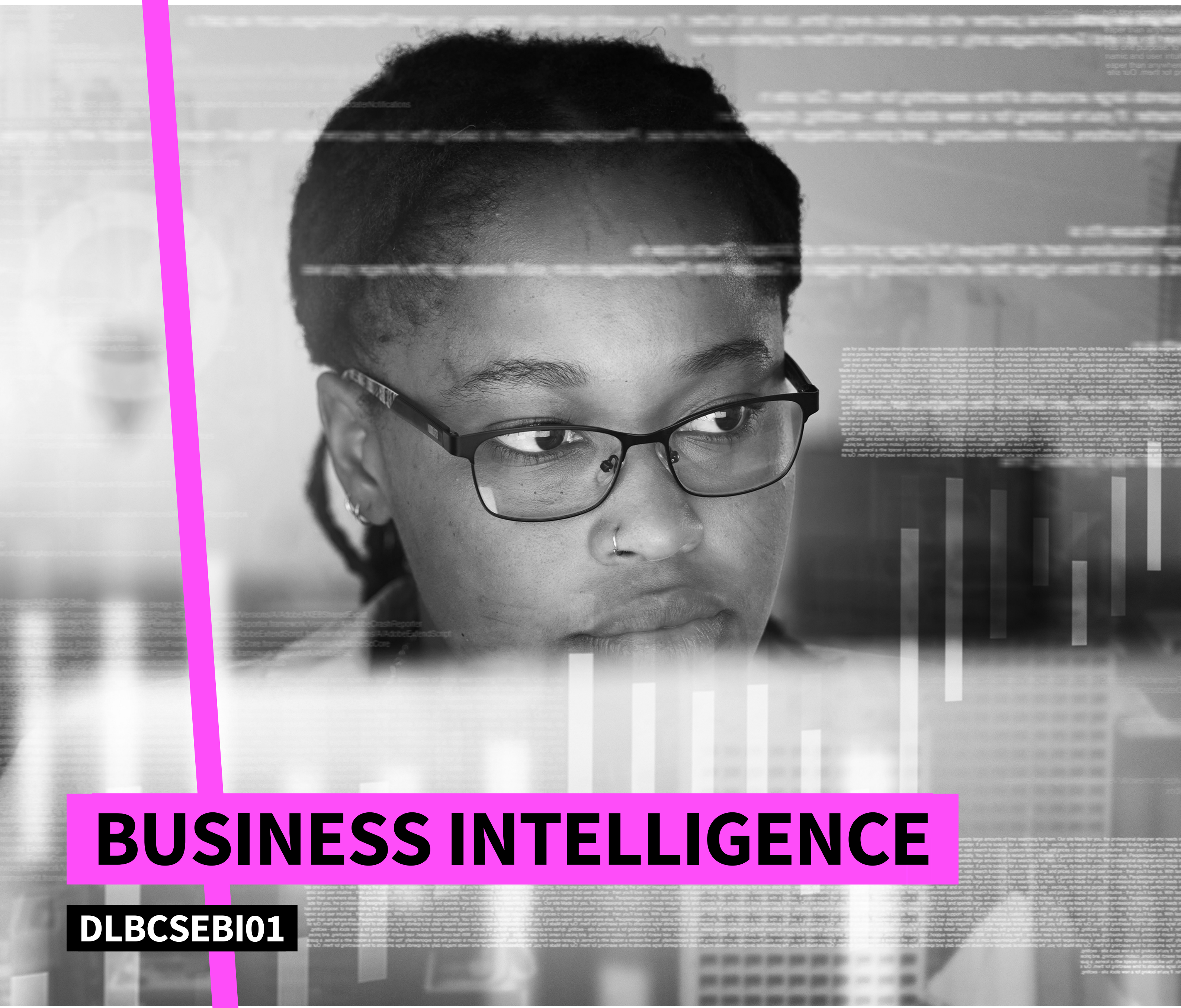
Course Book



BUSINESS INTELLIGENCE

DLBCSEBI01

IU INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

BUSINESS INTELLIGENCE

MASTHEAD

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

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

DLBCSEBI01
Version No.: 001-2023-0818
Gerasimos Marketos

TABLE OF CONTENTS

BUSINESS INTELLIGENCE

INTRODUCTION

WELCOME

SIGNPOSTS THROUGHOUT THE COURSE BOOK

This course book contains the core content for this course. Additional learning materials can be found on the learning platform, but this course book should form the basis for your learning.

The content of this course book is divided into units, which are divided further into sections. Each section contains only one new key concept to allow you to quickly and efficiently add new learning material to your existing knowledge.

At the end of each section of the digital course book, you will find self-check questions. These questions are designed to help you check whether you have understood the concepts in each section.

For all modules with a final exam, you must complete the knowledge tests on the learning platform. You will pass the knowledge test for each unit when you answer at least 80% of the questions correctly.

When you have passed the knowledge tests for all the units, the course is considered finished and you will be able to register for the final assessment. Please ensure that you complete the evaluation prior to registering for the assessment.

Good luck!

