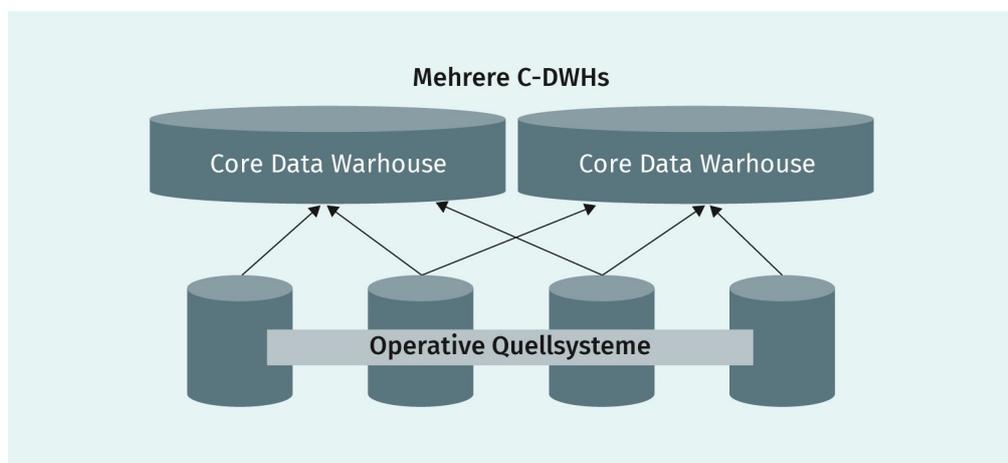
Bei diesem sogenannten monolithischen Ansatz wird die Auswertungsfunktion des Core Data Warehouse in den Vordergrund gestellt. Dieser Ansatz kann jedoch bei komplexen Lösungen mit erheblichen Nachteilen (Performance, Administrationsaufwand) verbunden sein (Kemper/Baars/Mehanna 2010).

Mehrere C-DWHs

Unter bestimmten Rahmenbedingungen im geschäftlichen Unternehmensfeld, z. B. unterschiedlichen Produkt- bzw. Marktstrukturen, können durchaus **mehrere Core Data Warehouses** aufgebaut werden.

Mehrere C-DWH
Diese Variante ist sinnvoll bei spartenorientierten Großunternehmen.

Abbildung 8: Mehrere C-DWHs



Quelle: Kemper/Baars/Mehanna 2010, S. 22.

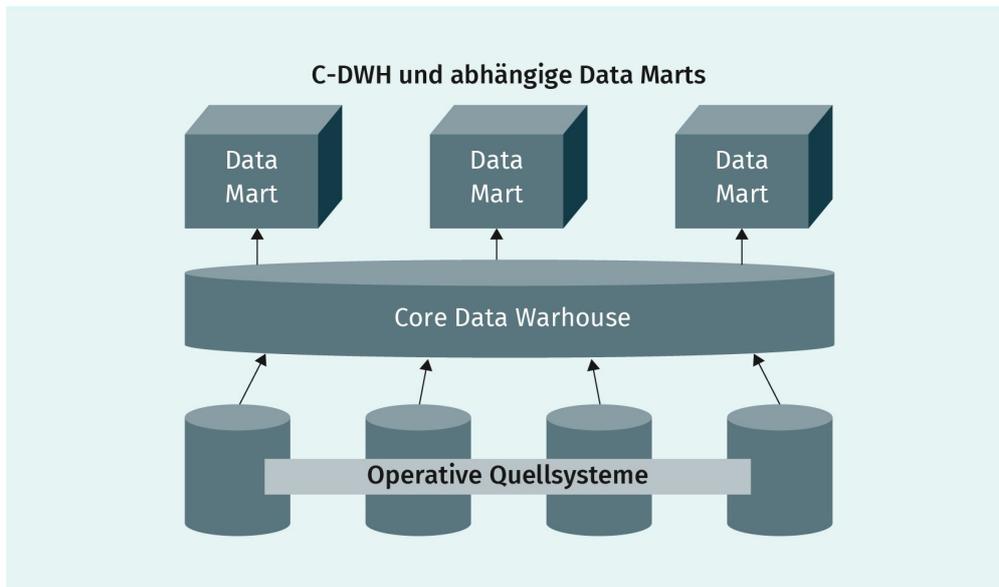
Die genannten Rahmenbedingungen sind insbesondere bei spartenorientierten Unternehmen, Konzernen und Großunternehmen mit diversifiziertem Angebotsspektrum vorzufinden (Kemper/Baars/Mehanna 2010).

C-DWH und abhängige Data Marts

Die Erweiterung des Core Data Warehouse um Data Marts ist die in Lehrbüchern am häufigsten dargestellte Architekturvariante. Die Data Marts werden dabei mithilfe von Transformationsprozessen und Daten aus dem Core Data Warehouse versorgt. In folgender Abbildung wird die Variante „**C-DWH und abhängige Data Marts**“ dargestellt.

C-DWH und abhängige Data Marts
Dies ist die in der Literatur am häufigsten dargestellte Architekturvariante. Ihr Vorteil sind kurze Antwortzeiten.

Abbildung 9: C-DWH und abhängige Data Marts



Quelle: Kemper/Baars/Mehanna 2010, S. 22.

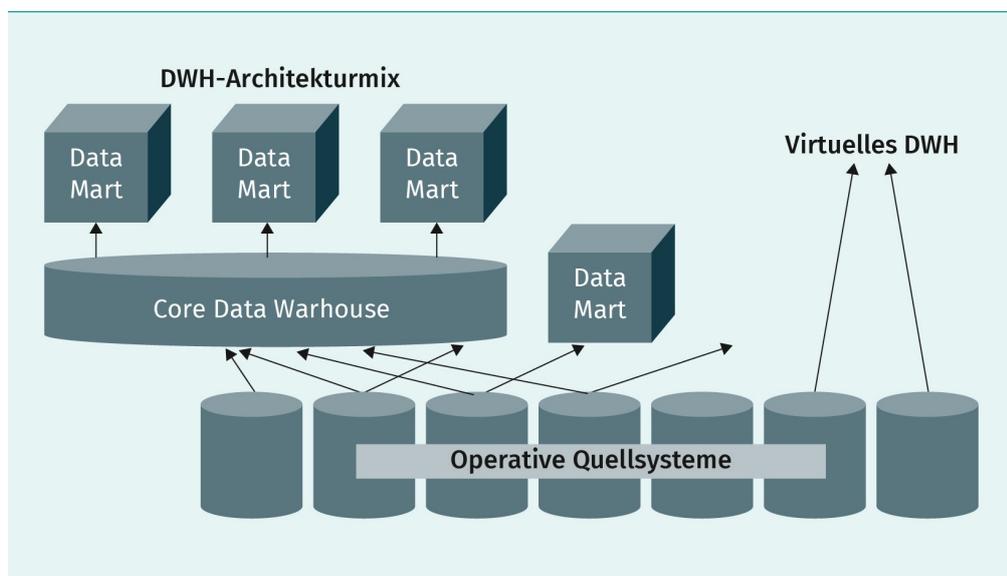
Bei abhängigen Data Marts werden Daten periodisch aus dem C-DWH extrahiert und in Data Marts gespeichert. Bei den extrahierten Daten handelt es sich um kleine, abteilungs-spezifische Datenextrakte aus dem Core Data Warehouse. Durch die Bildung dieser Extrakte ist das Datenvolumen eines Data Marts wesentlich geringer. Hierdurch sind bei Anfragen auf diesen Datenbestand schnellere Antwortzeiten zu erzielen (Kemper/Baars/Mehanna 2010). Die bei abhängigen Data Marts verwendete Struktur wird aufgrund ihres Aussehens oft „Nabe-Speiche-Architektur“ genannt (Bauer/Günzel 2013, S. 69).

DWH-Architektur-Mix

In der Praxis ist häufig ein **DWH-Architekturmix** vorzufinden, der aus C-DWHs, abhängigen und unabhängigen Data Marts sowie direkten Datendurchgriffen (Virtuelles DWH mit eigener Datentransformation) besteht.

DWH Architekturmix
Dieser Mix besteht aus C-DWH, abhängigen und unabhängigen Data Marts sowie direkten Datendurchgriffen.

Abbildung 10: DWH-Architekturmix



Quelle: Kemper/Baars/Mehanna 2010, S. 22.

In der Praxis kann der Ansatz durchaus das Ergebnis einer historisch gewachsenen BI-Landschaft darstellen. Der Ansatz kann aber auch das Ergebnis eines bewussten Gestaltungsprozesses sein, um die optimale Unterstützung wertschöpfender Primärprozesse und flankierender Querschnittsprozesse sicherzustellen (Kemper/Baars/Mehanna 2010).

Im Unternehmensalltag finden sich zahlreiche Data-Warehouse-Systeme, die die klassische Architektur modifizieren und an die jeweiligen Anforderungen anpassen. Beispielsweise können mehrere Data Marts und ein Operational Data Store (ODS) eingesetzt werden, die mit dem unternehmensweiten Core Data Warehouse kooperieren.



ZUSAMMENFASSUNG

In Anlehnung an die Unterteilung von Erich Gutenberg können Anwendungssysteme nach Art der unterstützenden Arbeiten in operative und dispositive Systeme unterschieden werden. Operative Systeme dienen dazu, alltägliche Informationen für ein Unternehmen abzuspeichern und zu verwalten. Das Data Warehouse wird den dispositiven Systemen zugeordnet. DWHs werden eingesetzt, um Informationen aus operativen Daten zu gewinnen. Die zugehörigen Daten können ebenfalls in oben genannter Form klassifiziert werden.

Von der Datenextraktion aus den operativen Quellen bis hin zur Entscheidungsfindung des Managements wird nach den Prozessphasen (1) Datenbereitstellung, (2) Informationsgenerierung, Speicherung, Distribution und (3) Informationszugriff unterschieden.

Eine Business-Intelligence-Architektur besteht aus verschiedenen Komponenten, z. B. Quellsystemen, Staging Area, ODS, C-DWH, Data Mart, ETL, Aggregationen und Frontend.

In der Praxis existieren eine Vielzahl von DWH- bzw. Data-Mart-Architekturvarianten, welche teilweise auch aus historisch gewachsenen BI-Landschaften entstanden sind.

LEKTION 3

DATA WAREHOUSE

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- wie Daten aus verschiedenen operativen Systemen unternehmensweit integriert werden.
- welche Transformationsschritte dafür notwendig sind.
- welche Unterscheidungsmerkmale zwischen C-DWH und Data Mart bestehen.
- welche Funktionen ein Operational Data Store (ODS) anbieten kann.
- inwiefern Metadaten unterstützen können.

3. DATA WAREHOUSE

Einführung

Bevor BI-relevante Daten im Data Warehouse bereitgestellt werden, sind umfangreiche Aktivitäten erforderlich. Analytische BI-Anwendungen verlangen themenbezogene, integrierte Datensammlungen, z. B. nach Kunde, Produkt oder Organisationseinheit. Mithilfe des ETL-Prozesses werden Daten aus operativen Systemen in betriebswirtschaftlich interpretierbare Daten transformiert. Dabei werden erforderliche Daten über lange Zeiträume gespeichert und dem Management in aggregierter Form bereitgestellt. Zudem werden große Datenmengen aus mehreren operationalen Datenbanken konsolidiert, um diese dann im Data Warehouse zu speichern.

Der ETL-Prozess übernimmt die Bereinigung und Transformation von operativen Daten, welche anschließend im Data Warehouse für weitergehende Analysen bereitgestellt werden. Nach der Extraktion von operativen Daten aus den Quellsystemen übernimmt der Transformationsprozess die entscheidende Aufbereitung.

Die Aufbereitung erfolgt in den vier Teilprozessen: Filterung, Harmonisierung, Aggregation und Anreicherung. Anschließend erfolgt das Laden in die Auswertungsebene des DWH.

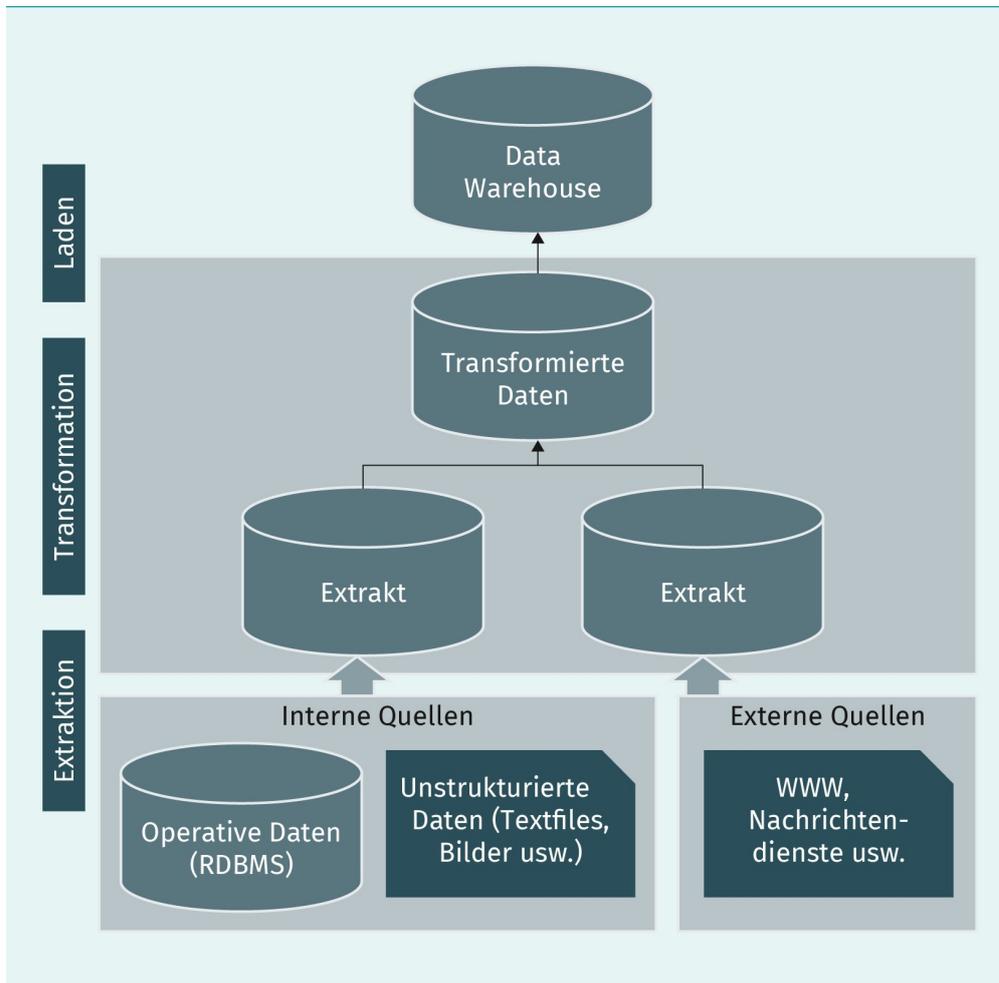
3.1 ETL-Prozess

Um Daten aus mehreren operativen Datenquellen zu vereinigen und aufzubereiten, werden diese mittels gezielter Umwandlungsaktionen in managementrelevante Informationen überführt. Diese Aktion wird in drei Schritten als sogenannter **ETL-Prozess** vollzogen, der in der folgenden Abbildung in abstrakter Form dargestellt wird:

ETL-Prozesse

Durch diese werden operative Daten in managementrelevante Informationen überführt.

Abbildung 11: ETL-Prozess



Quelle: erstellt im Auftrag der IU, 2016.

Dabei werden große Datenmengen aus den Quellsystemen extrahiert, den Anforderungen des DWH entsprechend verarbeitet und anschließend in das DWH geschrieben.

Der Prozess des Übertragens von Daten aus operativen Quellen in das DWH findet üblicherweise periodisch statt und besteht dabei aus folgenden drei Stufen:

1. Extraktion der relevanten Daten aus verschiedenen Quellen,
2. Transformation in ein einheitliches multidimensionales Format,
3. Laden in das Data Warehouse und Bereitstellen für Analysen.

Die Erstellung des ETL-Prozesses ist der aufwendigste Schritt bei der Data-Warehouse-Entwicklung. Der ETL-Prozess ist von zentraler Bedeutung. Dabei ist der Aufbau eines soliden Data Warehouse nur mit einer qualitativ hochwertigen Datenbasis möglich.

Grundsätzlich können ETL-Prozesse individuell programmiert oder mithilfe von Werkzeugen entwickelt werden. Durch die hohe Komplexität von ETL-Prozessen ist der Einsatz eines Werkzeugs in vielen Fällen empfehlenswert (Kimball/Caserta 2004). Die folgenden Abschnitte beschreiben den detaillierten Transformationsprozess.

Bestandteile des Transformationsprozesses

Der Transformationsschritt ist der aufwendigste und komplexeste Teil des Integrationsprozesses. Nach Kemper setzt sich die **Transformation** aus den vier Teilprozessen Filterung, Harmonisierung, Aggregation und Anreicherung zusammen, welche im Folgenden skizziert werden.

Transformation
Sie besteht aus: Filterung,
Harmonisierung, Aggregation
und Anreicherung.

Tabelle 2: Teilprozesse der Transformation

Bestandteile des Transformationsprozesses	
Filterung	Extraktion und Bereinigung syntaktischer und inhaltlicher Defekte der Daten
Harmonisierung	Betriebswirtschaftliche Abstimmung der gefilterten Daten
Aggregation	Verdichtung der gefilterten und harmonisierten Daten
Anreicherung	Berechnung und Speicherung betriebswirtschaftlicher Kennzahlen

Quelle: erstellt im Auftrag der IU, 2016.

Die einzelnen Bestandteile werden nachfolgend detailliert beschrieben. Die beiden ersten Transformationsschritte – Filterung und Harmonisierung – sind für die Bereinigung und Aufbereitung von Daten zuständig, etwa die Angleichung unterschiedlicher Kodierungen und Währungen. Danach sind die Daten prinzipiell bereits für BI-Analysen verwertbar.

Die beiden folgenden Schritte der Aggregation und Anreicherung fassen Daten themenspezifisch zusammen. Zudem werden die Daten um betriebswirtschaftliche Kennzahlen erweitert. Die so erzeugten und ins Data Warehouse geladenen Daten erhalten damit bereits eine bestimmte Ausrichtung auf die Bedürfnisse einzelner Nutzergruppen und deren Analyseziele.

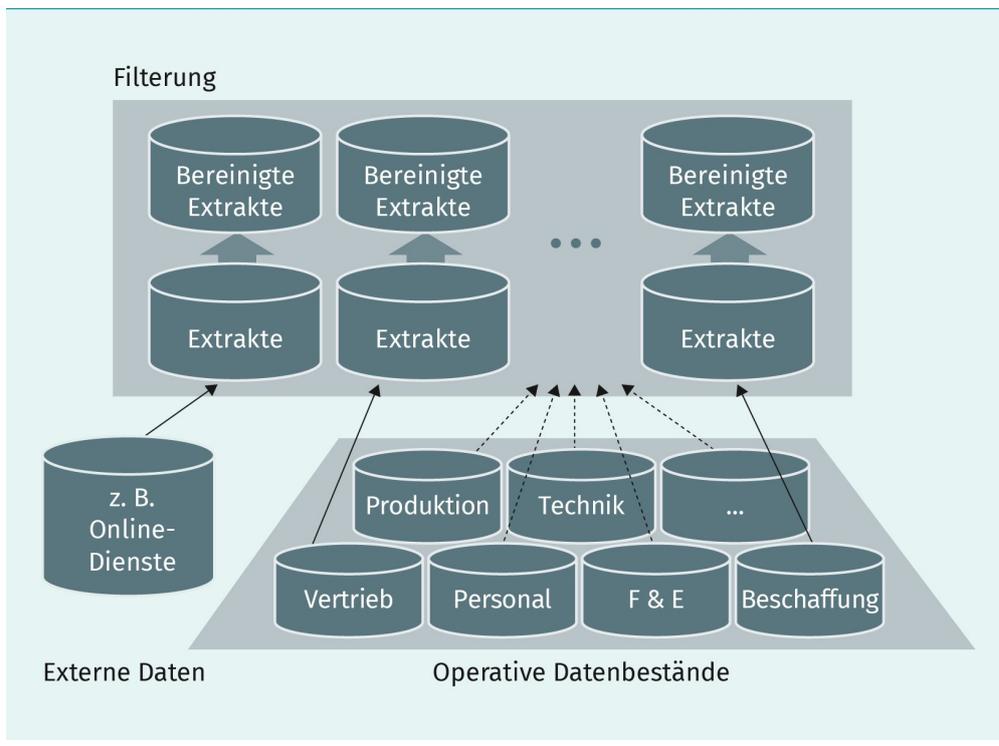
Transformation 1: Filterung

Filterung
Dieser Teilprozess umfasst die Zwischenspeicherung von Extrakten und Bereinigung der Daten.