BASIC READING

Grossmann, W., & Rinderle-Ma, S. (2015). *Fundamentals of business intelligence*. Springer.

Kolb, J. (2013). *Business intelligence in plain language: A practical guide to data mining and business analytics*. Createspace.

Sharda, R., Delen, D., & Turban, E. (2014). *Business intelligence and analytics: Systems for decision support*. Pearson.

Sherman, R. (2014). *Business intelligence guidebook: From data integration to analytics*. Morgan Kaufmann.

Vaisman, A., & Zimányi, E. (2016). *Data warehouse systems: Design and implementation*. Springer.

REQUIRED READING

UNIT 1

Simon, A. (2014). *Modern enterprise business intelligence and data management.* Elsevier MK. Chapter 5

UNIT 2

Clegg, D. (2015). Evolving data warehouse and BI architectures: The big data challenge. *TDWI Business Intelligence Journal*, *20*(1), 19—24.

UNIT 3

Ankorion, I. (2005). Change data capture: Efficient ETL for real-time BI. *DM Review*, *15*(1), 36—43.

UNIT 4

Kimball, R. (2008). Slowly changing dimensions, types 2 and 3. *DM Review*, *18*(10), 19—38.

UNIT 5

Abkay, S. (2015). How big data applications are revolutionizing decision making. *Business Intelligence Journal*, *20*(1), 25—29.

UNIT 6

Gangadharan, G. R., & Swami, S. N. (2004). Business intelligence systems: Design and implementation strategies. *26th international conference, information technology interfaces (ITI 2004)* (pp. 139—144). The University of Zagreb Computing Centre (SRCE).

FURTHER READING

UNIT 1

Chaudhuri, S., Dayal, U., & Narasayya, V. R. (2011). An overview of business intelligence technology. *Communications of the ACM*, *51*(8), 88—98. Available online

Kawatzeck, R., & Dinter, B. (2015). Agile business intelligence: Collection and classification of agile business intelligence actions by means of a catalog and a selection guide. *Information Systems Management*, *32*(3), 177—191.

UNIT 2

Arizachandra, T., & Watson, H. J. (2008). Which data warehouse architecture is best? *Communications of the ACM*, *51*(10), 146—147.

Clegg, D. (2015). Evolving data warehouse and BI architectures: The big data challenge. *Business Intelligence Journal*, *20*(1), 19—24.

Ivan, M.-L. (2014). Characteristics of in-memory business intelligence. *Informatica Economica*, *18*(3), 17—25.

Zafary, F. (2020). Implementation of business intelligence considering the role of information systems integration and enterprise resource planning. *Journal of Intelligence Studies in Business*, *10*(1), 59—74.

UNIT 3

Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 134—142. Available online

Philip Chen, C. L., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, *275*, 314—347. Available online

UNIT 4

Franconi, E., & Kamblet, A. (2004). A data warehouse conceptual data model. *Proceedings of the 16th international conference on scientific and statistical database management (SSDBM 2004)* (pp. 435—436). IEEE.

Stiglich, P. (2014). Data modeling in the age of big data. *Business Intelligence Journal*, *19*(4), 17—

UNIT 5

Allio, M. (2012). Strategic dashboards: Designing and deploying them to improve implementation. *Strategy & Leadership*, *40*(5), 24—31.

Vincentdo, V., Pratama, A. R., Girsang, A. S., Suwandi, R., & Andrean, Y. P. (2019). Reporting and decision support using data warehouse for e-commerce top-up cell-phone credit transaction. *7th international conference on cyber and IT service management (CITSM)* (pp. 1—4). IEEE.

UNIT 6

Alpar, P., & Schulz, M. (2016). Self-service business intelligence. *Business & Information Systems Engineering*, *58*, 151—155.

Lennerholt, C., van Laere, J., & Söderström, E. (2018). Implementation challenges of self service business intelligence: A literature review. *Proceedings of the 51st Hawaii international conference on system sciences* (pp. 5055—5063).

Lennerholt, C., van Laere, J., & Söderstrom, E. (2018). User related challenges of self-service business intelligence. *Proceedings of the 53rd Hawaii international conference on system sciences* (pp. 188—197).

LEARNING OBJECTIVES

Business intelligence (BI) is a process used to extract information from company data that supports informed corporate management and the optimization of business activities. In the course **Business Intelligence**, the techniques, procedures, and models used in BI for data provision, information generation, and analysis, as well as the distribution of the information gained through BI processes, are presented and discussed. At the end of the course, you will be able to explain the various aspects of data warehousing and independently select methods or techniques to meet specific BI requirements. Ultimately, you will be able to independently design and prototype business intelligence applications based on concrete requirements.

UNIT 1

MOTIVATION AND INTRODUCTION

STUDY GOALS

On completion of this unit, you will have learned …

- what the term business intelligence (BI) means.
- how the term business intelligence was developed.
- the characteristics of a data warehouse.
- how the term business intelligence is defined in practice.

1. MOTIVATION AND INTRODUCTION

Introduction

For several years now, there has been a growing trend towards the globalization and dynamization of markets. As a result of greater competition, many companies now seek to create information advantages in order to establish overall competitive advantages. Information has thus become a managerial resource that is of strategic and tactical importance. An effective supply of relevant information is a prerequisite for improving the quality of corporate decision-making.

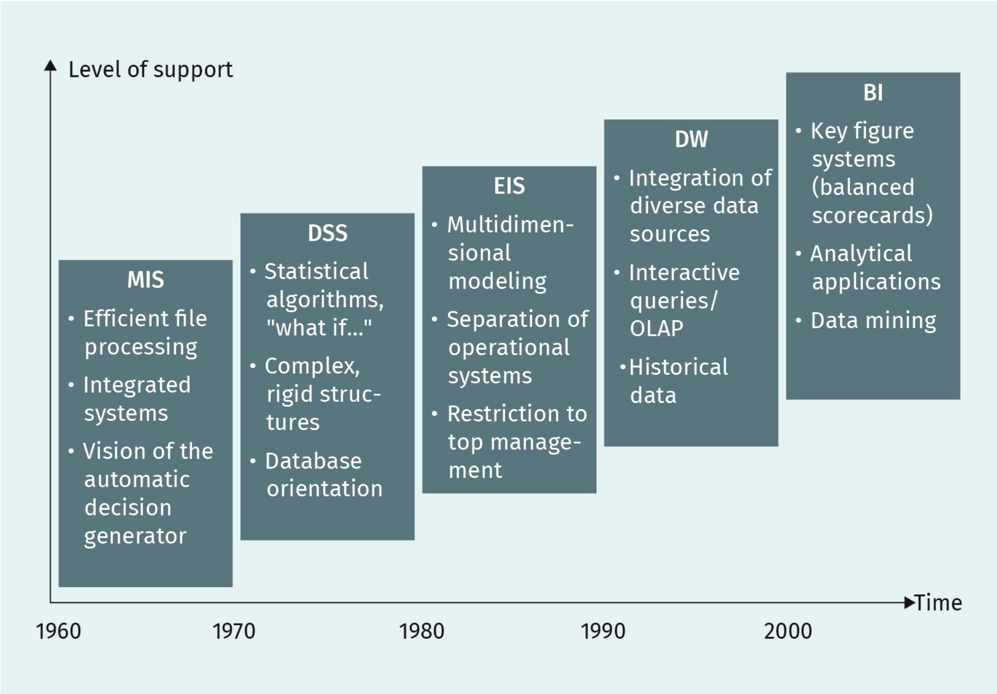
Business intelligence (BI) involves integrating strategies, processes, and technologies to generate critical knowledge about the current status and potential of the often fragmented divisions of a company. These perspectives on the company are then combined with market and competitor data in decision support systems which present this newly-acquired knowledge in such a way that it can be used directly for analysis, planning, and control purposes.

Located within the overarching concept of business intelligence is the data warehouse. The term data warehouse (DWH) is often understood differently due to the various definitions and interpretations that exist in the literature and in practice. In the following sections, we will explore the historical development of business intelligence. The term data warehouse is then described and positioned within the context of BI.

1.1 Motivation and Historical Development of the Field

The historical development of business intelligence goes back to the 1960s. As seen in the following figure, a number of different systems for supporting managerial decision-making have existed from that time. The term data warehouse was initially abbreviated to DW, but today, DWH is the more common abbreviation. You will find different abbreviations for data warehouse in the literature, particularly in some of the classic texts.

Figure 1: Historical Development



Source: Humm & Wietek, 2005, p. 4.

Management Information System (MIS)

At the end of the 1960s, the first information systems were introduced along with the term management information system (MIS). According to Grothe, the goal of a MIS was to “provide managers of companies with the information they need to make decisions. Time, content, and the way information was presented were to be optimized as secondary conditions” (Grothe, 2000, p. 65) (translated by author). However, according to Gluchowski et al. (2008), these goals could only be met to a limited extent because of limitations in the technology available at that time.

Decision Support System (DSS)

In the mid-1970s, the management information system was largely replaced by the decision support system (DSS). With the advent of interactive electronic data processing (EDP) systems, additional models, methods, and scenarios were made available to companies which enabled individual analyses of information (Gluchowski et al., 2008). Advances in hardware also made it possible to process information more efficiently. It was here that the basis for data-based decision support was laid (Grothe, 2000).

Unfortunately, decision support systems for the most part did not meet the high expectations associated with them. Thanks to technical progress, a DSS meant that structured data could be analyzed. However, the analysis of data was only possible for parts of the company. Moreover, this could only be done with operational data (Gluchowski et al.m

2008; Grothe, 2000). According to Hannig (2002), a further problem was that managers for the most part did not accept decision support systems as they did not trust computers to support creative decision-making processes.