Mithilfe der **Filterung** werden die für das DWH benötigten Daten selektiert, zwischengespeichert und von Mängeln befreit. Die Filterung unterteilt sich in Extraktion und Bereinigung. Bei der Extraktion werden die Daten in die speziell hierfür vorgesehenen Extraktionsbereiche (Staging Area) eingestellt. Die Bereinigung dient der Befreiung von syntaktischen und semantischen Mängeln.

In folgender Abbildung wird die Filterung als Teilprozess des Transformationsprozesses abgebildet.

Abbildung 12: Erste Transformationsschicht - Filterung



Quelle: Kemper/Baars/Mehanna 2010, S. 28.

Die Bereinigung der Daten dient der Behebung von Mängeln und somit der Erreichung einer definierten Datenqualität. Die Bereinigungen ist notwendig, da operative Systeme nicht unbedingt korrekte Daten enthalten. Die Ursachen für fehlerhafte Daten können vielfältig sein, z. B. falsche Eingaben durch Anwender, Systemfehler oder Evolution der Systeme.

Bei den zu behebbenden Mängeltypen lassen sich syntaktische (technische) und semantische (inhaltliche) Mängel unterscheiden. Syntaktische Mängel sind formale Fehler wie falsche Steuerzeichen, alphanumerische Werte in einem numerischen Feld, NULL-Werte in einem NOT NULL-Feld oder Werte außerhalb des Wertebereiches. Semantische Mängel sind Fehler betriebswirtschaftlicher Art, z. B. offensichtlich falsche Umsatzzahlen.

Mängelklassen

In der Literatur werden Mängel der 1., 2. und 3. Klasse unterschieden. Mängel der 1. Klasse lassen sich automatisch erkennen und während des Extraktionsvorgangs auch automatisch korrigieren. Bei Mängeln der 2. Klasse erfolgt zwar die Defekterkennung automatisch, die Korrektur muss aber manuell nach dem Extraktionsvorgang vorgenommen werden. Mängel der 3. Klasse können ausschließlich manuell erkannt und korrigiert werden.

Tabelle 3: Mängelklassifikation im Rahmen der Bereinigung

	Klasse 1	Klasse 2	Klasse 3
Bereinigung	Automatische Erkennung und Behebung	Automatische Erkennung/ manuelle Korrektur	Manuelle Erkennung und manuelle Korrektur
Syntaktische Mängel	Bekannte Formatanpassung	Erkennbare Formatinkompatibilitäten	-
Semantische Mängel	Fehlende Datenwerte	Ausreißerwerte/ unstimmmige Wertkonstellationen	Unerkannte semantische Fehler in Quelldaten

Quelle: Chamoni/Gluchowski 2015, S. 135.

Die grundsätzlich automatisch erkennbaren Mängel der 1. Klasse sind durch bestimmte Algorithmen zu korrigieren. Beispielsweise können auf syntaktischer Ebene interne Format-, Steuer- und Sonderzeichen während der Extraktion identifiziert und durch Zuordnungstabellen (Mapping-Tabellen) in den extrahierten Daten bearbeitet werden.

Analoges gilt für semantische Fehler. Wurden beispielsweise bei der Übertragung von Umsatzdaten einzelne Filialen vergessen, können diese durch Äquivalenzwerte wie monatliche Planwerte oder Ist-Werte des Vormonates ergänzt werden.

Mängel der 2. Klasse können ebenfalls automatisch erkannt werden, müssen aber manuell von Technikern oder Betriebswirten korrigiert werden. Bei syntaktischen Mängeln sind dies etwa bislang nicht berücksichtigte Syntaxvarianten in den operativen Datenquellen. Künftig können diese dann automatisiert behandelt werden.

Auf semantischer Ebene können automatisierte Plausibilitätskontrollen und Wertebereichsüberprüfungen ungültige Datenfelder entdecken. Beispielsweise durch Vergleich von Bilanz- und Kontrollsummen. Je nach Schwere des Fehlers müssen eventuell auch die operativen Quellen korrigiert werden.

Während syntaktische Mängel immer automatisiert erfasst werden können, trifft dies nicht auf semantische Mängel zu. Mängel der 3. Klasse betreffen deshalb ausschließlich semantische Fehler. Dies sind alle Mängel, die sich nicht durch die Prüfverfahren von Klasse-2-Fehlern entdecken lassen, also weder durch Plausibilitäts- noch Wertebereichsprüfungen. Vielmehr lassen sich diese Fehler nur durch betriebswirtschaftliche Experten identifizieren. Auch hier müssen eventuell die operativen Quellen korrigiert werden.

Transformation 2: Harmonisierung

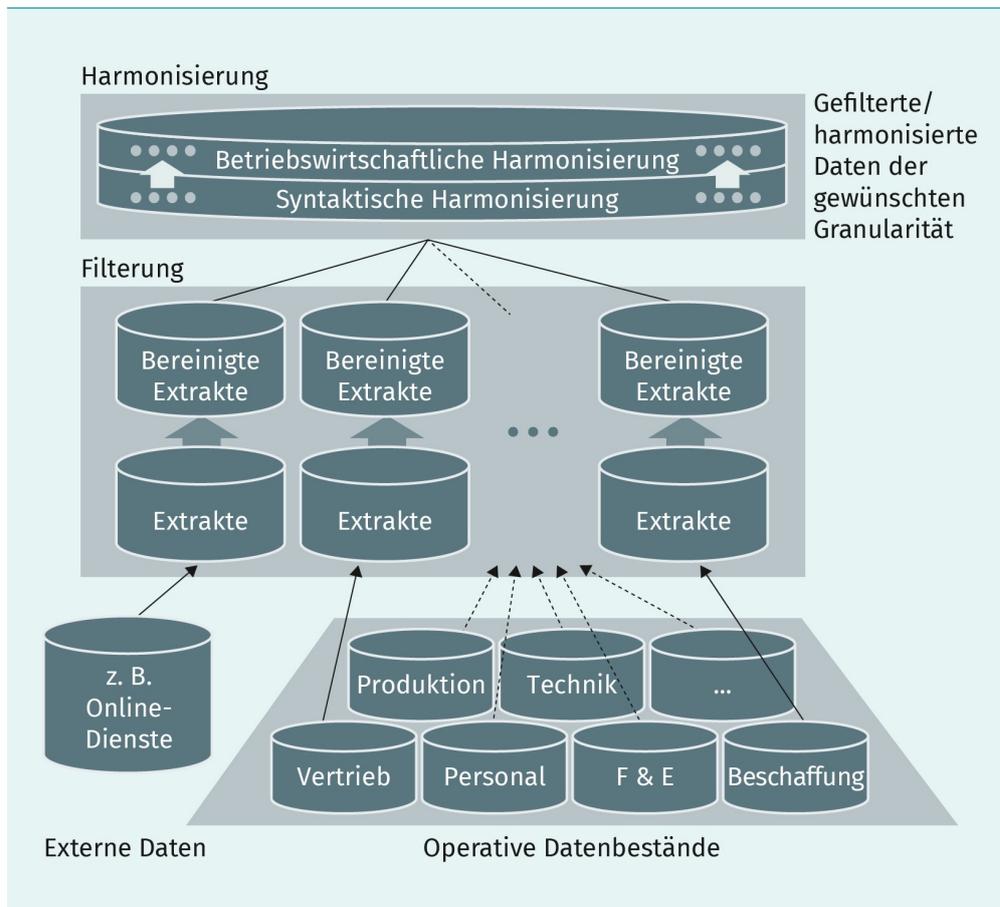
Harmonisierung
Dieser Teilprozess umfasst die betriebswirtschaftliche Abstimmung gefilterter Daten.

Der zweite Transformationsschritt nach der Filterung behandelt die **Harmonisierung** der Daten. Die Harmonisierung bezeichnet dabei den Prozess der betriebswirtschaftlichen Abstimmung gefilterter Daten. Die Harmonisierung ist notwendig, falls Daten aus unterschiedlichen Quellsystemen integriert werden. In den heterogen gewachsenen Quellsystemen werden für gleiche Sachverhalte bzw. Eigenschaften oftmals unterschiedliche Schlüssel oder Ausprägungen angewandt. Das klassische Beispiel ist der Schlüssel für das

Geschlecht, der in drei Systemen unterschiedlich als männlich/weiblich, M/W und als 0/1 dargestellt wird. Das Ziel der Harmonisierung ist demnach, die gleichen Sachverhalte auf einen gemeinsamen Schlüssel zu bringen.

Weiterhin werden unterschiedliche Messgrößen in ein gemeinsames Maß übertragen, z. B. werden verschiedene Währungen in eine einheitliche Währung umgerechnet (Kimball/Caserta 2004).

Abbildung 13: Zweite Transformationsschicht - Harmonisierung



Quelle: Kemper/Baars/Mehanna 2010, S. 32.

Im Teilprozess „Harmonisierung“ werden die gefilterten und bereinigten Daten zusammengeführt. Dabei wird nach folgender syntaktischer und betriebswirtschaftlicher Harmonisierung unterschieden:

Syntaktische Harmonisierung

Zur syntaktischen Harmonisierung gehören die sogenannten Schlüsselharmonien sowie Kodierungen, Synonyme und Homonyme, auf die wir hier detailliert eingehen.

Schlüsselharmonien: Bei der Harmonisierung müssen unter anderem Schlüsselharmonien aufgelöst werden. Grundsätzlich ist ein gemeinsamer Schlüssel notwendig, wenn Daten aus mehreren Datenbanken zusammengeführt werden. Das Problem wird zumeist mithilfe einer Mapping-Tabelle gelöst, die beispielsweise für jeden Kunden einen neuen, künstlichen Primärschlüssel generiert.

Die Primärschlüssel der operativen Systeme werden dann als Fremdschlüssel mitgeführt, sodass hierüber Auswertungen möglich sind. Die Schlüssel von Datensätzen müssen innerhalb der Basisdatenbank, DWH und Data Mart eindeutig sein. Die in Quellsystemen vorhandenen Schlüssel erfüllen dieses Kriterium aufgrund ihrer Heterogenität und der Verteilung der Daten in der Regel nicht. Während der Transformationsphase müssen daher globale, eindeutige Schlüssel zugewiesen werden. Diese globalen Schlüssel werden als „Surrogate Keys“ bezeichnet. Moderne ETL-Werkzeuge verfügen über standardisierte Transformationen, die eindeutige Surrogate Keys generieren.

Die Abbildung von lokalen Schlüsseln auf globale Surrogate Keys ist zu dokumentieren, um flexibel auf Änderungen reagieren zu können. Neben der Vereinheitlichung der Daten spielen Surrogate Keys bei der Historisierung eine wichtige Rolle (Kimball/Caserta 2004). Beispielsweise liegen in zwei Datenquellen (CRM, ERP) Kundendaten vor, die in eine Tabelle integriert werden müssen. Beide Quelldatensätze verfügen über einen Kundenschlüssel. Für die Zieldatenbank ist ein globaler Schlüssel für Kundendaten einzuführen. Hierfür werden während der Transformation die Schlüssel der Quelltabellen entfernt und durch Surrogate ersetzt.

Kodierungen, Synonyme und Homonyme: Neben den Schlüsselharmonien müssen zudem Kodierungen, Synonyme und Homonyme aufgelöst werden. Hier dazu einige Beispiele:

- Einzelne Datenbestände können unterschiedlich kodiert sein (Kodierung). So können Attribute wie Geschlecht in Datenquelle 1 als M/W kodiert sein, in Datenquelle 2 als 0/1.
- Unterschiedliche Attributnamen können die gleiche Bedeutung haben (Synonymie). Beispielsweise kann in Datenquelle 1 für den Namen von Betriebsmitarbeitern das Attribut „Personal“ vorgesehen sein, in Datenquelle 2 aber „Mitarbeiter“.
- Umgekehrt können gleiche Attributnamen unterschiedliche Bedeutungen haben (Homonymie). In Datenquelle 1 kann „Partner“ beispielsweise den Namen von Kunden bezeichnen, in Datenquelle 2 den Namen von Lieferanten.

In allen drei Fällen müssen die Daten harmonisiert werden. Im ersten Fall muss der Attributwert einheitlich z. B. auf 0/1-Werte festgelegt werden, im zweiten Fall ist ein identischer Attributname zu wählen und im dritten Fall ein unterschiedlicher Attributname.

Auch hier werden für den Abgleich in der Regel Mapping-Tabellen implementiert, welche die gefilterten Daten über Namensabgleichungen und Kodierungsabstimmungen zu themenorientierten Datensammlungen zusammenführen.

Betriebswirtschaftliche Harmonisierung

Neben der syntaktischen Abgleichung erfolgt die Vereinheitlichung betriebswirtschaftlicher Begriffe. Hierbei handelt es sich weniger um ein technisches, sondern um ein betriebswirtschaftlich-organisatorisches Problem. Dabei müssen die operativen Daten (z. B. Währung) in einheitliche Werte überführt werden. Monetäre Größen verschiedener Währungen müssen in einem einheitlichen Währungssystem vorliegen.

Für die entsprechenden Aktivitäten lassen sich Transformationsregeln implementieren, welche die Harmonisierung umsetzen. Nach Abschluss des Teilprozesses „Harmonisierung“ liegen im Data Warehouse bereinigte und konsistente Daten vor, die bereits für Analysezwecke nutzbar sind.

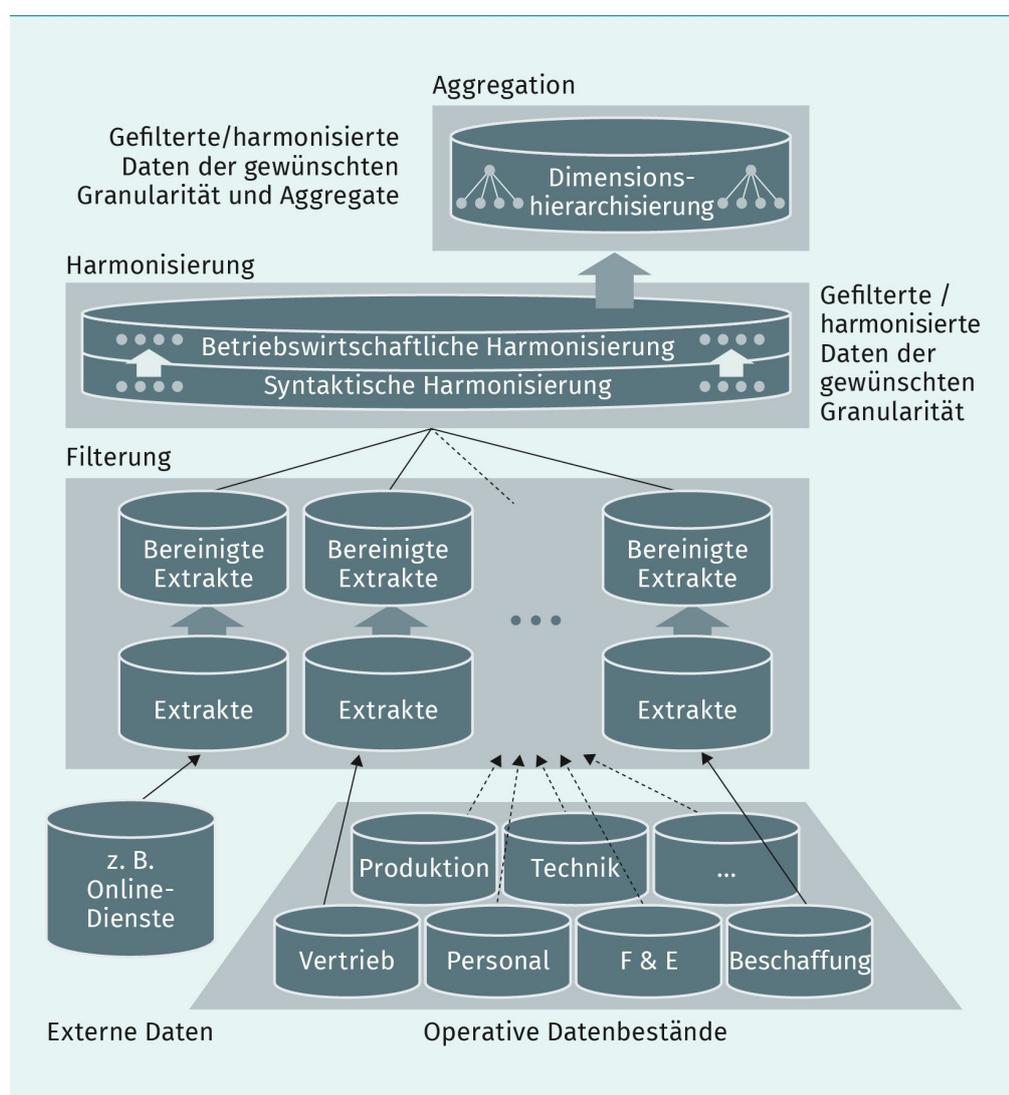
Transformation 3: Aggregation

Mithilfe des Teilprozesses **Aggregation** werden gefilterte und harmonisierte Daten verdichtet und in die gewünschte Granularität überführt.

Aggregation

Dieser Teilprozess umfasst die Verdichtung von gefilterten und harmonisierten Daten.

Abbildung 14: Dritte Transformationsschicht - Aggregation



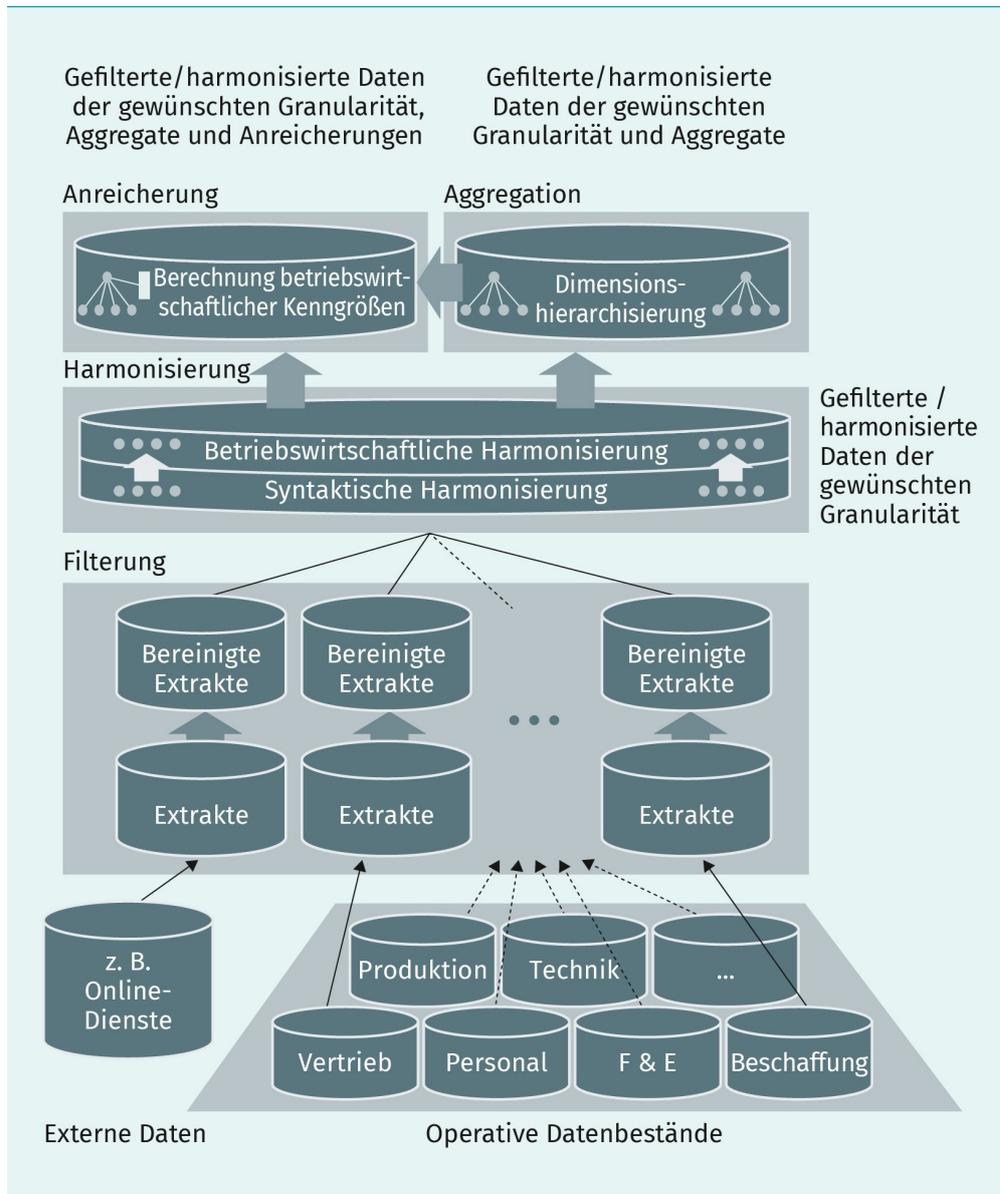
Quelle: Kemper/Baars/Mehanna 2010, S. 37.