Executive Information System (EIS)

In the mid-1980s, executive information systems (EIS) emerged at the same time as the arrival of powerful personal computers (PCs) in companies (Gluchowski et al., 2008; Hannig, 2002). The target users of executive information systems were primarily upper management and staff working in controlling functions. The EIS was comprised of individual systems that presented decision-relevant, multidimensional data to management in a more up-to-date and improved way than previous information systems (Gluchowski et al., 2008). In contrast to its predecessors (i.e., MIS and DSS), the EIS was easier to implement due to the spread of PCs in companies; with the MIS and DSS, central computers had to be used compared to the EIS operated using a PC. However, the disadvantage of using individual systems in the EIS was that they could only be used within a single department or company site, as they were developed individually for this purpose. As with decision support systems, the potential of executive information systems was not realized as they were not accepted by end users (Hannig 2002) and making any changes to the EIS—due to the individualized development of each system—was expensive (Grothe 2000).

Data Warehouse (DWH)

The main breakthrough in the acceptance and use of information systems came as a result of globalization, which accelerated at the beginning of the 1990s. Prior to this, managers had been largely skeptical about adopting such systems. However, as operations and supply chains spread across the globe, managers became more dependent on available information. Decision-making had also fundamentally changed as a result of decentralization. Decisions were no longer made in the head office (which could be located on the other side of the world), but rather were made promptly, locally, and using up-to-date information. Another reason for the increased demand for effective information systems was the flood of data facing companies, resulting from internationalization and the associated spread of company locations around the world. Previous systems (MIS, DSS, and EIS) were simply not able to meet these requirements.

A significant problem for many companies was that they now had to manage several inconsistent or non-compatible data sources. A new type of information management system was required: a complete, uniform, and consistent database (Hannig, 2002). A central database was developed that brought together data from the different systems used throughout a company; thus emerged the term data warehouse (DWH) (Grothe, 2000; Hannig, 2002).

In the 1990s, the creation of DWH analysis tools, often referred to as BI, was a major influence on the development of data warehouses. Today, the term BI is mostly used as a generic term (Grothe, 2000).

1.2 Business Intelligence as a Framework

Many companies today face the same scenario: a constantly increasing flood of data paired with insufficient useful information. Ensuring the effective supply of information to management is an important competitive factor. Often, however, the right information is not delivered in the right quantity, at the right place, at the right time. The aim of business intelligence is to ensure that it is. With the help of the data warehouse, operational information logistics can be improved and valuable and specific information can be delivered in a timely manner to management.

Features of a DWH

A DWH addresses the data problems experienced by management thus described. The father of data warehousing, W. H. Inmon, coined the following definition:

DATA WAREHOUSE

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

The four basic characteristics of a data warehouse are described below:

1. Subject-oriented (theme-focused) means that the data stock of a DWH are selected and organized according to profession or business criteria.
2. Integrated (unified) refers to the integration of data from heterogeneous source systems. The data must be standardized with regard to structure and format.
3. Nonvolatile (persistent) refers to the permanent storage of data in the DWH. Stored data are not changed or deleted.
4. Time-variant (historicization) means that time series analyses (comparison of data over time) are possible in the DWH. Data are stored as they existed at specific points in time. As a result, changes and developments over time can be analyzed.

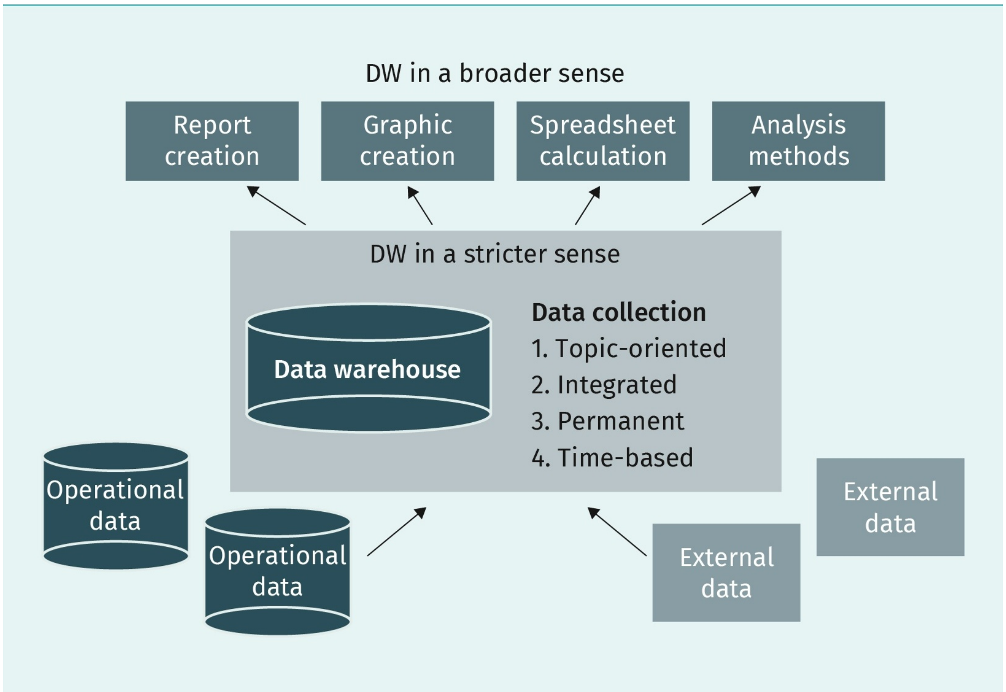
Definitions

There are a number of DWH and BI definitions found in the literature. Different interpretations of DWH and BI are therefore summarized in the following sections.

DWH

The aforementioned definition of a data warehouse supplied by Inmon has been extended by several authors, who describe additional tasks such as the connection, extraction, and transformation of external data as well as data collection and administration. According to Schinzer et al. (1999), the concept of the data warehouse also relates to the analysis and presentation of data with the help of appropriate tools.

Figure 2: Delimitation of the term "DWH"



Source: Glasker, 2017.

Data warehouse (DWH) in the narrower sense
This involves purely data collection.

This figure indicates that **data warehouse (DWH), in the narrower sense** of the word, covers purely data collection. DWH, in the broader sense, includes report generation, graph generation, spreadsheet analysis, and analysis methods.

BI

Archiving data alone does not bring about any competitive advantages. These are only realized through the creative and intelligent application of data (Muksch & Behme, 1996). The use of knowledge available across the company is known as BI. BI thus represents a further extension of the DWH concept in the broader sense and can be thought of as the front end of the DWH. The term was originally coined by the Gartner Group and is defined as follows:

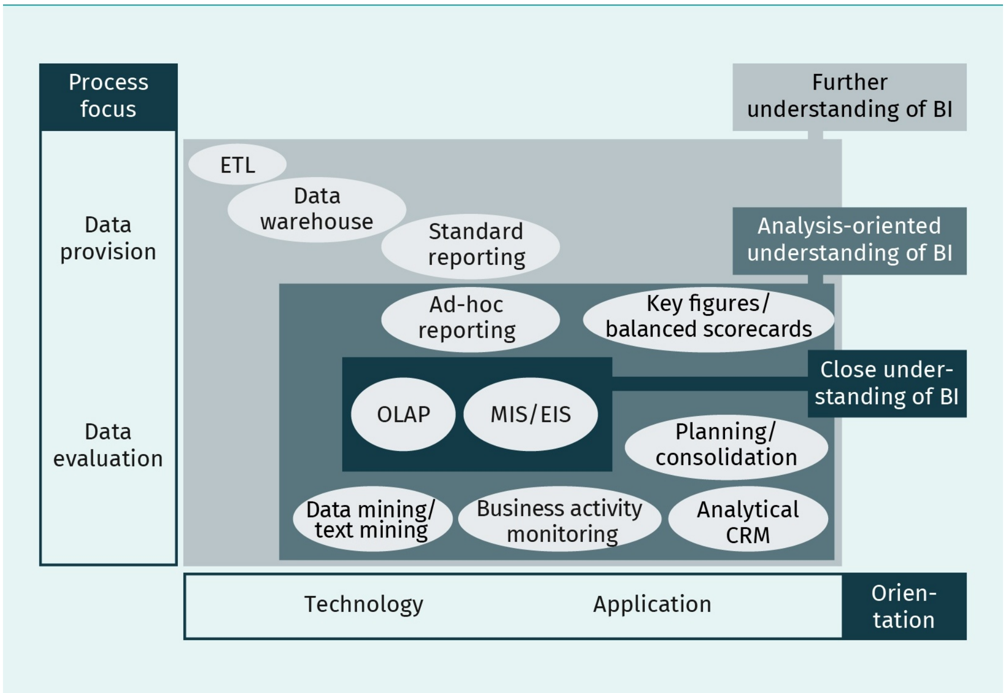
BUSINESS INTELLIGENCE

Business intelligence is the process of transforming data into information and, through discovery, into knowledge (Muksch & Behme, 1996, p. 37).

According to Gluchowski et al. (2008), BI involves techniques and applications that are designed to support decision-making and lead to a better understanding of the mechanisms driving outcomes.

The following figure shows how BI can be classified using a narrow understanding, an analysis-oriented understanding, and a broader understanding of the concept of BI.

Figure 3: Classification of BI



Source: Gluchowski et al., 2008, p. 92.

BI, in the narrower sense, refers to core applications that support decision-making without the additional input of more advanced methods or modeling. These include online analytical processing (OLAP), MIS, and EIS.

Analysis-oriented BI refers to the applications that allow decision makers to analyze existing data directly on the system using a user interface and various methods and models. These include OLAP, MIS and EIS, text mining, data mining, and ad hoc reporting.

Business intelligence (BI) in the broader sense
This includes all BI applications that are used directly or indirectly for decision-making.

Business intelligence (BI), in the broader sense, covers all applications that are used directly or indirectly for decision making. This includes evaluation and presentation functions as well as data preparation and storage (Gluchowski et al., 2008; Kemper et al., 2010).

SUMMARY

With the help of business intelligence, companies try to gain information advantages and thus establish competitive advantages. Information is now a managerial resource that has immense strategic value. The historical development of business intelligence goes back to the 1960s and information systems have gone through various iterations (BI, DWH, EIS, DSS, and MIS).