Für die Bildung von tagesaktuellen Daten auf der Basis von Produkt- und Kundengruppen müssen beispielsweise alle Einzeldaten über Aggregationsalgorithmen zu produktgruppen- und kundengruppenspezifischen Werten auf Tagesbasis zusammengefasst werden. Zudem werden Vorkommungen, z. B. für betriebswirtschaftliche Kennzahlen, durchgeführt. Das Ziel von Verdichtungen ist, Summenwerte zu erzeugen, die aus Performancegründen vorberechnet im Data Warehouse vorgehalten werden.

Transformation 4: Anreicherung

Die Bildung und Speicherung betriebswirtschaftlicher Kennzahlen aus gefilterten und harmonisierten Daten wird als Anreicherung bezeichnet. Der Teilprozess **Anreicherung** ist der letzte Schritt im Transformationsprozess.

Abbildung 15: Vierte Transformationsschicht - Anreicherung



Anreicherung

Dieser Teilprozess umfasst die Bildung von betriebswirtschaftlichen Kennzahlen nach der Harmonisierung bzw. Aggregation.

Quelle: Kemper/Baars/Mehanna 2010, S. 38.

Wie bereits angedeutet, werden die vorhandenen Daten um betriebswirtschaftliche Kennziffern angereichert. Dabei werden Berechnungen durchgeführt und Ergebnisse den übrigen Daten hinzugefügt. Hierbei werden insbesondere Kennziffern gespeichert, die für mehrere Anwender relevant sind. Beispielsweise können wöchentliche Deckungsbeiträge auf Produktebene oder jährliche Deckungsbeiträge auf Filialebene kalkuliert und integriert werden. Ersteres kann für den Produktmanager interessant sein, letzteres für Filialleiter und Geschäftsführung. Die Aufnahme dieser Kenngrößen in den Datenbestand hat

mehrere Vorteile. Aufgrund der Vorberechnungen können Abfragen effizienter durchgeführt werden. Zudem sind die errechneten Werte aufgrund der einmaligen Berechnung konsistent. Außerdem liegt ein abgestimmtes betriebswirtschaftliches Konzept vor.

Die Hauptaktivitäten des ETL-Prozesses sind mit der Extraktion und insbesondere der Transformation abgeschlossen. Die genannten Teilprozesse werden in der sogenannten Staging Area durchgeführt. Nach der Transformation werden die Daten während der Lade-phase in das Zielsystem geschrieben.

3.2 DWH und Data Mart

Das DWH umfasst im engeren Sinne die Datenhaltung. Einzelne Komponenten des DWH sind demnach die Staging Area, Basisdatenbank, Data Mart, ODS und Metadaten, welche im Folgenden detailliert dar- und teilweise auch gegenübergestellt werden.

Staging Area

Staging Area
In dieser werden Extrakte zwischengespeichert, um nachgelagerte Systeme zu entlasten.

Nach Inmon (2005) ist die Staging Area ein Arbeitsbereich, in welchem Daten temporär zwischengespeichert werden. Die Extraktion in die **Staging Area** findet üblicherweise periodisch statt. Anschließend werden Transformationen innerhalb der Staging Area durchgeführt. Grundsätzlich werden die Daten aus der Staging Area gelöscht, nachdem diese in nachgelagerte Systeme geladen wurden.

Der Zweck der Staging Area ist es, nachgelagerte Systeme (z. B. die Basisdatenbank) zu entlasten. Daher ist der separate Arbeitsbereich, insbesondere bei der Verarbeitung großer Datenmengen und komplexer Transformationen, von besonderer Bedeutung (Inmon 2005).

C-DWH

Funktionalität

Basisdatenbank
Sie enthält integrierte Daten für nachgelagerte Systeme.

Die **Basisdatenbank** ist ein Datenspeicher, der zwischen Staging Area und Auswertungsdatenbank angesiedelt ist. Dabei dient die Datenbank dem Zweck, detaillierte, historische, konsistente und normalisierte Daten für nachgelagerte Systeme bereitzustellen. Die Abgrenzung zur Staging Area liegt im Zustand der Daten. Diese werden in der Basisdatenbank integriert und bereinigt gespeichert.

In Bezug auf die Auswertungsdatenbank wird die Basisdatenbank in erster Linie über das Datenmodell abgegrenzt. In der Basisdatenbank werden Daten normalisiert und abfrage-neutral vorgehalten. Außerdem werden die Daten in der kleinsten erforderlichen Granularität gespeichert. Je nach Anforderungen werden Daten auch historisiert (Bauer/Günzel 2013, S. 586).

Das Core DWH enthält somit die folgenden Funktionen:

- **Sammel- und Integrationsfunktion:** Sie steht für die Aufnahme von unternehmensweiten Daten, die für die spätere Analyse erforderlich sind.
- **Distributionsfunktion:** Sie steht für die Datenversorgung von nachgelagerten Systemen (z. B. Data Marts).
- **Qualitätssicherungsfunktion:** Die transformierten Daten sichern die syntaktische und semantische Stimmigkeit der dispositiven Datenbasis.

Aktualisierungsstrategie

Die Aktualisierung des Core Data Warehouse erfolgt nach den jeweiligen Anforderungen an das BI-System. Grundsätzlich werden die folgenden drei Aktualisierungsvarianten unterschieden:

- **Aktualisierung in Abhängigkeit der Änderungsquantität:** Hier werden die aufgelaufenen Änderungen aus den Quellsystemen zunächst gesammelt. Die Daten werden in das C-DWH geladen, sobald eine definierte Anzahl von Änderungen erreicht ist.
- **Aktualisierung in periodischen Zeitabständen:** Die Aktualisierung erfolgt nach Bedarf des jeweiligen Anwendungsbereichs, z. B. stündlich, täglich, wöchentlich oder monatlich.
- **Echtzeit-Aktualisierung:** Dabei können Daten transaktionssynchron in das C-DWH geladen werden. Bei diesem komplexen Verfahren werden in der Regel spezielle Beladungssysteme eingesetzt, die einer ereignisgesteuerten Logik folgen. Dabei werden Daten im sogenannten Push-Betrieb in das Data Warehouse geladen.

Data Mart

Definition

Um das Handling mit komplexen C-DWH zu vereinfachen, werden **Auswertungsdatenbanken** eingeführt. Data Marts sind Ausschnitte eines C-DWH, kleinere Datenpools für Anwendungen, die spezifische Nutzergruppen wie bestimmte Abteilungen oder Aufgabenbereiche bedienen.

Auswertungsdatenbank
Diese enthält Ausschnitte des C-DWH.

Die Daten für Data Marts werden meist mit speziellen Transformationsprozessen aus dem Core Data Warehouse in kleine, überschaubare Einheiten extrahiert. Dies können beispielsweise alle relevanten Daten einer Region oder einer bestimmten Produktgruppe sein. Der Vorteil liegt darin, dass nicht die komplette Datenbasis eines Unternehmens abgebildet wird, sondern nur die Daten, welche für Fragestellungen des jeweiligen Bereichs oder der jeweiligen Abteilung benötigt werden. Die Vertriebsabteilung hat meist lediglich Interesse an Kennzahlen wie Verkaufszahlen, Umsatz oder Provisionen. Eine Fertigungsabteilung hingegen interessiert sich für Produktionsstückzahlen oder Fertigungszeiten.

Die Auswertungsdatenbank bildet demnach die Grundlage für nachgelagerte Analysewerkzeuge. Im Data Mart werden Daten vollständig integriert und bereinigt gespeichert. Die Daten werden analyseorientiert vorgehalten. Technisch liegen Auswertungsdatenbanken in der Regel relationale Datenbanken zugrunde. Multidimensionale Speichermodelle werden seltener eingesetzt.

Grundsätzlich werden die Daten in einem multidimensionalen Modell strukturiert. Die für Analysen benötigten Daten werden durch ETL-Prozesse aus der Basisdatenbank extrahiert und in die Auswertungsdatenbank geladen. Da die Basisdatenbank bereits integrierte und bereinigte Daten vorhält, müssen die Daten vor dem Laden nur noch in das Zielschema transformiert und eventuell aggregiert werden.

In der Praxis werden häufig mehrere Auswertungsdatenbanken (Data Marts) eingesetzt. Dabei werden Daten nach Analyseanforderungen oder Organisationseinheiten aufgeteilt. Außerdem können Sicherheits- oder Datenschutzaspekte eine Verteilung der Daten auf mehrere Data Marts notwendig machen. Wird in einer konkreten Architektur keine Basisdatenbank (Core Data Warehouse) betrieben, müssen Integration und Bereinigung bei der Überführung in die Auswertungsdatenbank durchgeführt werden (Bauer/Günzel 2013, S. 587).

Charakteristika

In der folgenden Tabelle werden die Charakteristika von Data Marts und C-DWH zusammenfassend gegenübergestellt.

Tabelle 4: Data Marts und Core Data Warehouse

Charakteristika	Data Markt	Core Data Warehouse
Betriebswirtschaftliches Ziel	Effiziente Unterstützung der Entscheider einer Abteilung, ausgerichtet alleinig auf deren Analyseanforderungen	Effiziente Managementunterstützung durch strategische, taktische und operative Informationsobjekte für alle Entscheider in einem Unternehmen
Ausrichtung	Abteilungsbezogen	Zentral, unternehmensweit
Granularität der Daten	Zumeist höher aggregierte Daten	Kleinster Grad der Detaillierung
Semantisches Datenmodell	Semantisches Modell ist auf vorab modellierte Analyseanforderungen festgelegt	Semantisches Modell ist auch für zukünftige Analyseanforderungen offen
Modellierungskonventionen	<ul style="list-style-type: none"> • Heterogen (proprietäre Data Marts, jede Abteilung hat ihre eigenen Konventionen); • Einheitlich (abgeleitete Data Marts, Konventionen des Core Data Warehouse werden übernommen) 	Einheitlich
Verwendete OLAP-Technologie	M-OLAP (proprietäre Data Marts); R-OLAP bzw. H-OLAP (abgeleitete Data Marts)	R-OLAP
Direkter Zugriff durch Endanwender	In der Regel möglich	Häufig nicht erlaubt; zentraler Betrieb des C-DWH durch IT-Abteilung; dient als Quelldatensystem für Data Marts

Charakteristika	Data Markt	Core Data Warehouse
Freiheitsgrade der Analysen	Eher gering (Anwender kann über die Abteilungsgrenzen nicht hinaussehen)	Flexibel; sämtliche zugänglichen (Sicherheit) Informationen können in Analysen einfließen
Einfluss von externen Datenquellen	Zumeist nicht gegeben, wenn ja, dann nur spezifischer Ausschnitt	Hoch; sämtliche verfügbaren externen Datenquellen werden integriert, um die Qualität der Analysen verbessern zu können
Datenvolumen	Gering bis moderat	Von moderat bis sehr umfangreich (bis in den Petabyte-Bereich)

Quelle: Kurz 1999, S. 110f.

Abschließend ist festzuhalten, dass in der Praxis Auswertungsdatenbanken eingeführt werden, um das Handling mit komplexen C-DWH zu vereinfachen.

3.3 ODS und Metadaten

ODS – Definition und Merkmale

In neuen Data-Warehouse-Ansätzen ist häufig zusätzlich ein spezieller Daten-Pool integriert, der als **Operational Data Store (ODS)** bezeichnet wird. Grundsätzlich kann der ODS aus architektonischer Sicht als Vorstufe eines DWH aufgefasst werden.

Der ODS beinhaltet aktuelle Daten auf Transaktionsbasis, die aus verschiedenen operativen Quellsystemen entstammen. Die Daten werden für Anwendungs- und Auswertungsdienste bereitgestellt. In der Regel werden die Daten mithilfe zusätzlicher Transformationsprozesse aus dem Core Data Warehouse extrahiert.

In den ODS wird ein sehr kleiner und zeitpunktaktueller Ausschnitt entscheidungsrelevanter Daten übertragen. Diese Daten sind oft schon an die Anforderungen der Analysesysteme angepasst.

Im Folgenden werden die einzelnen Merkmale von ODS erläutert:

- Themenorientierung,
- Integration,
- Zeitpunktbezug,
- Volatilität,
- hoher Detaillierungsgrad.

Der Operational Data Store wird anhand von entscheidungsorientierten Perspektiven konzipiert. Dimensionen sind z. B. Produkte und Regionen (Themenorientierung). Die unternehmensweiten Daten wurden mithilfe von entsprechenden Transformationen einheitlich in den ODS übertragen. Der Transformationsprozess in den ODS beinhaltet primär die Filterung und Harmonisierung (Integration).

Operational Data Store (ODS)
Dieser beinhaltet – als Vorstufe des DWH – aktuelle Transaktionsdaten für Auswertungszwecke.

Im ODS werden grundsätzlich keine Historisierungen durchgeführt. Aufgrund dessen sind keine zeitraumbezogenen Auswertungen möglich. Aus Recovery-Gründen werden Daten jedoch über eine Zeitspanne von mehreren Tagen oder Wochen vorgehalten (Zeitpunktbezug). Grundsätzlich wird eine regelmäßige Aktualisierung durchgeführt, wobei diese konsequenterweise (keine Historisierung) zu einem Überschreiben führt (Volatilität).

Die Daten werden detailliert vorgehalten, da Analysen im ODS zumeist im operativen Kontext stehen. Eine detaillierte Speicherung bedeutet, dass Daten auf der Transaktionsebene abgelegt werden (hoher Detaillierungsgrad).

Metadaten

Für die Analyse von Daten ist von besonderer Bedeutung, dass Anwender wissen, was sich eigentlich hinter den jeweiligen Datenfeldern verbirgt. Die Informationen zu diesen Daten werden in Metadaten bereitgestellt.

Differenzierung

Metadaten enthalten Informationen über die im Data Warehouse gespeicherten Daten und deren Verarbeitung und bieten Unterstützung bei Aufbau, Verwaltung und Betrieb von DWH-Systemen.

Grundsätzlich können Metadaten in passive und aktive Metadaten unterschieden werden. Die hier beschriebene Unterscheidung bezieht sich auf die Nutzung.

Passive Metadaten

Diese dokumentieren Daten und deren Verhältnis zur Umwelt.

- Die **passiven Metadaten** dokumentieren in erster Linie die ihnen zugrunde liegenden Daten sowie deren Verhältnis zur Umwelt. Sie dienen zur Speicherung der Definitionen. Definiert werden Struktur, Entwicklungsprozess und Datenverwendung. Nutzer passiver Metadaten sind alle im BI-Umfeld Tätigen: Anwender, Administratoren, Entwickler (Kemper/Baars/Mehanna 2010).
- Demgegenüber stellen **aktive Metadaten** Methoden dar, die auf Daten ausgeführt werden. Transformationsregeln für ETL-Vorgänge können als aktive Metadaten verstanden werden. Aktive Metadaten können zur Laufzeit interpretiert werden und werden genutzt, um auf Transformations- und Analyseprozesse Einfluss zu nehmen (Kemper/Baars/Mehanna 2010).

Aktive Metadaten

Diese stellen Methoden dar, die auf Daten ausgeführt werden.

Des Weiteren lassen sich technische und betriebswirtschaftliche Metadaten unterscheiden:

Technische Metadaten

Diese fokussieren die Filterung.

Betriebswirtschaftliche Metadaten

Diese fokussieren die Harmonisierung, Aggregation, Anreicherung und Berechtigungsverwaltung.

- **Technische Metadaten** konzentrieren sich auf IT-orientierte Aspekte der Transformationsschicht 1 (Filterung).
- **Betriebswirtschaftliche Metadaten** fokussieren demgegenüber die Schichten 2–4 (Harmonisierung, Aggregation, Anreicherung) und die Berechtigungsverwaltung. (Kemper/Baars/Mehanna 2010).

Vorteile

Mithilfe der Metadaten kann eine effiziente Gestaltung von Entwicklungs- und Betriebsprozessen gewährleistet werden. Außerdem erfolgt eine Effektivitätssteigerung bei der Nutzung von BI-Systemen.

Neben ihrer Aufgabe, Informationen bereitzustellen, dienen die Metadaten darüber hinaus dem Data-Warehouse-Manager als Steuerungselement. Beispielsweise werden vollständig ausführbare Spezifikationen von Datenverarbeitungsschritten als Metadaten gespeichert und zum Ausführungszeitpunkt vom entsprechenden Werkzeug interpretiert und ausgeführt.

Im Folgenden werden die Vorteile der Metadaten, insbesondere für Entwicklung und Betrieb, aufgezeigt:

- Anpassung an Quellsysteme,
- Harmonisierung der Daten aus heterogenen Quellsystemen,
- Wartung und Wiederverwendung,
- Berechtigungsverwaltung,
- Datenqualität,
- Begriffsverständnis.

In der Transformationsschicht 1 (Filterung) werden operative Daten in das DWH übernommen. Die durchzuführenden Extraktions- und Bereinigungsprozesse können mithilfe von Metadaten dokumentiert werden. Hierdurch können Prozesse einfacher modifiziert bzw. erweitert werden (Anpassung der Quellsysteme).

In der Transformationsschicht 2 (Harmonisierung) werden Daten syntaktisch und semantisch integriert. Die Transformationsaktivitäten können durch Informationen über Struktur und Bedeutung der Quellsysteme erleichtert werden. Hierdurch kann eine effiziente Weiterentwicklung des BI-Systems sichergestellt werden. (Harmonisierung der Daten aus heterogenen Quellsystemen).

Durch die Speicherung von Metadaten wird die Wartung und Weiterentwicklung von BI-Systemen vereinfacht. Betriebswirtschaftliche und technische Änderungen können, aufgrund konsistenter Metadaten, schnell und widerspruchsfrei durchgeführt werden. Hierbei kann beispielsweise die Wiederverwendbarkeit von Datenmodellen und Transformationsprozessen unterstützt werden (Wartung und Wiederverwendung).

Die Berechtigungsverwaltung stellt eine zentrale Komponente der dispositiven Datenhaltung dar. Mithilfe von Metadaten werden Benutzerrollen beschrieben, die eine konsistente Zugriffsadministration erlauben. Dadurch wird eine einfache Administration von Beziehungen zwischen BI-Nutzern, Anwendungen und Datenberechtigungen ermöglicht (Berechtigungsverwaltung).

Grundsätzlich können Metadaten über den gesamten Transformationsprozess bereitgestellt werden, um eine möglichst hohe Transparenz für den Anwender herzustellen. Beispielsweise können Verantwortlichkeiten, Qualität der Quellsysteme, Harmonisierungs-

prozesse und Anreicherungen dokumentiert werden. Hierdurch kann insbesondere die Datenqualität bezüglich Konsistenz, Aktualität, Genauigkeit und Vollständigkeit sichergestellt werden (Datenqualität).

Betriebswirtschaftliche Kennzahlen können bezüglich ihrer Bezeichnung, Abgrenzung, Herkunft und Verwendung mithilfe von Metadaten beschrieben werden. Die Metadaten der dispositiven Datenhaltung stellen damit einen „single point of truth“ im Unternehmenskontext dar (Begriffsverständnis).

Architekturvarianten

Die zumeist komplexen, individuellen BI-Ansätze von Unternehmen bestehen aus einer Vielzahl spezieller Softwarekomponenten. Grundsätzlich kann zwischen End-to-End- und Best-of-Breed-Ansätzen unterschieden werden.

Bei End-to-End-Ansätzen bieten die Softwarehersteller aufeinander abgestimmte Werkzeuge an. Die Tools unterstützen sämtliche Entwicklungs- und Betriebsprozesse, von der ETL-Konzeption bis hin zur Berichterstellung bzw. der Portalintegration von Berichten. Im Gegensatz dazu bieten Softwarehersteller von Best-of-Breed-Lösungen spezielle Werkzeuge, mit denen leistungsfähige Komponenten eines unternehmensspezifischen BI-Konzepts entwickelt werden. Die konsistente Metadatenverwaltung aller Komponenten eines integrierten BI-Konzepts ist ein komplexes Vorhaben.

Im Folgenden werden die drei Architekturvarianten beschrieben:

Zentrales Metadatenmanagement

Dabei werden die Metadaten aller Komponenten und Berechtigungsstrukturen in einer Datenbank vorgehalten.

1. **Zentrales Metadatenmanagement:** Hier wird eine zentrale Datenbank verwendet, welche für die zentrale Metadatenverwaltung eingesetzt wird. In der Datenbank werden Metadaten aller Komponenten und Berechtigungsstrukturen gespeichert. Die sogenannte monolithische Lösung wird insbesondere bei End-to-End-Ansätzen angewandt.

In der Praxis existieren kaum zentrale Metadaten-Lösungen.

Vorteile	Nachteile
<ul style="list-style-type: none"> • redundanzfreie und konsistente Metadatenverwaltung 	<ul style="list-style-type: none"> • Abhängigkeit von der zentralen Datenhaltungskomponente
<ul style="list-style-type: none"> • globaler Zugriff auf alle Metadaten 	<ul style="list-style-type: none"> • komplexe, zentrale Wartung komponentenspezifischer Metadaten
<ul style="list-style-type: none"> • Verzicht auf Austauschmechanismen (Metadaten) 	<ul style="list-style-type: none"> • schlechte Performance großer, zentraler Lösungen

Dezentrales Metadatenmanagement

Dabei verfügen alle Komponenten über eigene Metadaten-Repository.

2. **Dezentrales Metadatenmanagement** ist gegenüber dem vorigen Konzept der gegen teilige Ansatz. Im Prinzip verfügen alle Komponenten über eine eigene Metadaten-Repository und kommunizieren miteinander, um Metadaten auszutauschen. BI-Konzepte werden zumeist mit der dezentralen Metadatenverwaltung implementiert.

Vorteile	Nachteile
<ul style="list-style-type: none"> • Autonomie der Anwendungen 	<ul style="list-style-type: none"> • viele Schnittstellen
<ul style="list-style-type: none"> • schneller, lokaler Zugriff 	<ul style="list-style-type: none"> • redundante Datenhaltung

3. **Föderiertes Metadatenmanagement** stellt eine Kombination der zuvor vorgestellten Ansätze dar. Die Komponenten verwalten jeweils die eigenen Metadaten. Darüber hinaus existiert ein zentrales Repository, in welchem gemeinsam genutzte Metadaten verwaltet werden. Mithilfe einer standardisierten Schnittstelle erfolgt der Austausch von Metadaten zwischen den einzelnen Komponenten und dem zentralen Repository.

Vorteile:

- einheitliche Präsentation gemeinsam benutzter Metadaten,
- Autonomie der lokalen Repository,
- reduzierte Anzahl an Schnittstellen zwischen Repositories,
- kontrollierte Redundanz.

Föderiertes Metadatenmanagement

Bei diesem existiert über die einzelnen Metadaten hinaus ein zentrales Repository mit gemeinsam genutzten Metadaten.

Mithilfe von Standardschnittstellen kann der Austausch von Metadaten zwischen BI-Werkzeugen und Metadaten-Repository ermöglicht werden.

Berechtigungsstrukturen

Die Regelung der Zugriffsberechtigungen erfolgt zumeist innerhalb der einzelnen Systeme. Im Gegensatz dazu erlauben BI-Konzepte mit integrierter Datenhaltung eine zentrale Berechtigungsverwaltung. Hierdurch entfällt die getrennte Berechtigungsverwaltung von verschiedenen Systemen, wie z. B. ETL, C-DWH oder Analysesystemen.

In der Praxis etablieren sich zunehmend sogenannte rollenbasierte Zugriffskontrollen, insbesondere aufgrund deren hoher Flexibilität. Benutzer bzw. Benutzergruppen erhalten Mitgliedschaften in entsprechenden Rollen. In Rollen werden Rechte zusammengefasst, die zur Erfüllung definierter Aufgaben und Funktionen nach dem Need-to-know-Prinzip notwendig sind. Durch die Zuweisung von Datensichten zu den Rollen wird sichergestellt, dass Anwender lediglich auf bestimmte Datenfelder, z. B. Umsätze der eigenen Filiale, zugreifen dürfen.

Administrationsschnittstellen

Mithilfe von Administrationsschnittstellen können technische und betriebswirtschaftliche Spezialisten sämtliche Bereiche der dispositiven Datenhaltung pflegen. Die entsprechenden Personen können beispielsweise

- Transformationsregeln,
- dispositive Daten und
- rollenbasierte Zugriffsberechtigungen

Technische Administrationsschnittstelle

Diese dient zur Modifizierung von Daten und erster Transformationsschicht (Filterung).

Fachliche Administrationsschnittstelle

Diese dient zur Modifizierung der Transformationsschichten (Harmonisierung, Aggregation, Anreicherung) und Berechtigungsstrukturen.

generieren, modifizieren und löschen. Grundsätzlich kann nach der technischen sowie der fachlichen Administrationsschnittstelle unterschieden werden. Mithilfe der **technischen Administrationsschnittstelle** können Daten und die erste Transformationsschicht (Filterung) modifiziert werden. Hierzu gehört neben der Datenmanipulation auch die Bearbeitung sämtlicher Strukturen zur Extraktion und Bereinigung von Daten.

Die **fachliche Administrationsschnittstelle** dient dagegen zur Pflege der oberen drei Transformationsschichten (Harmonisierung, Aggregation, Anreicherung) und der Verwaltung des Berechtigungssystems. Beispielsweise verwenden die betriebswirtschaftlichen Spezialisten die fachliche Administrationsschnittstelle, um syntaktische und semantische Harmonisierungsprozesse, Hierarchieebäume, Aggregationen oder betriebswirtschaftliche Kennzahlen intuitiv zu bearbeiten.

Im Rahmen der Berechtigungsverwaltung erfolgt mithilfe der fachlichen Administrationsschnittstelle die intuitive Zuordnung von Rollen und Mitarbeitern bzw. Mitarbeitergruppen.



ZUSAMMENFASSUNG

Der ETL-Prozess übernimmt die Bereinigung und Transformation von operativen Daten, welche anschließend im Data Warehouse für weitergehende Analysen bereitgestellt werden. Die Aufbereitung erfolgt in den vier Teilprozessen: Filterung, Harmonisierung, Aggregation und Anreicherung.

Das DWH umfasst im engeren Sinne die Datenhaltung. Einzelne Komponenten des DWH sind demnach die Staging Area, Basisdatenbank, Data Mart, ODS und Metadaten.

Die Staging Area ist ein Arbeitsbereich, in welchem Daten temporär zwischengespeichert werden, um beispielsweise nachgelagerte Systeme zu entlasten.

Die Basisdatenbank ist ein Datenspeicher, der zwischen der Staging Area und der Auswertungsdatenbank angesiedelt ist. Das C-DWH stellt unternehmensweite, integrierte Daten von nachgelagerten Systemen zur Verfügung.

Data Marts sind Ausschnitte eines C-DWH, die Daten für spezifische Anwendergruppen (z. B. Abteilungen) bereitstellen.

Das ODS beinhaltet aktuelle Daten auf Transaktionsbasis, die verschiedenen operativen Quellsystemen entstammen. Die klassischen Merkmale eines ODS sind Themenorientierung, Integration, Zeitpunktbezug, Volatilität und hoher Detaillierungsgrad. Metadaten enthalten Informationen über die im Data Warehouse gespeicherten Daten und deren Verar-

beitung und bieten Unterstützung bei Aufbau, Verwaltung und Betrieb von DWH-Systemen. Hierbei kann nach aktiven und passiven bzw. technischen und fachlichen Metadaten unterschieden werden.

LEKTION 4

MODELLIERUNG MULTIDIMENSIONALER DATENRÄUME

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- welche grundlegenden Modellierungstechniken existieren.
- welche Analysemöglichkeiten durch OLAP-Würfel geboten werden.
- wie multidimensionale Modelle physisch gespeichert werden.
- welche Historisierungsmöglichkeiten von Dimensionen existieren.

4. MODELLIERUNG MULTIDIMENSIONALER DATENRÄUME

Einführung

Ein Data-Warehouse-System unterstützt unternehmensweit Entscheidungsträger bei ihrer Arbeit. Entsprechend der Aufgabengebiete haben diese meist unterschiedliche Interessen an den einzelnen Daten. Ein Mitarbeiter aus dem Controlling interessiert sich in der Regel für betriebswirtschaftliche Kennzahlen, ein Arzt hingegen für den Erfolg einer bestimmten Therapie. Aus diesem Grund besteht die Notwendigkeit, flexible Sichtweisen auf Daten zu gewähren.

Für Data-Warehouse-Anwendungen ist ein sogenanntes multidimensionales Datenmodell flexibler als ein relationales Datenmodell. In multidimensionalen Modellen werden Unternehmensdaten in einem mehrdimensionalen Datenraum angeordnet.

4.1 Datenmodellierung

Relationale und multidimensionale Modelle

Neben dem relationalen Datenmodell spielen im BI-Kontext vor allem multidimensionale Datenräume eine bedeutende Rolle. Eine performanceorientierte Modellierung multidimensionaler Räume erlauben vor allem Star- und Snowflake-Schemata.

Grundsätzlich können Datenmodelle semantisch, logisch oder physisch ausgerichtet sein. Physische Datenmodelle sind techniknah orientiert und spezifizieren, wie Daten physisch abgespeichert werden. Für Anwender sind logische und semantische Sichten interessanter. Logische Datenmodelle beschreiben alle Daten auf logischer Ebene, unabhängig von ihrer Speicherung. Semantische Modelle besitzen hingegen die größte Nähe zur Realität. Sie bilden die Daten in einer völlig technikneutralen Ebene ab.

Eines der bekanntesten semantischen Datenmodelle ist das Entity-Relationship-Modell (ERM) von Peter Chen. Es wurde in den siebziger Jahren entwickelt und im Laufe der Zeit modifiziert und erweitert. Das ERM dient zum einen in der konzeptionellen Phase der Anwendungsentwicklung der Verständigung zwischen Anwendern und Entwicklern, zum anderen in der Implementierungsphase als Grundlage für das Design der Datenbank. Mit ERM lassen sich operative Datenstrukturen gut modellieren. Auch mehrdimensionale Datenstrukturen, die weiter unten vorgestellt werden, können damit abgebildet werden.

Redundanzen und Normalformen

In der Praxis werden relationale Datenbanken nicht immer aus einem Modell (z. B. ERM) erstellt. Der Nachteil ist, dass redundante Informationen gespeichert werden können. Mit Redundanz ist das mehrfache Speichern identischer Attributwerte ein und derselben Objektausprägung gemeint. Dies ist beispielsweise der Fall, wenn in einer Mitarbeiterdatenbank der Name des Mitarbeiters zusammen mit der Abteilung und der Abteilungsnummer gespeichert wird. Redundanzen gefährden die Konsistenz der Datenbasis und können zu sogenannten Anomalien führen. Ändert sich beispielsweise die Abteilungsnummer, müssen mehrere Tupel gleichzeitig geändert werden. Dies ist nicht nur aufwendig, sondern birgt die Gefahr der Inkonsistenz (Update-Anomalie). Verlassen alle Mitarbeiter eine Abteilung, geht auch die Information verloren, welche Abteilungsnummer der Abteilung zugeordnet ist (Deletion-Anomalie).

Um Anomalien zu verhindern, gibt es eine Vielzahl von Vorschriften und Prinzipien. Ein weit verbreitetes Verfahren, Redundanzen und Inkonsistenzen zu vermeiden, ist die Normalisierung.

Primärschlüssel und erste Normalform

Normalisierte Daten sind Daten, die frei von Redundanzen und Inkonsistenzen sind, und dadurch effizienter verwaltet werden können. Eine starke Normalisierung kann allerdings auch zu Lasten der Performance gehen. Die Normalformenlehre geht auf Edgar F. Codd, Erfinder der relationalen Datenbank, zurück. Er entwickelte mathematische Regeln, die relationale Datenbanken in redundanzfreie oder zumindest redundanzärmere Datenstrukturen überführen. Im Kern besteht dieser Prozess aus drei Normalisierungsschritten.

Die Definition der ersten Normalform (1NF) lautet: Eine Tabellenzeile darf nur einen Attributwert enthalten (technisch: Der Attributwert muss atomar sein.). Kommen in einer Tabelle Wiederholungsgruppen vor, so muss für jeden Wert der Wiederholungsgruppe eine eigene Tabellenzeile geschaffen werden.

Zweite und Dritte Normalform

Die Definition der zweiten Normalform lautet: Eine Relation ist in der zweiten Normalform, wenn sie in der ersten Normalform ist und alle Nicht-Schlüsselattribute funktional vom gesamten Schlüssel abhängen. Damit muss ausgeschlossen sein, dass bereits Schlüsselteile bestimmte Attribute der Relation identifizieren können. Somit ist aus jenen Attributen, die nur von einem Teil des zusammengesetzten Primärschlüssels abhängen, eine neue Tabelle zu erzeugen.

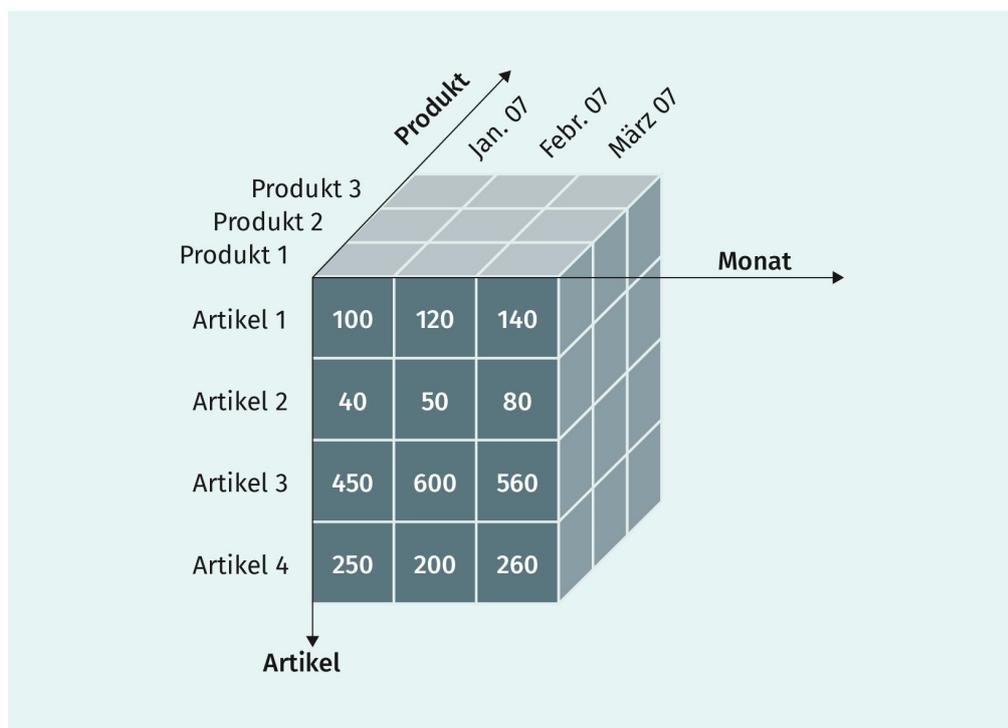
Die Definition der dritten Normalform lautet: Eine Relation der dritten Normalform liegt vor, wenn die zweite Normalform besteht und zusätzlich keine funktionalen Abhängigkeiten zwischen Nicht-Schlüsselattributen existieren.

4.2 OLAP-Würfel

Im Mittelpunkt dieses Modells stehen meist betriebswirtschaftliche Kennzahlen als Träger der quantitativen Informationen, die durch die dazugehörige Menge an Dimensionen eindeutig in ihrem Kontext beschrieben werden. Jede Dimension erklärt sich durch eine Menge von Attributen. Die Dimension Zeit wird beispielsweise durch die Attribute Jahr, Monat, Quartal, Woche und Tag beschrieben. Die Attribute können innerhalb einer Dimension miteinander in Beziehung stehen und eine Beziehungshierarchie bilden. Die Hierarchie ermöglicht zum einen die Aggregation von Daten und zum anderen die Navigation durch die Daten.

Im allgemeinen Sprachgebrauch wird der so entstehende multidimensionale Datenraum auch als OLAP-Würfel bzw. Cube bezeichnet, der nachfolgend dargestellt wird:

Abbildung 16: Cube und Dimensionen



Quelle: Gluchowski/Gabriel/Dittmar 2008, S. 156.

Ein Würfel besteht immer aus Dimensionen (Kanten des Würfels) und Kennzahlen bzw. Fakten (Werte an den Koordinaten innerhalb des Würfels). Die **Fakten** sind die betriebsbedingten Größen wie etwa Umsatz, Kosten oder Stückzahlen. Demgegenüber besitzen **Dimensionen** einen beschreibenden Charakter für die Kennzahlen, z. B. die Zeit, der Kunde und das Produkt für einen Umsatz.

Fakten

Diese sind unternehmerische Kennzahlen, wie z. B. Umsatz, Kosten oder Stückzahlen.

Die Elemente einer Dimension lassen sich aufgrund funktionaler Abhängigkeiten oft in Hierarchien gruppieren. In der Zeitdimension können beispielsweise Monate zu Quartalen oder in der Produktdimension Produkte zu Produktgruppen zusammengefasst werden. Entlang der Dimensionshierarchie lassen sich die Kennzahlen konsolidieren (z. B. Summenbildung). Für die Navigation innerhalb des multidimensionalen Datenmodells stehen die Funktionen Roll-Up und Drill-Down, Slice und Dice, Drill-Across, Pivotierung und Rotation zur Verfügung, die im Folgenden erläutert werden.

Dimensionen

Diese beschreiben unternehmerische Kennzahlen, z. B. durch die Angabe von Zeit, Region oder Produkt.

Roll-Up und Drill-Down

Mithilfe der Funktionen Roll-Up und Drill-Down können Anwender entlang der Beziehungshierarchie navigieren.

Abbildung 17: Roll-Up & Drill-Down

	Produkt A	Produkt B	Produkt C	Produkt D
1. Quartal	140.000	100.000	200.000	120.000
	Drill-down 			
Januar	40.000	30.000	70.000	40.000
Februar	45.000	35.000	60.000	35.000
März	55.000	35.000	70.000	45.000
	 Roll-up			

Quelle: Kemper/Baars/Mehanna 2010, S. 103.

Beim **Drill-Down** wird der Detaillierungsgrad erhöht, das heißt, der Anwender hat die Möglichkeit, die Kennzahlen der einzelnen Produkte auch auf Monatsebene zu betrachten.

Der **Roll-Up** ist die inverse Funktion zum Drill-Down, das heißt, der Detaillierungsgrad wird verringert, sodass Daten in aggregierter Form (Quartalsebene) betrachtet werden.

Slice und Dice

Durch Slice und Dice werden individuelle Sichten auf das Datenmodell erzeugt. Im Folgenden wird der Slice-Operator dargestellt:

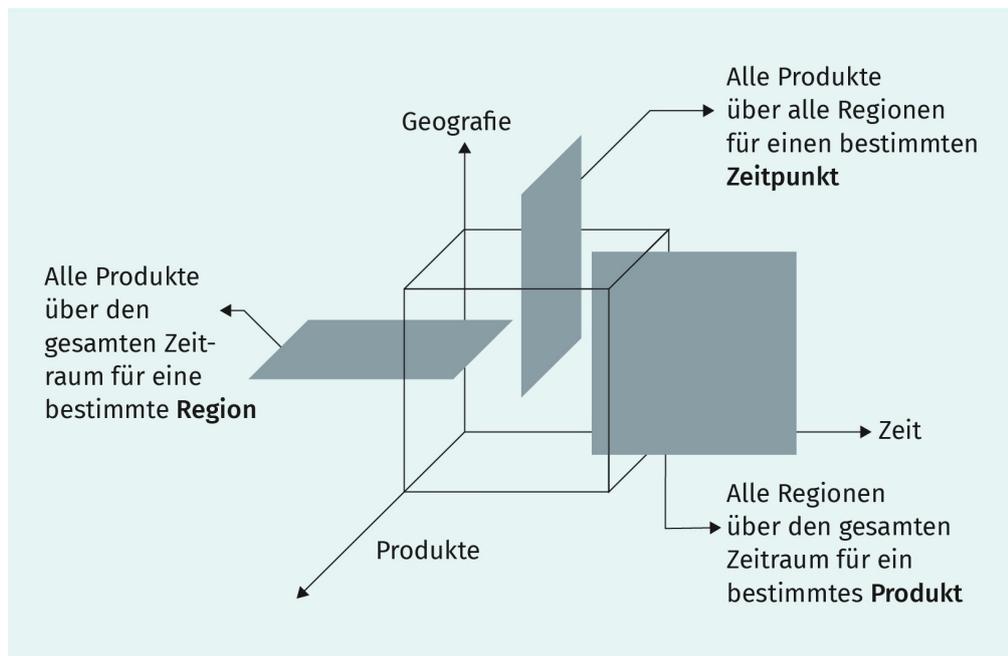
Drill-Down

Hierbei springt man in eine tiefere Detaillierungsebene eines Berichts.