The primary characteristics of a DWH are the collection of subject-oriented, integrated, nonvolatile, and time-variant data. In the literature there is a multitude of DWH and BI definitions. These definitions characterize DWH or BI in both a narrower and broader sense. BI, in a broader sense, includes all applications that support decision-making either directly (e.g., OLAP) or indirectly (e.g., data extraction).

UNIT 2

DATA PROVISIONING

STUDY GOALS

On completion of this unit, you will have learned …

- how operational and dispositive systems differ from one another.
- what typical BI reference architecture looks like.
- which basic BI components exist.
- which architecture variants are possible.

2. DATA PROVISIONING

Introduction

The term business intelligence (BI) refers to procedures and processes that facilitate the systematic analysis of electronic data. Insights derived from the data then enable companies to make better operational or strategic decisions. The basic prerequisite for the use of powerful BI tools is the preparation and storage of consistent data that meets the business needs of managers.

From a technical perspective, a data warehouse (DWH) is realized via a database system. These often specialized database systems are configured for the needs of complex queries, since loading processes with high data volumes have additional requirements. The storage of data in tables based on rational database systems is widespread.

2.1 Operational and Dispositive Systems

Dispositive
Something is said to be dispositive if it relates to the management and control of operational processes.

Erich Gutenberg, a German economist and key figure in post-war, modern business studies, classified activities of the firm as operational and **dispositive** (Schmidt, 1998). According to Gutenberg, work is operational if it directly relates to the provision of goods and services, the utilization of goods and services, and the performance of financial tasks that are not of a planning nature. Activities are deemed to be dispositive if they relate to the management and control of operational processes.

In line with this distinction offered by Erich Gutenberg (1983), application systems and the data that exist within these systems can be distinguished as either operational or dispositive systems. Operational systems are about capturing and recording data, whereas dispositive systems are about analyzing data.

OLTP and OLAP

Operational systems
These are systems (e.g., ERP, CRM) that manage up-to-date information.

Operational systems are used to store and manage information necessary for the everyday operations of a company, e.g., a customer database or an employee directory. Information in these systems is regularly changed and frequently queried. Only current data records are of interest; past address data of a customer, for example, are of little value and can be deleted or overwritten. The data models used in such operational systems must be optimized for a high number of transactions. The processing method used for operational systems is known as online transactional processing (OLTP).

The DWH falls under the branch of dispositive systems. Dispositive systems are used to extract information from operational data. For example, it might be determined that a significant number of customers have relocated in the last six months and this information can be used to adapt and optimize sales structures. By using a DWH, operational systems are relieved of analytical queries, which might otherwise reduce processing capacities because of their complexity.

There are different types of queries conducted in operational and dispositive systems that aim to meet different objectives. Operational systems have a relatively large number of users. During business hours, numerous read requests are made for individual data records. Dispositive systems are generally queried by fewer people, but these people are individual experts seeking to address complex issues. They undertake sophisticated queries that evaluate a large number of data sets. The systems are optimized according to their respective application purpose in order to achieve an improved processing speed. The processing method used for dispositive systems is known as online analytical processing (OLAP). Due to the heterogeneity of transaction-oriented operational systems and the analytic-oriented data warehouse, both the systems and the data warehouse are physically separate from one another.

Operational and Dispositive Data

When considering a DWH, the question of technical necessity arises. After all, a DWH is only a replication of data that a company generates and stores in its data processing systems. However, the need for a DWH is understood when two different views of the data—operational and dispositive—are considered.

Operational data are directly related to the company's service provision activities. Dispositive data, on the other hand, are of an analytical nature and are used to manage and control the company (Kemper et al., 2010). In the following table, the most important differences between the two views are described.

Table 1: Characteristics of Operational and Dispositive Data

	Characteristics of operational data	**Characteristics of dispositive data**
Objective	Handling of business processes	Information for management; decision support
Alignment	Detailed, granular business transaction data	Mostly condensed, transformed data; comprehensive metadata
Time frame	Up-to-date; time-related; transaction-oriented	Different, task-dependent; history review
Modeling	Old stocks often not modeled (function-oriented)	Subject or topic-related, standardized and suitable for end users
Status	Often redundant; inconsistent	Consistently modeled; controlled redundancy
Update	Running and competing	Complementary; updating of derived, aggregated data

	Characteristics of operational data	Characteristics of dispositive data
Queries	Structured; mostly static in the program code	Ad-hoc for complex, constantly changing questions and ready-made standard evaluations

Source: Kemper et al., 2010, p. 16.

The differences between operational and dispositive data can be illustrated using the example of an insurance company.

Examples of operational data in this context are:

- detailed information on individual insurance contracts,
- continuous data changes recorded via the online portal, and
- storage of contracts in two different systems: one for motor vehicles and one for life insurance.

Examples of dispositive data are:

- summaries of sales and profits for each customer group,
- presentation of temporal changes compared to the previous year, and
- comparison of motor vehicle and life insurance product lines.