Roll-Up

Hierbei springt man auf eine höhere Aggregatensebene.

Abbildung 18: Slice-Operator

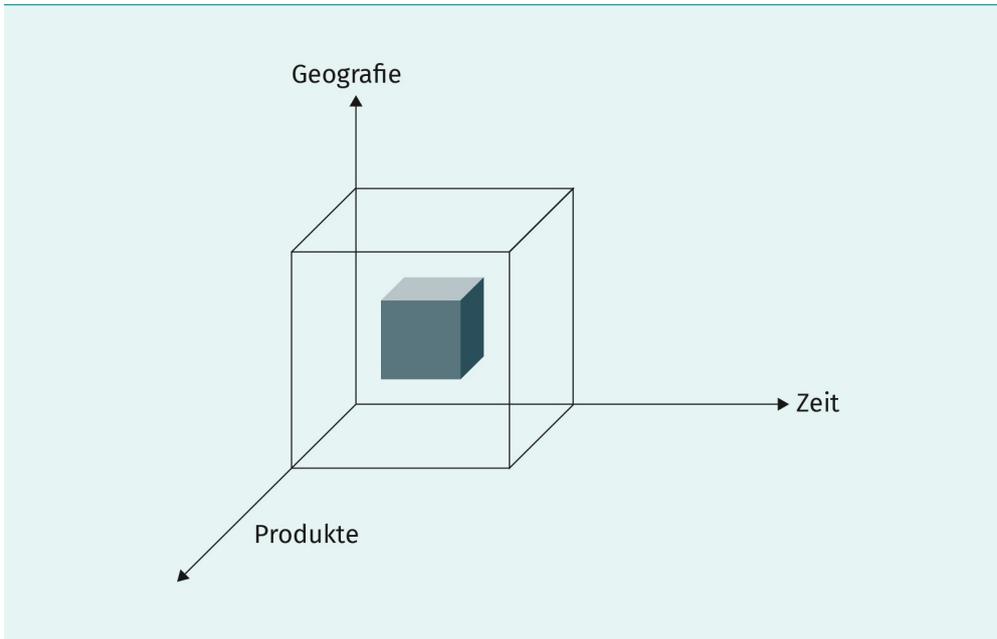


Quelle: Kemper/Baars/Mehanna 2010, S. 105.

Slice-Funktion
Diese Funktion ermöglicht die Betrachtung einer einzelnen Ebene.

Die **Slice-Funktion** schneidet einzelne Scheiben aus dem Datenwürfel heraus und ermöglicht damit die Betrachtung einer einzelnen Ebene des Würfels. Der Anwender kann sich mithilfe dieser Funktion bei seiner Analyse auf die Betrachtung des Umsatzes einer einzelnen Region beschränken, indem z. B. die Scheibe der Region „herausgeschnitten“ wird.

Abbildung 19: Dice-Operator



Quelle: Kemper/Baars/Mehanna 2010, S. 105.

Die **Dice-Funktion** schneidet einen Teilwürfel aus dem Gesamtwürfel heraus. Der Operator bietet dem Anwender die Möglichkeit, die Kennzahlen bei einer konkreten Kombination von Dimensionselementen zu betrachten.

Dice-Funktion
Diese Funktion ermöglicht die Betrachtung einzelner Teilwürfel.

Beim **Drill-Across** wird zwischen verschiedenen Würfeln gewechselt.

Drill-Across
Dies ermöglicht einen Wechsel in andere Würfel.

Pivotierung

Die **Pivotierung** ermöglicht das virtuelle Drehen bzw. Rotieren von Datenwürfeln, um Berichtsdaten aus verschiedenen Perspektiven zu betrachten. Die Reihenfolge der dargestellten Dimensionen wird während der Ausführung der Pivotierung vertauscht.

Pivotierung
Hierbei wird der Würfel gedreht, um Daten aus verschiedenen Perspektiven zu betrachten.

4.3 Physische Speicherung

Für die Speicherung von Daten in Strukturen eines DWH existieren verschiedene Konzepte. Die Verwendung des multidimensionalen Datenmodells verlangt nicht zwangsläufig eine multidimensionale Verwaltung der Daten. Das alternativ zum multidimensionalen einsetzbare relationale Speichermodell findet häufiger Verwendung. Als weitere Möglichkeit bietet sich die hybride Speicherung an, welche eine Mischung aus multidimensionaler und relationaler Speicherung darstellt.

Relationale Speicherung (ROLAP)

ROLAP
Hierbei werden multidimensionale Datenmodelle in relationale Speicherkonzepte umgesetzt.

Am weitesten verbreitet ist die Verwendung des **relationalen Speichermodells (ROLAP)**, bei welchem die Speicherung von Daten aus dem multidimensionalen Datenmodell in zweidimensionalen Tabellen erfolgt. Für das Gesamtsystem muss jedoch die multidimensionale Schnittstelle erhalten bleiben. Daher erfolgt eine Abbildung der Daten des multidimensionalen Datenmodells auf das Datenmodell des relationalen Datenbanksystems (Bauer/Günzel 2013, S. 241f.).

Multidimensionale Speicherung (MOLAP)

MOLAP
Hierbei erfolgt die physische Speicherung in einem multidimensionalen Datenbankmanagementsystem.

Bei der **multidimensionalen Speicherung (MOLAP)** erfolgt die physische Speicherung in einem multidimensionalen Datenbankmanagementsystem. Die Speicherung wird durch die direkte Übertragung der Modellelemente in physische Objekte ermöglicht. Hierbei erfolgt die Speicherung der Datenelemente in Arrays.

Durch die multidimensionale Speicherung kann eine sehr hohe Anfragegeschwindigkeit erzielt werden. Allerdings gibt es beim multidimensionalen Datenbankmanagementsystem das Problem, dass ein solches System keine sehr großen Datenbestände verwalten kann. Daher sollte ein umfangreiches Data Warehouse mit niedrigen Granularitäten eher mit relationaler Speicherung, ein kleines, bereits aus aggregierten Werten gebildetes Data Warehouse (z. B. Data Mart) dagegen mithilfe multidimensionaler Speicherung realisiert werden.

Hybride Speicherung (HOLAP)

HOLAP
Hierbei werden die Stärken des relationalen und multidimensionalen Konzepts verknüpft.

Das **hybride Speichermodell (HOLAP)** versucht, die Stärken des relationalen und des multidimensionalen Konzepts zu verbinden. Für die Speicherung wird sowohl eine relationale als auch eine multidimensionale Datenbank verwendet. In der relationalen Datenbank werden die Daten gespeichert, die detailliert und in großen Mengen vorliegen. Die aggregierten Daten werden dagegen in einer multidimensionalen Datenbank abgelegt. Der Datenzugriff erfolgt mithilfe eines multidimensionalen Abfragewerkzeugs.

Durch die Verwendung beider Technologien ist es möglich, die jeweiligen Vorteile aususchöpfen und die bestehenden Nachteile zu überwinden. Jedoch sind umfassende Kenntnisse auf beiden Gebieten sowie zusätzlicher Implementierungsaufwand notwendig.

4.4 Star- und Snowflake-Schema

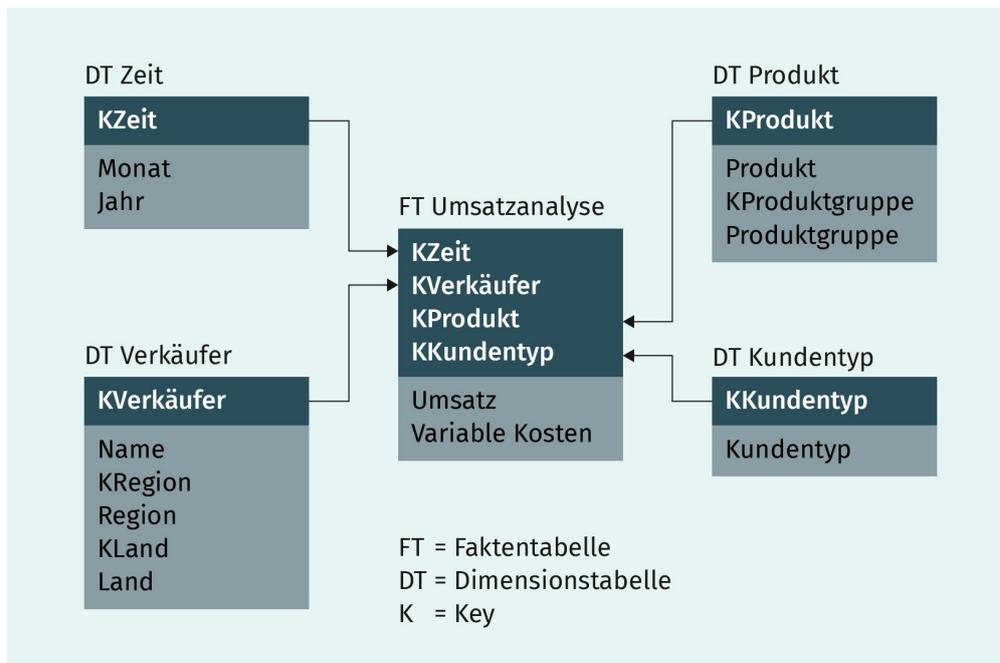
Bei der physischen Umsetzung eines multidimensionalen Datenmodells kann nach Star-Schema und Snowflake-Schema unterschieden werden, welche im Folgenden erläutert werden.

Star-Schema

Das **Star-Schema** besteht aus einer Faktentabelle, in der die Kenngrößen eines Datenwürfels verwaltet werden.

Star-Schema
Hierbei werden Dimensionen sternförmig um die Faktentabelle angeordnet.

Abbildung 20: Star-Schema



Quelle: Kemper/Baars/Mehanna 2010, S. 68.

In der Abbildung ist zu erkennen, dass die Dimensionstabellen sternförmig um die Faktentabelle angeordnet sind. Für jede Dimension existiert eine eigene Tabelle. Beziehungen bestehen lediglich zwischen Fakten- und Dimensionstabelle und nicht zwischen den Dimensionstabellen.

Die Verknüpfung zwischen der Fakten- und Dimensionstabelle wird erreicht, indem der Primärschlüssel der Faktentabelle als Fremdschlüssel in die Primärschlüssel der Dimensionstabellen aufgenommen wird.

Snowflake-Schema

Eine andere Möglichkeit ist die Abbildung mithilfe des **Snowflake-Schemas**. Im Gegensatz zum Star-Schema ist das Snowflake-Schema bezüglich der funktionalen Abhängigkeiten normalisiert, was zu einer Vielzahl von Tabellen führt, die bei einer Abfrage miteinander verbunden werden müssen (SQL-Join). Hierdurch ergeben sich bei Abfragen Performance-defizite. Aus diesem Grund wird häufiger das Star-Schema verwendet, welches sich aus dem Snowflake-Schema durch Denormalisierung der zu jeder Dimension gehörenden Tabellen ergibt (Bauer/Günzel 2013, S. 243).

Snowflake-Schema
Dieses Schema ist bezüglich der funktionalen Abhängigkeiten normalisiert.

4.5 Historisierung

In einem multidimensionalen Datenmodell wird zwischen Fakten- und Dimensionstabellen unterschieden. Die Historisierung von Fakten geschieht in der Regel über die Fremdschlüsselbeziehungen der Daten zur Zeitdimension. Die Historisierung von Dimensionen ist dagegen eine regelmäßig auftretende Herausforderung im DWH-Kontext. Aufgrund dessen wird die Problematik nachfolgend detailliert beschrieben.

Slowly Changing Dimensions (SCD)
Diese beschreiben Konzepte zur Historisierung von Dimensionen.

Im Data-Warehouse-Kontext müssen neben der Historisierung von Fakten auch Dimensionsänderungen berücksichtigt werden. Zur Historisierung von Dimensionen führte Kimball den Begriff „**Slowly Changing Dimensions (SCD)**“ ein. Änderungen von Dimensionen treten beispielsweise bei Umbenennung von Produkten auf.

Slowly Changing Dimensions können nach Kimball auf drei Arten verarbeitet werden. Folgendes Beispiel verdeutlicht die drei Typen in ihrer Behandlung. Gegeben ist eine Produkttabelle, welche eine Dimension im Data Warehouse beschreibt. Die Schlüsselattribute sind in den Darstellungen jeweils unterstrichen.

Tabelle 5: Dimensionstabelle Produkte

<u>Produktnummer</u>	<u>Produktname</u>	<u>Produktgruppe</u>
1	Telefon1	Schnurgebundene Tel.
2	Telefon2	Schnurlose Tel.
3	Telefon3	Schnurlose Tel.

Quelle: erstellt im Auftrag der IU, 2016.

Aus den entsprechenden Datenquellen wird am 02.06.2015 der folgende Datensatz extrahiert, bei dem sich die Produktgruppe eines Produkts geändert hat.

Tabelle 6: Extraktionsdaten Produkte

<u>Produktnummer</u>	<u>Produktname</u>	<u>Produktgruppe</u>
3	Telefon3	Schnurlose ISDN Tel.

Quelle: erstellt im Auftrag der IU, 2016.

Die Produktgruppe wurde nach „Schnurlose ISDN Tel.“ modifiziert.

SCD Typ 1

Bei **SCD Typ 1** wird der entsprechende Datensatz überschrieben. Dadurch gehen Informationen verloren. Bei der Analyse über einen Zeitraum ist die Änderung der Dimension nicht mehr nachvollziehbar. Bei einer Dimensionshistorisierung nach SCD Typ 1 wird der bisherige Datensatz „Telefon3“ durch den neuen Datensatz überschrieben. Hierbei geht die Information über die bisherige Produktgruppe verloren.

SCD Typ 1
Hierbei wird der Datensatz überschrieben.

Tabelle 7: Historisierung SCD Typ 1

Produktnummer	Produktname	Produktgruppe
1	Telefon1	Schnurgebundene Tel.
2	Telefon2	Schnurlose Tel.
3	Telefon3	Schnurlose ISDN Tel.

Quelle: erstellt im Auftrag der IU, 2016.

Im dargestellten SCD Typ 1 findet dementsprechend keine Historisierung statt. Der entsprechende Datensatz wird überschrieben.

SCD Typ 2

Für eine Dimensionshistorisierung nach **SCD Typ 2** werden zwei weitere Attribute benötigt, die das Gültigkeitsintervall der Dimension definieren. Bei Änderung einer Dimension wird die obere Grenze der Gültigkeit der bisherigen Dimension auf das Datum der Änderung gesetzt. Die neue Dimension wird als neuer Datensatz mit einer Gültigkeit vom Datum der Änderung bis unendlich eingefügt.

SCD Typ 2
Hierbei wird jedem Datensatz ein Gültigkeitsintervall hinzugefügt.

Bei SCD Typ 2 wird jedem Datensatz ein Gültigkeitsintervall hinzugefügt, welches durch die Attribute GültigVon und GültigBis definiert wird. Mit Eintreffen des neuen Datensatzes wird der bisherige Eintrag für Telefon3 historisiert, indem das Attribut GültigBis auf den Tag vor der Änderung gesetzt wird. Der neue Datensatz für Telefon3 wird mit dem Gültigkeitsintervall vom Datum der Änderung bis unendlich in die Tabelle eingefügt. Die ursprüngliche Produktgruppe bleibt durch die Historisierung nachvollziehbar. Ebenso das Datum der Änderung.

Damit der aktuelle Datensatz vom historisierten unterschieden werden kann, muss das Gültigkeitsintervall zusammen mit der Produktnummer in einen neuen, zusammengesetzten Primärschlüssel einbezogen werden. Alternativ hierzu kann die Unterscheidung auch durch Austausch des Schlüssels durch einen Surrogate Key geschehen. Der neue Schlüssel muss anschließend in die Faktentabellen propagiert werden, damit neue Fakten die korrekte Dimension referenzieren (Kimball/Caserta 2004, S. 185). Die Abbildung vom bisherigen Produkt auf das geänderte muss dabei über die Metadaten des ETL-Prozesses gesteuert werden.

Tabelle 8: Historisierung SCD Typ 2

Produktnummer	Produktname	Produktgruppe	Gültig von	Gültig bis
1	Telefon1	Schnurgebundene Tel.	01.01.1999	31.12.9999
2	Telefon2	Schnurlose Tel.	01.01.1999	31.12.9999
3	Telefon3	Schnurlose Tel.	01.01.1999	01.06.2015
4	Telefon3	Schnurlose ISDN Tel.	02.06.2015	31.12.9999

Quelle: erstellt im Auftrag der IU, 2016.

In der dargestellten Tabelle wurde eine Historisierung nach SCD Typ2 mittels Gültigkeitsintervall und Surrogate Key durchgeführt.

SCD Typ 3

SCD Typ 3
Hierbei wird der Datensatz um ein Attribut mit der neuen Bezeichnung erweitert.

Der **SCD Typ 3** erweitert den bestehenden Datensatz um ein Attribut, das die neue Bezeichnung enthält. Dadurch wird nicht vollständig historisiert, sondern nur der Ursprungszustand und der aktuelle Zustand gespeichert.

Bei SCD Typ 3 wird die Produkttabelle um das zusätzliche Attribut „NeueProdGruppe“ erweitert, in dem die neue Produktgruppe gespeichert wird.

Tabelle 9: Historisierung SCD Typ 3

Produktnummer	Produktname	Produktgruppe	Neue Produktgruppe
1	Telefon1	Schnurgebundene Tel.	Schnurgebundene Tel.
2	Telefon2	Schnurlose Tel.	Schnurlose Tel.
3	Telefon3	Schnurlose Tel.	Schnurlose ISDN Tel.

Quelle: erstellt im Auftrag der IU, 2016.

In der dargestellten Tabelle wurde eine Historisierung (SCD Typ 3) über das zusätzliche Attribut „NeueProdGruppe“ durchgeführt. Durch das zusätzliche Attribut können alte und neue Produktgruppe nachvollzogen werden. Mehrfache Änderungen können nicht gespeichert werden. Das Datum der Änderung wird nicht gespeichert.

Vergleich der SCD Typen 1 bis 3

SCD Typ 1 ist am einfachsten zu implementieren und hält das Datenaufkommen gering. Allerdings findet keine Historisierung statt. Durch das Überschreiben gehen eventuell wichtige Informationen verloren. In Anbetracht des analytischen Charakters der meisten Anwendungen, die auf dem Data Warehouse basieren, wird dieser Informationsverlust in den meisten Fällen inakzeptabel sein.

SCD Typ 2 bietet eine volle Historisierung von Dimensionsänderungen. Die Implementierung ist aufwendig. Ebenso ist die notwendige Erkennung der Änderung schwierig. Dafür gehen bei SCD Typ 2 keine Informationen verloren.

SCD Typ 3 historisiert nur den ursprünglichen und aktuellen Wert eines Attributs. Informationen gehen verloren, wenn ein Attribut mehrfach verändert wird. Der Typ ist ebenfalls aufwendig zu implementieren. Dadurch ist SCD Typ 3 lediglich für sehr spezielle Fälle anzuwenden. Ein Beispiel für den Typ ist die Umstellung der Postleitzahlen. Da davon ausgegangen werden kann, dass eine solche Änderung selten auftritt und die Änderung nicht voll historisiert nachvollziehbar sein muss, kann hier der Informationsverlust in Kauf genommen werden.

Die Modellierung multidimensionaler Daten erfolgt innerhalb des relationalen Datenmodells mit Star- und Snowflake-Modellen. Die Schemata erlauben eine performanceoptimierte Modellierung multidimensionaler Datenräume. Dabei werden zum Teil die strikten theoretischen Anforderungen an relationale Systeme (wissentlich) verletzt.



ZUSAMMENFASSUNG

Anwender haben zumeist unterschiedliche Interessen in Bezug auf einzelne Daten. Daher sind flexible Sichtweisen notwendig. Mithilfe von multidimensionalen Datenmodellen können Daten in einem multidimensionalen Datenraum angeordnet werden.

Im Mittelpunkt des Modells stehen zumeist betriebswirtschaftliche Kennzahlen als Träger der quantitativen Informationen, die durch die dazugehörige Menge an Dimensionen eindeutig in ihrem Kontext beschrieben werden. Für die physische Speicherung von multidimensionalen Modellen existieren verschiedene Konzepte (ROLAP, MOLAP und HOLAP).

Bei der physischen Umsetzung eines multidimensionalen Datenmodells kann nach dem Star- oder Snowflake-Schema unterschieden werden.

Die Historisierung von Dimensionen ist eine regelmäßig auftretende Herausforderung im DWH-Kontext. Unter dem Begriff „Slowly Changing Dimensions (SCD)“ werden Methoden zusammengefasst, um Änderungen an Dimensionstabellen zu erfassen und gegebenenfalls historisch zu dokumentieren. Im Wesentlichen werden drei Verfahren bzw. Typen unterschieden.

LEKTION 5

ANALYSESYSTEME

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- wie Analysesysteme bzw. Frontends klassifiziert werden können.
- was unter freier Datenrecherche und Ad-hoc-Analysesystemen zu verstehen ist.
- welche Berichtssysteme angewandt werden.
- welche modellgestützten und konzeptorientierten Analysesysteme existieren.

5. ANALYSESYSTEME

Einführung

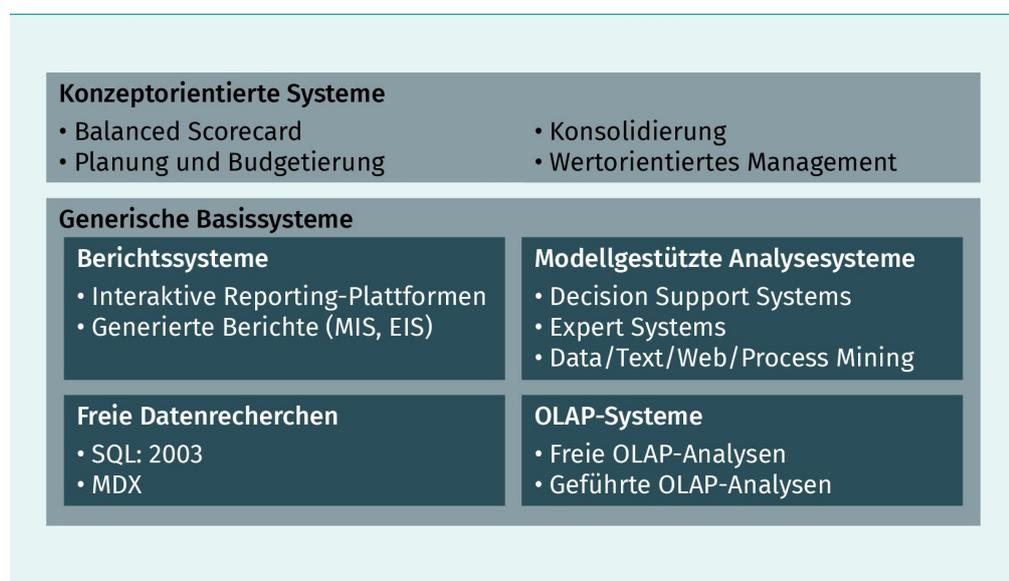
Auf der Basis des Datenbestandes einer implementierten Data-Warehouse-Architektur können mit speziellen IT-Systemen Analysen zur Informationsgewinnung durchgeführt werden. Ein Datenzugriff auf ein Data Warehouse oder der zugrunde liegenden Data-Marts kann auf verschiedene Weisen und mit unterschiedlichsten Anwendungen erfolgen und ist häufig abhängig von den IT-Kenntnissen des jeweiligen Anwenders.

Um den diversen Anforderungen der Anwender gerecht zu werden, wurde eine Vielzahl an Verfahren und Tools zur Datenanalyse entwickelt, die sehr unterschiedliche Eigenschaften besitzen, sodass sich eine eindeutige Klassifizierung derselben als schwierig erweist. Eine mögliche Klassifizierung unterscheidet die

- freie Datenrecherche,
- Ad-hoc-Analysesysteme (z. B. OLAP),
- Berichtssysteme,
- modellgestützten Analysesysteme,
- konzeptorientierten Systeme,

wobei die ersten vier Klassen auch als generische Systeme bezeichnet werden. Abbildung 21 veranschaulicht den Sachverhalt und zeigt zusätzlich einige der gängigen Vertreter dieser fünf Verfahren.

Abbildung 21: Analysesysteme für das Management



Quelle: Kemper/Baars/Mehanna 2010, S. 90.

5.1 Freie Datenrecherche und OLAP

Freie Datenrecherche

Da es sich bei einem Data Warehouse bzw. bei den Data-Marts um Datenbankobjekte handelt, kann eine entsprechende Datenabfrage auch mit einer dem Datenbankmanagementsystem (DBMS) zugrunde liegenden Datenmanipulationssprache (DML), sehr häufig mit Structured Query Language (SQL), durchgeführt werden. Eine solche Anfrage wird als freie Datenrecherche bezeichnet, da sie im Allgemeinen sehr flexibel, insbesondere aber unabhängig der zugrunde liegenden Modelle, formuliert werden kann (vgl. Kemper/Baars/Mehanna 2010). Solche SQL-Anfragen können beispielsweise direkt in Programmierumgebungen eingebunden und dort ausgeführt werden.

Eine zweite Variante der freien Datenrecherche ist die von Microsoft entwickelte Sprache Multidimensional Expressions (MDX), die speziell für Data-Marts entwickelt wurde und im Kontext eines Data Warehouses als eine Weiterführung von SQL verstanden werden kann, da MDX-Anfragen statt der Attribute von Tabellen, wie etwa beim SQL, die Dimensionen eines Würfels (Cubes) abfragen.

Beide Varianten, sowohl SQL als auch MDX, richten sich an technikversierte Anwender, die unter anderem mit der Struktur bzw. dem zugrunde liegenden Datenmodell vertraut sein müssen und sind daher für eine breite Anwendergruppe ungeeignet.

OLAP – Online Analytical Processing

Unter dem Begriff OLAP werden Methoden und Technologien zusammengefasst, welche die Ad-hoc-Analyse auf der Basis multidimensionaler Informationen erlauben. Für den Anwender eines solchen Verfahrens besteht die Möglichkeit, Daten unter verschiedensten Gesichtspunkten, sogenannten Dimensionen, zu analysieren.

Der OLAP-Begriff geht auf den britischen Mathematiker Edgar F. Codd zurück, der zwölf Regeln formuliert, welche die OLAP-Fähigkeit eines IT-Systems charakterisieren (Codd/Codd/Salley 1993). Diese Regeln lassen sich in vier Gruppen zusammenfassen und werden im Folgenden kurz aufgeführt.

Allgemeine Anforderungen

- Multidimensionale konzeptionelle Sichtweise: Es wird ein multidimensionales konzeptionelles Datenmodell zugrunde gelegt, das eine intuitive Analyse für den Anwender möglich macht.
- Transparenz für Datenbestände: Die Daten werden im Allgemeinen aus heterogenen Quellen zusammengeführt und verfügbar gemacht.
- Intuitive Datenbearbeitung: Zur Analyse werden sogenannte Slice-und-Dice-Funktionen oder auch Drill-Down-Operationen über Dimensionen zur Verfügung gestellt.
- Zugriffsmöglichkeit: Der Zugriff auf die Daten soll in einer einheitlichen und konsistenten Form ermöglicht werden.

Anforderung an die Berichterstellung

- Gleichbleibende Antwortzeiten bei der Berichterstellung: Unabhängig der zugrunde liegenden Datenmodelle, der Datenmenge oder auch der Anzahl der Dimensionen, sollen konsistente und schnelle Antwortzeiten ermöglicht werden.
- Flexible Berichterstellung: Sicherstellung der Vergleichbarkeit verschiedenster Dimensionen.

Dimensionsaspekte

- Generische Dimensionalität: Die Struktur der Dimensionen sollte einheitlich operational verfügbar sein.
- Uneingeschränkte kreuzdimensionale Operationen: Unterstützung von Berechnungen und Aggregationsfunktionen für Dimensionen.
- Unbegrenzte Anzahl von Dimensionen und Klassifikationsebenen: Es sollten zwischen 15 und 20 Dimensionen eines Datenmodells unterstützt werden.

Technologische Anforderungen

- Dynamische Behandlung dünn besetzter Matrizen: Dynamische Speicherstrukturanpassung.
- Mehrbenutzerunterstützung: Gemäß den Anforderungen einer Client-Server-Architektur.
- Client-Server-Architektur: Optimierte Lastenverteilung durch den Zugriff unterschiedlicher Front-Ends auf diverse Back-End-Server.

Da Codd/Codd/Salley diese Regeln zusammen mit verschiedenen Firmen veröffentlichten, wurde ihnen eine Herstellerabhängigkeit vorgeworfen. Auf Grund dessen, aber auch wegen der technischen Anforderungen, hat die Bedeutung der Regeln zur Bewertung von OLAP-Systemen an Bedeutung verloren.

Nigel Pendse und Richard Creeth entwickelten zwei Jahre später unter dem Begriff **Fast Analysis of Shared Multidimensional Information (FASMI)**, vgl. Pendse/Creeth 1995) fünf Regeln, die OLAP-Systeme auf der Basis von anwenderspezifischen Anforderungen wie folgt beschreiben:

FASMI

Diese Abkürzung steht für **F**ast **A**nalysis of **S**hared **M**ultidimensional **I**nformation.

- **Fast:** Abfragen sollen in einem Zeitfenster zwischen fünf und maximal zwanzig Sekunden beantwortet werden.
- **Analysis:** Ein OLAP-System soll jegliche benötigte Logik bewältigen können. Dabei soll die Definition einer komplexeren Analyseabfrage durch den Anwender mit wenig Programmieraufwand zu realisieren sein.
- **Shared:** Ein OLAP-System soll einen Mehrbenutzerbetrieb ermöglichen, was die Verfügbarkeit geeigneter Zugriffsschutzmechanismen impliziert.
- **Multidimensional:** Als Hauptkriterium wird eine mehrdimensionale Strukturierung der Daten mit voller Unterstützung der Dimensionshierarchien gefordert.
- **Information:** Bei der Analyse sollen einem Anwender alle benötigten Daten transparent zur Verfügung stehen. Analysen dürfen nicht durch Beschränkungen des OLAP-Systems beeinflusst werden.

Daraus ergibt sich unmittelbar, dass OLAP-Systeme performant und einfach zu bedienen sind, ihre Anfragen aus den Datenquellen in einem multidimensionalen Datenwürfel zusammengefasst werden und sie ihre Berichte in Form von Tabellen und Grafiken darstellen. Der Anwender kann einzelne Kriterien selektieren und miteinander kombinieren.

5.2 Berichtssysteme

Berichtssysteme erlauben eine einfache, übersichtliche Auswertung und Präsentation von Unternehmensdaten.

Grundsätzlich existiert eine Vielzahl von Software-Lösungen am Markt zur Erstellung und Gestaltung von Berichten. Grafische Oberflächen sowie Drag-and-drop-Techniken unterstützen den Analysten bei der Berichterstellung.

Scorecards und Dashboards

Im Rahmen der Berichterstellung werden eine Reihe aussagekräftiger Darstellungsmittel verwendet. Verbreitete grafische Präsentationsmittel sind die sogenannten „Scorecards“ und „Dashboards“.

Darin werden zumeist die sogenannten **Key Performance Indikatoren (KPI)** dargestellt. Grundsätzlich sind KPI betriebswirtschaftliche Schlüsselkennziffern, welche die Erreichung der strategischen Ziele repräsentieren. Bei Unternehmens-Websites wäre ein solcher Indikator etwa die durchschnittliche Verweildauer von Besuchern oder der über die Website generierte Umsatz.

Key Performance Indikatoren (KPI)
Diese sind betriebswirtschaftliche Schlüsselkennziffern.

Die **Scorecards** bieten eine Momentaufnahme entscheidungsrelevanter Daten auf einen Blick. Scorecards können mit einem Armaturenbrett in Fahrzeugen verglichen werden. Zugrunde liegen dabei zumeist große Mengen von verteilten Informationen (z. B. KPI) in verdichteter Form. Der Grad und die Visualisierungsform von Verdichtungen sind kontextabhängig. Häufig ist die Darstellung von Kennzahlen als Ampel-, Tachometer- oder Thermometerdarstellung vorzufinden.

Scorecards
Hierbei werden Daten auf einen Blick in visualisierter Form angeboten.

Das **Dashboard** visualisiert die Schlüsselkennzahlen aus unterschiedlichen Unternehmensbereichen in einer einheitlichen Bildschirmdarstellung mittels einfacher Geschäftsgrafiken und Tabellen. Zudem kann in Form von E-Mails oder SMS automatisch Alarm ausgelöst werden, sobald ein bestimmter Grenzwert unter- bzw. überschritten wird.

Dashboard
Dieses visualisiert Kennzahlen mittels Grafiken und Tabellen.

In der Praxis werden die Begriffe „Dashboard“ und „Scorecard“ häufig synonym verwendet. Allerdings besteht ein wesentlicher Unterschied. Bei einer Scorecard handelt es sich um einen Berichtstyp, in dem eine Sammlung von KPIs (Key Performance Indicators) und die Leistungsziele für die einzelnen KPIs angezeigt werden. Bei einem Dashboard handelt es sich dagegen um einen Container für eine zusammengehörige Gruppe von Scorecard- und Berichtsansichten, die zusammen angezeigt werden. Ein Dashboard enthält also eine Sammlung anderer Elemente wie Scorecards, Berichte und Filter.

Im Folgenden werden zwei Arten von Berichtssystemen dargestellt. Außerdem wird der Unterschied von MIS und EIS aufgezeigt.

Generierte Berichte

Management-Informationssysteme (MIS)

Diese sind berichtsorientierte Analysesysteme.

Die **Management-Informationssysteme (MIS)** sind berichtsorientierte Analysesysteme, die auf die Planung, Steuerung und Kontrolle der operativen Wertschöpfungskette ausgerichtet sind. MIS werden eingesetzt, um der Unternehmensleitung den Überblick über Geschäftsprozesse zu erleichtern und damit erforderliche unternehmerische Entscheidungen treffen zu können. Die Systeme bieten leistungsstarke Visualisierungs- und Analysetools und liefern attraktive Grafiken und Planungsinstrumente. Darüber hinaus können Tendenzen visualisiert und „Was-Wäre-Wenn-Analysen“ durchgeführt werden.

Executive Information System (EIS)

Dieses ist ein bereichsübergreifendes Informationssystem für die oberste Managementebene.

Das **Executive Information System (EIS)** kann definiert werden als ein unternehmensspezifisches und bereichsübergreifendes integratives und dynamisches Informationssystem zur informationellen Unterstützung der obersten Managementebene, das über ein großes Maß an Flexibilität und einen hohen Bedienungskomfort verfügt (Kemper/Baars/Mehanna 2010).

Moderne MIS und EIS ähneln sich häufig im äußeren Erscheinungsbild bzw. in der Form der Präsentation und Oberfläche. In Abgrenzung zum MIS präsentiert EIS jedoch primär hochverdichtete, steuerungsrelevante interne Daten sowie unternehmensexterne Informationen, wobei auch unstrukturierte Informationen integrierbar sind. Aufgrund dessen wird EIS auf einer höheren hierarchischen Ebene eingesetzt (Kemper/Baars/Mehanna 2010).

5.3 Modellgestützte Analysesysteme

Modellgestützte Analysesysteme

Diesen liegen betriebswirtschaftliche Modelle zugrunde.

Während bei der freien Datenrecherche und den OLAP-Systemen meist kleinere Berechnungen durchgeführt werden, erfordern komplexe Auswertungen **modellgestützte Analysesysteme**, die eine ausgeprägte algorithmische oder regelbasierte Ausrichtung aufweisen. Zu dieser Kategorie gehören die Decision Support Systems, Expertensysteme und das Data Mining (Kemper/Baars/Mehanna 2010). Modellgestützten Analysesystemen liegen demnach betriebswirtschaftliche Modelle zugrunde, nach denen Berechnungen und Analysen vorgenommen werden. Es handelt sich daher um Auswertungen mit höherer Komplexität und höherem Abstraktionsgrad.

Entscheidungsunterstützung

Decision Support System (DSS)

Dieses ist ein interaktives, modell- und formelbasiertes System.

In Anlehnung an die allgemein akzeptierte Definition wird unter **Decision Support System (DSS)** ein interaktives, modell- und formelbasiertes System verstanden (Kemper/Baars/Mehanna 2010). Ein Decision-Support-System bedient sich also bestehender Modelle und Formeln, um das Management in mehr oder weniger strukturierten Entscheidungssituationen zu unterstützen.

Durch Entscheidungsunterstützungs-Systeme geben die Systeme unter Berücksichtigung der vorliegenden Daten auch Handlungsempfehlungen. Schwerpunkt ist die Planung im engeren Sinne und zwar die Untersuchung möglicher Handlungsalternativen durch mathematische Methoden und Modelle (Hansen/Mendling/Neumann 2019, S. 351).

Expertensysteme (XPS)

Das **Expertensystem** ist ein Informationssystem, das fachspezifische Kenntnisse in einem abgegrenzten Anwendungsbereich verfügbar macht. Wesentliche Bestandteile sind eine Wissensbasis (Datenbank mit Expertenwissen) und eine Problemlösungskomponente (Puppe 1991). Hierbei ist zu beachten, dass neben dem Fachwissen der Spezialisten auch deren Problemlösungstechniken in das Expertensystem einfließen. Somit kann die Erfahrung von Experten auf eine ganzheitliche Weise genutzt werden (Kemper/Baars/Mehanna 2010).

Expertensystem (XPS)
Dieses macht fachspezifische Kenntnisse in einem abgegrenzten Anwendungsbereich verfügbar.

Expertensysteme bringen die gelernten Heuristiken mit Fachwissen von Experten in Verbindung und sind dadurch in der Lage, neue Fragen und Probleme bis zu einem gewissen Grad eigenständig zu lösen. Außerdem kann das Wissen des Systems, aufgrund der integrierten Wissenserwerbskomponente, erweitert und veraltete bzw. falsche Informationen berichtigt werden (Puppe 1991). Die Erklärungskomponente dient dazu, die Prozeduren des Expertensystems durchschaubar zu machen. Daher werden XPS dem Forschungsgebiet der Künstlichen Intelligenz zugerechnet.

Expertensysteme werden heute zumeist im Banken- und Versicherungsgewerbe für die Kreditwürdigkeitsprüfung und Risikoanalysen angewandt. Häufig sind XPS Bestandteil integrierter Anwendungen, z. B. in Form von interaktiven Hilfesystemen.

Data Mining

Als **Data Mining** wird die softwaregestützte Ermittlung bisher unbekannter Zusammenhänge, Muster und Trends aus dem Datenbestand sehr großer Datenbanken bzw. DWHs bezeichnet. Im Gegensatz zu den klassischen Abfragewerkzeugen muss der Analyst nicht von vornherein wissen, wonach er sucht. Vielmehr wird der Anwender zu den interessantesten Informationen geführt.

Data Mining
Darunter versteht man die Ermittlung bisher unbekannter Zusammenhänge, Muster und Trends.

Typische Aufgabenstellungen des Data Mining sind:

- **Ausreißerererkennung:** dabei werden ungewöhnliche Datensätzen identifiziert (z. B. Ausreißer, Fehler, Änderungen).
- **Clusteranalyse:** Sie gruppiert Objekte aufgrund von Ähnlichkeiten.
- **Klassifikation:** Hier werden bisher nicht zugeordnete Elemente entsprechenden Klassen zugeordnet.
- **Assoziationsanalyse:** Durch sie erfolgt die Identifizierung von Zusammenhängen und Abhängigkeiten in den Daten in Form von Regeln wie „Aus A und B folgt normalerweise C“.

- **Regressionsanalyse:** Sie behandelt die Identifizierung von Beziehungen zwischen (mehreren) abhängigen und unabhängigen Variablen.
- **Zusammenfassung:** Mit ihrer Hilfe erfolgt die Reduktion des Datensatzes in eine kompaktere Beschreibung ohne wesentlichen Informationsverlust.