We can see that an evaluation based directly on operational data does not meet the requirements of the planning process. The heterogeneous system landscape in particular makes it difficult to compare information. Furthermore, the concept of concurrent “queries” and “transactions” would mean that conducting direct analyses on operational data would be problematic: resource-intensive queries with long runtimes have the potential to block the entire operational system and impair day-to-day business (Bauer & Günzel, 2008; Kemper et al., 2010).

2.2 The Data Warehouse Concept

In practice, the data warehouse can include different process phases, architectures, and BI components, depending on the requirements of the organization utilizing it. The following section explains basic concepts of reference architectures that must be customized for the actual project at hand.

Please note that the terms business intelligence and data warehouse (in its broader sense) have been used synonymously.

Process Phases and Reference Architecture

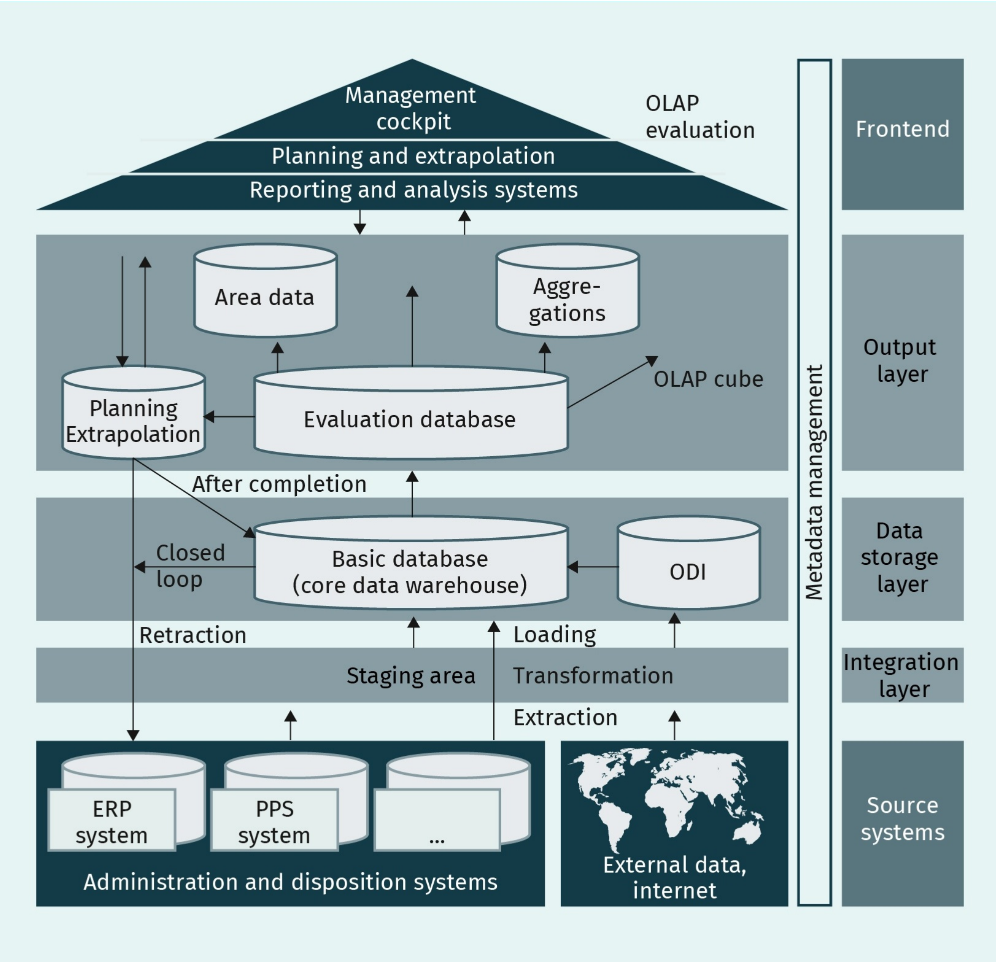
A large number of process phases and reference architectures exist in the literature. Process phases refer to the stages through which data passes. Data warehouse reference architecture refers to the template for designing the collection and storage of data using the data warehouse. According to Kemper et al. (2010), the process phases are as follows:

- data provision,
- information generation, storage, and distribution, and
- information access.

The first process phase of a DWH is to merge data and information from heterogeneous sources. The data can come from supply chain management (SCM), enterprise resource planning (ERP), customer relationship management (CRM), or external systems. All of these heterogeneous data are merged in the DWH. In the second step, data are analyzed using OLAP and data mining. In addition to extensive query options, these systems can also generate event-driven warning messages. In the third step, the findings from the second step are communicated to the company in the form of recommendations or actions.

The following figure from Gansor et al. (2010) illustrates the various components of BI reference architecture and provides context for their subsequent description.

Figure 4: BI Reference Architecture



Source: Gansor et al., 2010, p. 56.