Banken verwenden Data Mining beispielsweise zur Erkennung von Kreditkartenbetrug und für die Profilerstellung von Kunden, die mit gewisser Wahrscheinlichkeit ihre Kreditverbindlichkeiten nicht erfüllen können. Im Marketing wird Data Mining genutzt, um Absatzprognosen, Kundensegmentierungen, Warenkorbanalysen und Missbrauchserkennungen durchzuführen. Im Personalwesen können die Personalauswahl und Mitarbeiterfehleistungserkennung durch Data Mining unterstützt werden. Einen starken Aufwind erlebte Data Mining mit der Zunahme von Web-Anwendungen. Dabei hat sich auch der Begriff Web Mining etabliert.

5.4 Konzeptorientierte Systeme

Konzeptorientierte Systeme

Diese umfassen Werkzeuge, die umfangreiche betriebswirtschaftliche Konzepte in BI-Analysen umsetzen.

Die **konzeptorientierten Systeme** verstehen sich als diejenigen Tools innerhalb von BI, welche Analysen und Daten, aufbauend auf umfangreichen betriebswirtschaftlichen Konzepten oder Verfahren, liefern.

Die Balanced Scorecard als Managementkonzept gehört in den heutigen BI-Systemen zum Standard der konzeptorientierten Systeme. Daneben stellen BI-Lösungen diverse Tools für die Planung und Konsolidierung zur Verfügung sowie für weitere Bereiche, z. B. das wertorientierte Management (Kemper/Baars/Mehanna 2010).

Balanced Scorecard (BSC)

Mithilfe der Balanced Scorecard wird das Blickfeld des Managements auf alle relevanten Teile des Unternehmens gelenkt.

Die **Balanced Scorecard (BSC)** ist eine spezielle Form der Scorecards bzw. Cockpits. BSCs sind ursprünglich das Resultat eines Forschungsprojekts, das Anfang der neunziger Jahre unter der Leitung von Robert S. Kaplan und David P. Norton durchgeführt wurde. Der Auslöser für das Projekt war die Unzufriedenheit mit der eindimensionalen, finanziell orientierten Beschreibung und Steuerung von Unternehmen. Die Verwendung von nur einer Dimension wird der Realität nicht gerecht.

Das mit BSC neu eingeführte Element besteht darin, dass nicht nur auf die Finanzperspektive fokussiert wird, sondern auch andere, menschliche Aspekte betrachtet werden. Mit den Methoden der BSC wird das Blickfeld des Managements auf alle relevanten Teile von Unternehmenssichten gelenkt und zu einem ausgewogenen („balanced“) Bild geführt. Um eine ausgeglichene Steuerung zu gewährleisten wird das Unternehmen durch die Balanced Scorecard aus vier verschiedenen Perspektiven betrachtet: Neben der Finanzperspektive sind die internen Geschäftsprozessperspektiven, die Kundenperspektive und die Lern-

und Wachstumsperspektive zu berücksichtigen. Die Perspektiven sind allerdings nur ein bewährtes Basisgerüst, das im Einzelfall durch unternehmensindividuelle Perspektiven ergänzt werden kann.



ZUSAMMENFASSUNG

Um aus Unternehmensdaten geschäftsrelevante Informationen zu generieren, müssen diese entsprechend analysiert werden. Grundsätzlich können Analysesysteme nach freier Datenrecherche, nach Ad-hoc-Analysen oder Berichtssystemen sowie modellgestützten und konzeptorientierten Systemen unterschieden werden.

Bei der freien Datenrecherche können Daten relativ einfach mit SQL oder MDX abgefragt werden.

Mithilfe von Ad-hoc-Analysen bzw. OLAP erhält der Analyst die Möglichkeit, entsprechende Auswertungen „live“ bzw. „online“ anpassen zu können, um je nach geschäftlicher Anforderung verschiedene Sichten auf die Daten zu erhalten.

Die Berichtssysteme (z. B. Scorecards, Dashboards, MIS, EIS) erlauben eine einfache, übersichtliche Auswertung und Präsentation von Unternehmensdaten.

Während bei der freien Datenrecherche und OLAP-Systemen zumeist kleinere Berechnungen durchgeführt werden, erfordern komplexe Auswertungen modellgestützte Systeme, die eine ausgeprägte algorithmische oder regelbasierte Ausrichtung aufweisen. Zu dieser Kategorie gehören die Decision Support Systems, Expertensysteme und das Data Mining.

Unter konzeptorientierten Systemen sind diejenigen Tools zu verstehen, die Analysen und Daten, aufbauend auf umfangreichen, betriebswirtschaftlichen Konzepten oder Verfahren, liefern.

LEKTION 6

DISTRIBUTION UND ZUGRIFF

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- wie aus BI-Inhalten Wissen generiert wird.
- welche Systeme und Formate bei der Informationsdistribution unterstützen.
- wie mithilfe von Portalen der Informationszugriff erleichtert wird.

6. DISTRIBUTION UND ZUGRIFF

Einführung

BI-Analysen und deren Ergebnisse liefern wichtige Erkenntnisse, die unternehmensrelevante Entscheidungen wesentlich unterstützen. Jedoch werden die Erkenntnisse von Mitarbeitern nicht immer ausreichend berücksichtigt. In der Praxis kann es dazu kommen, dass Analyseergebnisse nur von wenigen Benutzern angewandt werden. BI-Analysen werden häufig nicht weitergegeben, sodass manchen Entscheidungsträgern und Abteilungen wichtige Analyseresultate verschlossen bleiben.

Die Gründe für die suboptimale Informationsversorgung sind vielfältig. Beispielsweise kann nur ein eingeschränkter Nutzerkreis auf die Analyseergebnisse von Data-Mining-Systemen zugreifen. Zudem kann die Auswahl von Adressaten für die manuelle Weitergabe der Analyseergebnisse subjektiv erfolgen. Die häufigste Ursache sind allerdings BI-Analysen, die von bestimmten Abteilungen durchgeführt werden, ohne dass andere davon Kenntnis besitzen.

Durch die mangelhafte Informationsversorgung wird der Erfolg von BI-Analysen langfristig untergraben. Aufgrund dessen muss die Unternehmenskommunikation verbessert und BI an das Wissensmanagement von Unternehmen angebunden werden. Die Ergebnisse von BI-Analysen müssen dokumentiert und in adäquater Form zur Verfügung gestellt werden, damit die Informationen unternehmensweit verwendet werden können. Content-Management- und Dokumenten-Management-Systeme sowie zentrale, personalisierbare BI-Portale leisten hier wertvolle Hilfe.

6.1 Informationsdistribution

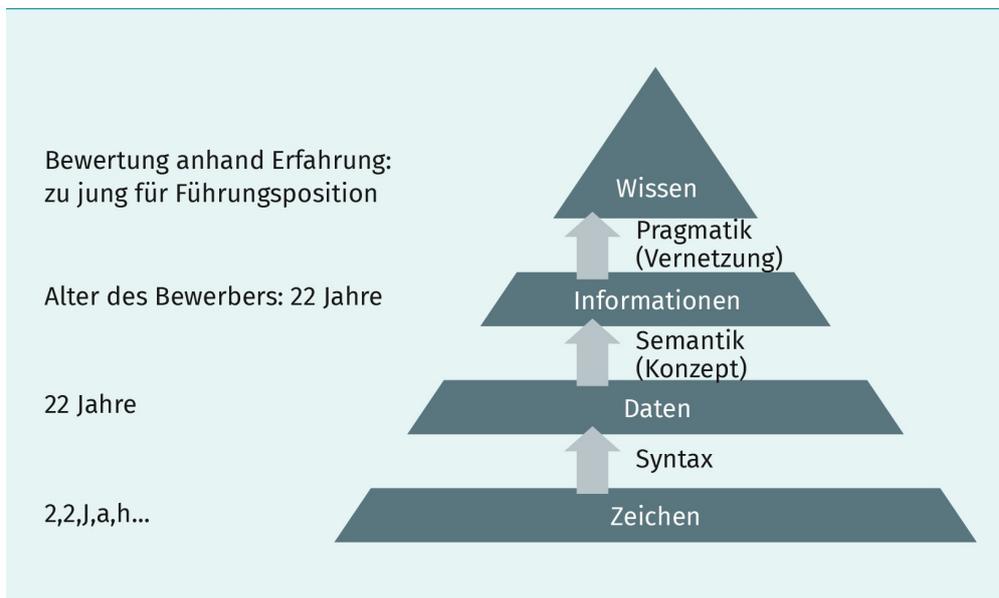
Im Rahmen der Entwicklung und Nutzung von BI-Lösungen wird wertvolles Wissen aufgebaut. Die zielgerichtete Wiederverwendung von Wissen verspricht enorme Effizienzgewinne. BI-Anwender können beim Zugriff auf vorbereitete Ergebnisse profitieren. Hierdurch wird Doppelarbeit vermieden. Zudem können Entwickler auf bestehende bzw. konsolidierte Anwendungen zugreifen. Für die systematische Verbreitung kann auf die Werkzeuge des betrieblichen Wissensmanagements zurückgegriffen werden.

Hierbei kommt insbesondere der Klasse der Content- und Dokumenten-Management-Systeme eine bedeutende Rolle zu. Im Folgenden werden Werkzeuge für das Wissensmanagement (Content- und Dokumenten-Management-Systeme) beschrieben.

Wissensmanagement

Für die Einordnung von Wissensmanagementwerkzeugen im BI-Kontext wird zunächst der Wissensbegriff im Einzelnen beschrieben und in der folgenden Abbildung abgegrenzt.

Abbildung 22: Abgrenzung des Wissensbegriffs



Quelle: Bodendorf 2006, S. 1.



DEFINITION WISSEN

Wissen wird definiert als die Gesamtheit der Kenntnisse und Fähigkeiten, welche zur Lösung von Problemen eingesetzt werden.

Im Vergleich zur Information wird ein übergeordneter Begründungszusammenhang hergestellt (Pragmatik). Bei den Informationen ist diese Eigenschaft nicht vorhanden. Informationen besitzen zwar kontextabhängige Bedeutung (Semantik), fokussieren aber lediglich einzelne Aspekte eines Themenbereichs. Daten sind Zeichen, die nach vorgegebenen Regeln (Syntax) zusammengesetzt wurden.

Wissensmanagement ermöglicht Unternehmen, betriebliches Wissen zu dokumentieren und interessierten Mitarbeitern zur Verfügung zu stellen. Dementsprechend liefern sogenannte Wissensmanagementsysteme als technische Komponente die IT-basierte Unterstützung für das betriebliche Wissensmanagement. Hierbei spielt insbesondere der Grad der Kodifizierbarkeit des Wissens eine bedeutende Rolle. Darunter wird das Wissen verstanden, welches in strukturierter Form gespeichert bzw. dokumentiert werden kann. Damit kann das Wissen entsprechenden Anwenderkreisen zugänglich gemacht werden. Im Gegensatz dazu wird von implizitem Wissen gesprochen, wenn sich dieses in den Köpfen der Mitarbeiter festgesetzt und aufgrund der Komplexität kaum in elektronischer Form abgelegt werden kann. Der Austausch von implizitem Wissen ist primär durch zwischenmenschliche Kommunikation möglich.

Für die Verwaltung von kodifiziertem Wissen werden Content-Management-Systeme (CMS) und Dokumenten-Management-Systeme (DMS) eingesetzt, welche für die Handhabung von unstrukturierten Daten konzipiert sind. Die genannten Systeme besitzen beim Aufbau integrierter BI- und Wissensmanagement-Lösungen einen hohen Stellenwert. Die Integration wird teilweise dadurch erleichtert, dass führende End-to-End-Anbieter entsprechende BI- und CMS-Werkzeuge als zusammengehörende Produktsuiten anbieten.

Content-Management- und Dokumenten-Management-Systeme

Grundsätzlich unterstützen sowohl Content-Management- als auch Dokumenten-Management-Systeme unstrukturierte Daten.

Dokumenten-Management-System

Dieses System dient der effektiven Verwaltung von papiergebundenen Dokumenten.

Die **Dokumenten-Management-Systeme** wurden ursprünglich für die effiziente Verwaltung papiergebundener Dokumente konzipiert. Die Funktionalität von DMS umfasst die Erfassung, Archivierung, Versionsverwaltung und Bereitstellung von Dokumenten in elektronischer Form.

Content-Management-System

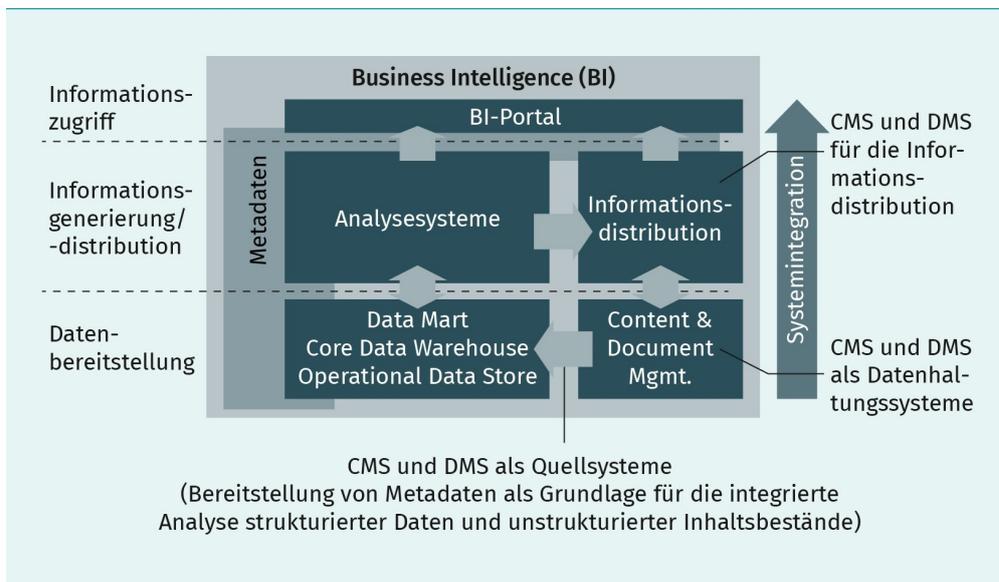
Dieses fokussiert auf den Umgang mit verschiedenen Medienformaten.

Demgegenüber fokussieren sich **Content-Management-Systeme** auf den Umgang mit heterogenen Medienformaten, wie z. B. numerischen Daten, Texten, Grafiken, Bildern, Audio- oder Videosequenzen.

Bei CMS erfolgt eine strikte Trennung von Inhalt, Struktur und Layout, um eine Mehrfachverwendung zu ermöglichen. Das CMS unterstützt das Einfügen, Aktualisieren und Archivieren von Beiträgen sowie deren inhaltliche Aufbereitung und die inhaltliche Zusammenstellung im Verwendungsfall. Dabei werden CMS unter anderem durch Verfahren zur Versionskontrolle, zur Berechtigungsvergabe sowie zur Qualitätssicherung unterstützt. Die verschiedenen Systeme wachsen aufgrund zunehmender webbasierter Infrastrukturen zusammen. DMS werden um CMS-Funktionen erweitert oder umgekehrt.

In folgender Abbildung ist zu erkennen, dass CMS und DMS in einem BI-Ansatz die Rollen der Informationsdistribution, Datenhaltungssysteme und Quellsysteme einnehmen können.

Abbildung 23: Mögliche Rollen von CMS und DMS in einem BI-Ansatz



Quelle: Kemper/Baars/Mehanna 2010, S. 146.

CMS und DMS verfügen über diverse Funktionen zur kontrollierten Verwaltung und Verbreitung von Inhalten. Die einzelnen Funktionen sowie deren Einsatzpotenziale im Kontext der Distribution von BI-Wissen werden im Folgenden beschrieben:

- dezidierte Zugriffssteuerung,
- Workflow-Steuerung,
- Lebenszykluskonzepte für Wissensseinheiten,
- Check-In/-Out: Steuerung konkurrierender Zugriffe,
- Zusammenfassung von Dokumenten,
- Versionsverwaltung,
- Information Retrieval.

Mithilfe entsprechender DMS- bzw. CMS-Funktionalität wird eine differenzierte Rollenzuweisung und Rechtevergabe ermöglicht. Dadurch kann der Zugriff auf verteiltes BI-Wissen zentral gesteuert werden (dedizierte Zugriffssteuerung).

Durch die Workflow-Steuerung werden mehrstufige Freigabeprozesse ermöglicht. Die Problematik eines „Information Overheads“ kann aufgrund der zugrunde liegenden Funktionalität vermieden werden.

Die verteilten BI-Inhalte besitzen zumeist keine unbegrenzte Gültigkeit. Analyseergebnisse und auch die dahinter stehenden Analysemodelle werden, z. B. aufgrund der Dynamik in Marktstrukturen, unbrauchbar. Aufgrund dessen werden Funktionalitäten angeboten, um BI-Dokumente mit einem Verfallsdatum zu versehen (Lebenszykluskonzepte für Wissensseinheiten).

Mit den sogenannten Check-In/-Out-Mechanismen können parallele Zugriffe vermieden werden, falls Inhalte nicht nur für den lesenden Zugriff genutzt werden. Beispielsweise sind Aktualisierungen erforderlich, um Modellparameter anzupassen oder obsoletere Ergebnisberichte zu aktualisieren.

Mithilfe von DMS können Einzeldokumente in einem Sammeldokument zusammengefasst werden. Korrespondierende Inhalte können dadurch gemeinsam hinterlegt werden, um z. B. für Balanced Scorecards entsprechende Berichte zu verschiedenen Detailsichten zu bündeln. Funktionen zur Versionsverwaltung erlauben die gezielte Rückverfolgung der Historie, sofern eingestellte BI-Inhalte bzw. Dokumente modifiziert werden dürfen.

Unter Information Retrieval werden Funktionen zum Auffinden von Dokumenten zu spezifischen Informationsbedarfen verstanden. Die entsprechende Funktionalität ist daher für die Informationsdistribution von zentraler Bedeutung. Der Integrationsnutzen wächst, wenn bereits ein entsprechender Dokumentenbestand in einem DMS oder CMS hinterlegt ist. Bei einer Anfrage werden sowohl die vorhandenen als auch ergänzte BI-Inhalte geliefert. Ein derartiger Ansatz ist dementsprechend auch als Schritt für die Zusammenführung von strukturierten und unstrukturierten Informationen zu verstehen.

Distribution

Die zu verteilenden Wissensinhalte lassen sich je nach Aufbereitung effizient und flexibel weinternutzen. Hierbei handelt es sich um eingeschränkt editierbare Dokumente bis hin zu parametrisierbaren Templates, die als Vorlagen für Anwendungen eingesetzt werden können. Grundsätzlich muss dabei unterschieden werden, ob die Analyseergebnisse selbst oder die Erstellung betroffen ist.

Distribution von Analyseergebnissen

DMS und CMS können genutzt werden, um die im BI-Kontext erstellten elektronischen Dokumente einer organisationsweiten Wiederverwendung zuzuführen. Hierbei sind insbesondere Ergebnisse betroffen, welche mit Analysesystemen erzeugt wurden. Diese werden typischerweise als MS-Excel-Dateien, PDF-Dokumente oder Webseiten aufbereitet.

Die Weiternutzung der Ergebnisse wird erleichtert, wenn die Vorhaltung der Informationen in einer maschinenlesbaren Form durchgeführt wird. Beispielsweise kann das CSV-Format von nahezu allen Programmen zur Datenaufbereitung und -analyse importiert werden. Hierbei werden Werte in einfachen Textdateien hinterlegt und durch Trennzeichen auf eine tabellarische Struktur abgebildet.

Eine weitergehende Flexibilität beim Datenaustausch kann durch Extensible Markup Language (XML) erzielt werden. Mit dem XML-Schema können weitreichende Struktur- und Formatvorgaben definiert werden. Die Mächtigkeit von XML wird dadurch begründet, dass verschiedene Dialekte kombiniert werden können.

Im BI-Bereich wird eine Vielzahl von Dialekten eingesetzt. Ein zunehmend verbreitetes Format ist hier XBRL. Die Extensible Business Reporting Language (XBRL) ist primär für den Austausch von Geschäftsdaten konzipiert. Mithilfe des Formats können Abschlussdaten

für bestimmte Rechnungslegungsstandards definiert und im Rahmen des externen Rechnungswesens veröffentlicht werden. Typische Adressaten sind Wirtschaftsprüfer, Anteilseigner, Analysten und staatliche Organisationen. Darüber hinaus eignet sich das Format für den Austausch zwischen Anwendungssystemen und das interne Reporting. Das Format wird zunehmend auch als Austauschformat im BI-Bereich angesehen, insbesondere aufgrund seiner Verbreitung, Einfachheit und Flexibilität.

Distribution von Teilsystemen

OLAP- und Data-Mining-Werkzeuge sind für die Lösung einmaliger und individueller Ergebnisse konzipiert. Eine Analyse handelt beispielsweise von einer einmaligen Verkaufsförderungsaktion oder der Ursachenaufdeckung bei einem regionalen Umsatzrückgang. Die einzelnen Ergebnisse besitzen in keinem anderen Kontext eine entsprechende Gültigkeit. Eine Verbreitung ist daher wenig sinnvoll.

Darüber hinaus ist jedoch das Wissen über die Vorbereitung und Durchführung der Analyse relevant. Festzuhalten sind das Analysemodell, Analyseverfahren, Dimensionen, Parameter, Datenanbindung und gegebenenfalls die Form und das Layout der Datenaufbereitung.

In ähnlicher Form kann auch bei der Berichtsentwicklung von der Wiederverwendung einer Reportdefinition profitiert werden. Das Schreiben einer Bedienungsanleitung ist mit erheblichem Aufwand verbunden. Die Wiederverwendbarkeit kann jedoch erleichtert werden, sofern maschinenlesbare Formate zur Definition von Anwendungsbausteinen ausgetauscht werden.

Neben den herstellerspezifischen Formaten können auch hier XML-Standards eingesetzt werden. Für die Beschreibung und den Austausch von Data-Mining-Modellen ist daher die Predictive Model Mining Language (PMML) konzipiert. Hier werden Datenquellen, vorbereitende Transformationen und Parameter des verwendeten Modells spezifiziert. In ähnlicher Form werden mithilfe der Report Definition Language (RDL) von Microsoft entsprechende Berichtsdefinitionen festgehalten. Hier werden Datenanbindung, Layout und Report-Metadaten berücksichtigt.

Die Professionalität der Wiederverwendung wird durch den Einsatz von individuell parametrisierten Anwendungsvorlagen (Templates) weiter erhöht. Templates sind im Idealfall angelegt, sodass manuelle Speicher-, Lade- oder Installationsaktivitäten entfallen. Führende Hersteller haben entsprechende Konzepte realisiert und liefern Standardtemplates für gängige Analyseanwendungen. Neben der Analyseschicht enthalten diese teilweise auch die Datenmodellierung und Anbindung der Datenquellen. In größeren BI-Umgebungen erlangt die zielgerichtete Distribution von Templates über DMS bzw. CMS eine zunehmende Relevanz.

6.2 Informationszugriff

Zugriff über Portale

Eine Benutzerschnittstelle für die Managementunterstützung muss die Komplexität der BI-Infrastruktur durch Integrationsleistung verbergen, um einen konsistenten und effizienten Zugriff zu gewährleisten. Zudem muss durch Personalisierung auf die verschiedenen Anforderungen eingegangen werden.

Grundsätzlich können BI-Inhalte auf verschiedenen Wegen zum Anwender gelangen, z. B. durch Analyse-Frontends, Bestandteil von Unternehmenswebseiten oder Integration in operative Anwendungen. Neuere Kanäle sind beispielsweise die sogenannten Feeds, welche für die Verteilung von Kurzinhalten an unterschiedliche Frontend-Anwendungen konzipiert wurden. Feeds werden beispielsweise für das Exception Reporting eingesetzt. Außerdem haben Widgets an Bedeutung gewonnen. Hierbei handelt es sich um eigenständige Programmkomponenten mit eigener Benutzerschnittstelle, die in eine grafische Bedienoberfläche integriert werden. Widgets werden häufig bei der Visualisierung von Prozesszuständen eingesetzt.

Neben stationären Geräten werden zunehmend auch internetfähige Mobilgeräte eingesetzt. Hier sind als Endgeräte insbesondere die Smartphones hervorzuheben. Die Realisierung von Lösungen wird durch offene und internetfähige Formate für die Übertragung und Konvertierung von Daten vereinfacht. Zudem erfolgt eine Erleichterung durch Technologien zur webbasierten Bereitstellung von Funktionen bzw. einzelnen Programmkomponenten. Eine besonders effektive und komfortable Variante, BI-Analysen den Endnutzern zu präsentieren, sind **Portale**. Hierbei handelt es sich zumeist um zentrale Web-Anwendungen, mit denen Unternehmen strukturierte Informationen anbieten können.

Portale

Dabei handelt es sich um zentrale Web-Anwendungen, mit denen Unternehmen strukturierte Informationen anbieten.

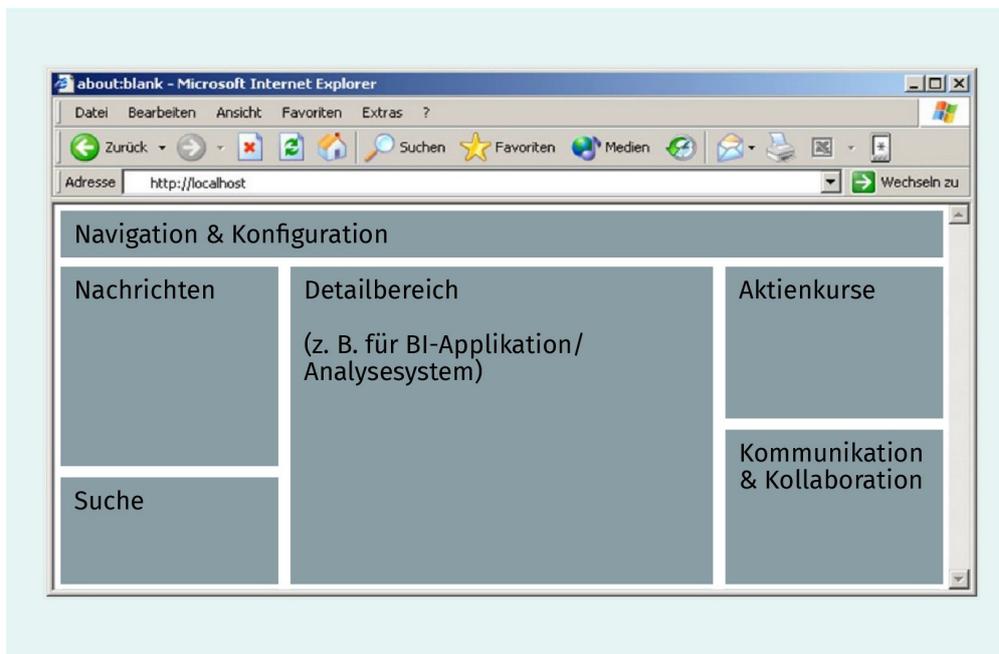
Integration von Inhalten

Das wichtigste Merkmal von Portalen ist, dass unterschiedliche Inhalte und Anwendungen unter einer gemeinsamen Oberfläche zusammengefasst werden. BI-Portale integrieren beispielsweise webbasierte Analysesysteme, die dem Nutzer ohne zusätzliche Installation von Client-Software zur Verfügung stehen. Das Management und andere Nutzer erhalten damit einen zentralisierten und strukturierten Zugriff auf die verfügbaren Informationen. Die Anwender verfügen über einen definierten Einstiegspunkt in das Unternehmenswissen und werden in die Lage versetzt, fundierte Unternehmensentscheidungen zu treffen. Technisch setzt sich ein BI-Portal aus mehreren Teilen zusammen, die als **Portlets** bezeichnet werden.

Portlets

Ein Portal setzt sich aus mehreren Teilen zusammen, die Portlets heißen.

Abbildung 24: Schematischer Aufbau eines Portals



Quelle: Kemper/Baars/Mehanna 2010, S. 158.

Ein Portlet kann beispielsweise für Nachrichten zuständig sein, ein anderes für BI-Applikationen und ein drittes für Kommunikation & Kollaboration. Die Inhalte können dabei aus internen oder externen Unternehmensquellen entstammen.

Zur Realisierung von BI-Portalen wird eine leistungsstarke Infrastruktur benötigt. Sie sollte zumindest aus folgenden Komponenten bestehen:

- Ein Portal-Framework mit Funktionen für Personalisierung, Single Sign-on (damit bekommt ein Portal-User entsprechend seines Berechtigungsprofils Zugriff auf alle benötigten Inhalte und Anwendungen) und geräteunabhängigen Zugriff. Letzteres soll mobilen Mitarbeitern den externen Zugriff von außerhalb des Unternehmens erlauben.
- Integration: Moderne Portale erlauben die Integration fast jeder Art von Datenquellen und Anwendungen: von beliebigen Datenbanken und einfachen Textdateien über Third-Party-Anwendungen wie ERP/CRM bis zu XML-Import via Internet. Die Verwendung von Portlets sorgt für eine schnelle Integration, wobei Nicht-Programmierer von zahlreichen Assistenten unterstützt werden.
- Folgende BI-Tools sollten enthalten sein: endbenutzergerechte Analyse und Reporting Werkzeuge, OLAP und Data Mining Tools sowie ETL-Tools.
- Basis-Dienste: Administration, Security, Directory Services (SSO) und andere Services erleichtern den Umgang mit dem Portal und sorgen für Sicherheit.
- Data-Warehouse-Infrastruktur: Sie erlaubt den Zugriff auf die dispositiven Daten und die Auswertung der Daten.

Um die Ergebnisse von BI-Analysen unternehmensweit nutzbar zu machen, müssen diese dokumentiert und in entsprechend aufbereiteter Form bestimmten Mitarbeitern zur Verfügung gestellt werden. Die in Unternehmen vorhandenen BI-Konzepte sollten deshalb in die Unternehmens-IT integriert werden.

Eine wichtige Rolle nehmen dabei Content-Management-Systeme ein. Sie erlauben es, Analyseergebnisse und Modelle berechtigten Interessenten gezielt und aufbereitet zur Verfügung zu stellen. Für den zentralen Informationszugriff bieten sich BI-Portale an. Mit diesen zentralen Web-Anwendungen lassen sich sämtliche BI-Analysen und Informationen unter einer einheitlichen Nutzeroberfläche zugänglich machen. Durch Personalisierung und Single Sign-on können BI-Portale zudem hoch individualisiert an die Bedürfnisse einzelner User angepasst werden.

Personalisierung

Personalisierung
Damit können Inhalte benutzerorientiert angeboten werden.

Eine **Personalisierung** kann grundsätzlich rollen- bzw. gruppenbezogen oder individuell eingeführt werden. Neben der Integration ist die Personalisierung ein weiteres zentrales Charakteristikum von Portalen. Die Personalisierung dient dazu, auf individuelle Bedürfnisse der Anwender zugeschnittene Ergebnisse anzubieten. Vertriebs- und Entwicklungsabteilung erhalten beispielsweise lediglich die Daten, welche für deren Verantwortungsbereich interessant bzw. relevant sind.

Die Rollen- bzw. gruppenbezogene Personalisierung wird einheitlich für bestimmte Rollen oder Gruppen durchgeführt, sodass für alle User, die einer bestimmten Gruppe angehören, die gleichen Einstellungen gelten. Die individuelle Personalisierung hingegen ist benutzerspezifisch und immer auf eine Person zugeschnitten. Eine individuelle Personalisierung kann explizit und implizit erfolgen. Bei der expliziten Personalisierung legt der User aktiv Einstellungen wie das Portal-Layout und Inhalte wie bestimmte Channels oder Anwendungen fest. Die implizite Personalisierung hingegen erfolgt durch Rückgriff auf Nutzerprofile und Nutzungsdaten und gibt selbstständig Empfehlungen relevanter Portalinhalte.

Eine weitere Möglichkeit die Benutzerorientierung zu erhöhen, ist das Konzept des Single Sign-on. Grundlage ist ein Verzeichnisdienst wie LDAP.



ZUSAMMENFASSUNG

BI-Analysen und deren Ergebnisse liefern wichtige Erkenntnisse, die unternehmensrelevante Entscheidungen unterstützen können. Aufgrund dessen muss die Unternehmenskommunikation verbessert und BI an das Wissensmanagement von Unternehmen angebunden werden.

Die Ergebnisse von BI-Analysen müssen dokumentiert und in adäquater Form zur Verfügung gestellt werden, damit die Informationen unternehmensweit verwendet werden können. Content-Management- und Dokumenten-Management-Systeme sowie zentrale, personalisierbare BI-Portale leisten hier wertvolle Hilfe.

Dokumenten-Management-Systeme wurden ursprünglich für die effiziente Verwaltung papiergebundener Dokumente konzipiert. Die Funktionalität von DMS umfasst die Erfassung, Archivierung, Versionenverwaltung und Bereitstellung von Dokumenten in elektronischer Form.

Demgegenüber fokussieren Content-Management-Systeme den Umgang mit heterogenen Medienformaten, wie z. B. numerischen Daten, Texten, Grafiken, Bildern, Audio- oder Videosequenzen.

Die Weiternutzung der Ergebnisse wird erleichtert, wenn die Vorhaltung der Informationen in einer maschinenlesbaren Form durchgeführt wird.

Eine besonders effektive und komfortable Variante, BI-Analysen Endnutzern zu präsentieren, sind Portale. Hierbei handelt es sich zumeist um zentrale Web-Anwendungen, mit denen Unternehmen strukturierte Informationen anbieten können.

ANHANG

LITERATURVERZEICHNIS

- Bachmann, R./Kemper, G. (2011): *Raus aus der BI-Falle. Wie Business Intelligence zum Erfolg wird*. 2. Auflage, mitp, Heidelberg.
- Bauer, A./Günzel, H. (2013): *Data Warehouse Systeme. Architektur, Entwicklung, Anwendung*. 4. Auflage, dpunkt Verlag, Heidelberg.
- Bodendorf, F. (2006): *Daten- und Wissensmanagement*. 2. Auflage, Springer Verlag, Berlin.
- Chamoni, P./Gluchowski, P. (2015): *Analytische Informationssysteme Business Intelligence-Technologien und -Anwendungen*. 5. Auflage, Springer Verlag, Berlin.
- Codd E. F./Codd S. B./Salley C. T. (1993): *Providing OLAP to User-Analysts: An IT Mandate*. Codd & Associates, Ann Arbor, Michigan.
- Gansor, T./Totok, A./Stock, S. (2010): *Von der Strategie zum Business Intelligence Competency Center (BICC). Konzeption – Betrieb – Praxis*. Carl Hanser Verlag, München.
- Gluchowski, P./Gabriel, R./Dittmar, C. (2008): *Management Support Systeme und Business Intelligence. Computergestützte Informationssysteme für Fach- und Führungskräfte*. 2. Auflage, Springer Verlag, Berlin, Heidelberg.
- Grothe, M. (2000): *Business Intelligence. Aus Informationen Wettbewerbsvorteile gewinnen*. Addison-Wesley, München.
- Grossmann, W./Rinderle-Ma, S. (2015): *Fundamentals of Business Intelligence*. Springer, Berlin/Heidelberg.
- Gutenberg, E. (1983): *Grundlagen der Betriebswirtschaft*. Band 1: Die Produktion. 18. Auflage, Springer Verlag, Berlin, Heidelberg.
- Hannig, U. (2002): *Knowledge Management und Business Intelligence*. Springer Verlag, Berlin.
- Hansen, H.-R./Medling, J./Neumann, G. (2019): *Wirtschaftsinformatik I. Grundlagen und Anwendungen*. 12. Auflage, De Gruyter Oldenbourg, Berlin/Boston.
- Humm, B./Wietek, F. (2005): *Architektur von Data Warehouses und Business Intelligence Systemen*. In: Informatik Spektrum, 23.02.2005, Heidelberg, S. 3–14.
- Inmon, W. (2005): *Building the Data Warehouse*. 4. Auflage, John Wiley & Sons, New York u. a.
- Kaplan, R./Norton, D. (1996): *Using the Balanced Scorecard as a Strategic Management System*. In: Harvard Business Review, Vol. 74, Issue 2, Boston, MA, pp. 75–85.

- Kemper, H.-G./Baars, H./Mehanna, W. (2010): *Business Intelligence – Grundlagen und praktische Anwendungen. Eine Einführung in die IT-basierte Managementunterstützung*. 3. Auflage, Vieweg + Teubner, Stuttgart.
- Keyes, J. (2006): *Knowledge Management, Business Intelligence, and Content Management: The IT Practitioner's Guide*. Taylor & Francis Ltd, London.
- Kimball, R./Caserta, J. (2004): *The Data Warehouse ETL Toolkit*. Wiley Publishing, Indianapolis.
- Kimball, R./Ross M. (2013): *The Data Warehouse Toolkit – The Definitive Guide to Dimensional Modeling*. 3. Auflage, Wiley Publishing, Indianapolis.
- Kurz, A. (1999): *Datawarehousing. Enabling technology*. mitp, Bonn.
- Muhsch, H./Behme, W. (Hrsg.) (1998): *Das Data-Warehouse-Konzept. Architektur – Datenmodelle – Anwendungen*. Springer Gabler, Wiesbaden.
- Müller, R. M./Lenz, H.-J. (2013): *Business Intelligence*. Springer Vieweg, Berlin/Heidelberg.
- Pendse, N./Creeth, R. (1995): *The OLAP-Report. Succeeding with On-Line Analytical Processing*. o.O.
- Puppe, F. (1991): *Einführung in Expertensysteme*. Springer, Berlin.
- Schinzer, H./Bange, C./Mertens, H. (1999): *Data Warehouse und Data Mining. Marktführende Produkte im Vergleich*. 2. Auflage, Vahlen, München.
- Schnider, D. et al. (2016): *Data Warehouse Blueprints: Business Intelligence in der Praxis*. Hanser Verlag, München.
- Turban, E./Sharda, R./Delen, D. (2014): *Business Intelligence and Analytics*. Pearson Education Canada, Toronto.
- Sherman, R. (2014): *Business Intelligence Guidebook: From Data Integration to Analytics*. Morgan Kaufmann, Burlington, Massachusetts.

ABBILDUNGSVERZEICHNIS

Abbildung 1: Historische Entwicklung	17
Abbildung 2: Abgrenzungen des DWH-Begriffs	20
Abbildung 3: Einordnung von BI	21
Tabelle 1: Charakteristika operativer und dispositiver Daten	25
Abbildung 4: BI-Referenzarchitektur	27
Abbildung 5: Unabhängige Data Marts	30
Abbildung 6: Data Marts mit abgestimmten Datenmodellen	31
Abbildung 7: Zentrales C-DWH	31
Abbildung 8: Mehrere C-DWHs	32
Abbildung 9: C-DWH und abhängige Data Marts	33
Abbildung 10: DWH-Architekturmix	34
Abbildung 11: ETL-Prozess	39
Tabelle 2: Teilprozesse der Transformation	40
Abbildung 12: Erste Transformationsschicht - Filterung	41
Tabelle 3: Mängelklassifikation im Rahmen der Bereinigung	42
Abbildung 13: Zweite Transformationsschicht - Harmonisierung	43
Abbildung 14: Dritte Transformationsschicht - Aggregation	46
Abbildung 15: Vierte Transformationsschicht - Anreicherung	47
Tabelle 4: Data Marts und Core Data Warehouse	50
Abbildung 16: Cube und Dimensionen	62
Abbildung 17: Roll-Up & Drill-Down	63

Abbildung 18: Slice-Operator	64
Abbildung 19: Dice-Operator	65
Abbildung 20: Star-Schema	67
Tabelle 5: Dimensionstabelle Produkte	68
Tabelle 6: Extraktionsdaten Produkte	68
Tabelle 7: Tabelle 7: Historisierung SCD Typ 1	69
Tabelle 8: Historisierung SCD Typ 2	70
Tabelle 9: Historisierung SCD Typ 3	70
Abbildung 21: Analysesysteme für das Management	74
Abbildung 22: Abgrenzung des Wissensbegriffs	85
Abbildung 23: Mögliche Rollen von CMS und DMS in einem BI-Ansatz	87
Abbildung 24: Schematischer Aufbau eines Portals	91

 **IU Internationale Hochschule GmbH**
IU International University of Applied Sciences
Juri-Gagarin-Ring 152
D-99084 Erfurt

 **Postanschrift**
Albert-Proeller-Straße 15-19
D-86675 Buchdorf

 media@iu.org
www.iu.org

 **Hilfe & Kontakt (FAQ)**
Antworten rund um Dein Studium findest
Du jederzeit auf myCampus.