BI components

Source systems

In classic BI reference architecture, data from a number of heterogeneous sources is imported, all with different structures, content, and access interfaces. OLTP systems are usually used, however, in principle, any type of source system can be conceivably included, e.g., semi-structured websites or unstructured text files. Source systems can include both internal company data (e.g., ERP, PPS system) and externally procured data (e.g., stock market prices, current raw material prices).

Staging area

The staging area is a work area in which data is temporarily stored. The staging area is necessary to relieve downstream systems when processing large amounts of data (Inmon, 2005).

Operational data store (ODS)

In contrast to the classic DWH approach, the ODS does not have aggregated data and longer history considerations. It is frequently used as a preliminary stage for supplying data for conventional DWH approaches (Kemper et al., 2010).

Basic database (core data warehouse)

The basic database is the central database within the DWH. After the initial transformation process, data are made available for various evaluation purposes or downstream systems.

Evaluation database (data mart)

The evaluation database forms the basis for downstream analysis tools. The data are stored with the help of a multidimensional model. From a technical point of view, evaluation databases are usually based on relational databases. Often, several evaluation databases are used and data are divided according to analysis requirements or organizational units (Bauer & Günzel, 2008).

Extracting, transforming, and loading (ETL) process

The extracting, transforming, and loading process integrates the data from the source systems into the DWH. The processing steps of extracting, transforming, and loading are carried out using ETL tools. From the data source systems, the data are transferred to the staging area via the data extraction step. After extracting the data from the data sources and loading them into the work area, the data must be converted according to the requirements of the company. Transformation affects both the structure and the content of the data. Data that comes from different sources must be converted into a uniform format. Plausibility checks can be used to improve data quality. The data are then transferred to the basic database (loading) as soon as they are available in a cleansed state following transformation. Since the basic database already contains integrated and cleansed data, data only need to be transformed into the target schema and possibly enriched or aggregated before they are loaded into the evaluation database (Bauer & Günzel, 2008).

Aggregation

Data are aggregated if they are required at a lower **granularity** than in the source systems. The process of aggregation considerably reduces the amount of data. For performance reasons, data are usually aggregated to the minimum required granularity. An example of aggregating data is combining daily sales into monthly sales.

Granularity
This is the level of detail data has within a data structure (“Granularity”, 2020).

Front end

Analysis tools form the front end of BI architecture. Front end tools can be more or less complex, depending on their application. Tools for data mining and OLAP are used to analyze the dataset and extract information from the mass of available data. Previously unknown relationships can be then uncovered, particularly through the use of data mining, where techniques such as classification and clustering are used. OLAP tools make the dataset accessible in an interactive way. The choice to aggregate data and the degree of aggregation for displayed data can be determined by the user. Portal systems are usually used to access information (Kemper et al., 2010).

2.3 Architecture Variants

In practice, there is a large number of architecture variants used for constructing data warehouses, some of which have been borrowed from other areas of data management while others have actually emerged from the BI field itself. In this section, several known architecture variants have been listed and then described in further detail. The basic architectural variants which can be used to create a DWH include

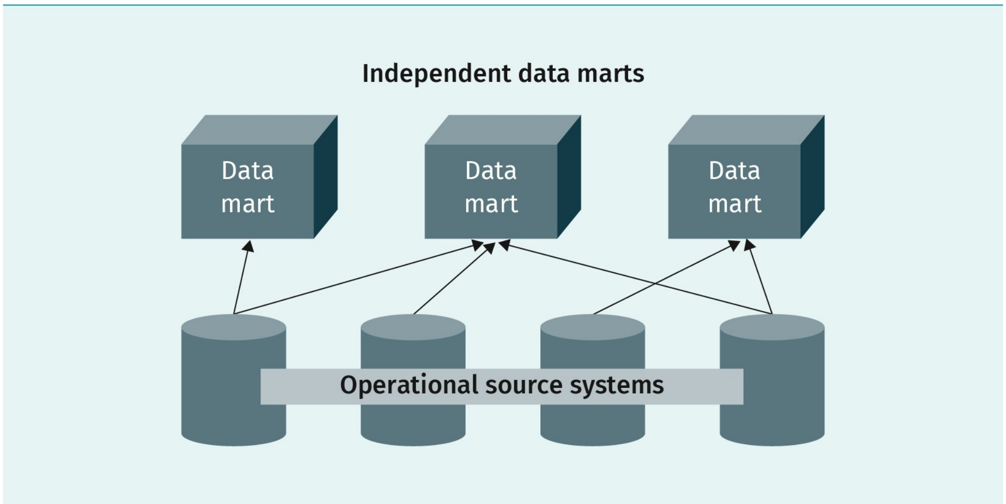
- independent data marts,
- data marts with coordinated data models,
- central core data warehouse (C-DWH) (no data marts),
- several C-DWHs,
- C-DWH and dependent data marts, and
- DWH architecture mix.

Independent Data Marts

Independent data marts
These are where independent DWHs are created in individual departments.

In practice, the architecture form of **independent data marts** is often created by individual departments building their own DWHs independently of each other, as seen in the following figure.

Figure 5: Independent Data Marts



Source: Kemper et al., 2010, p. 22.

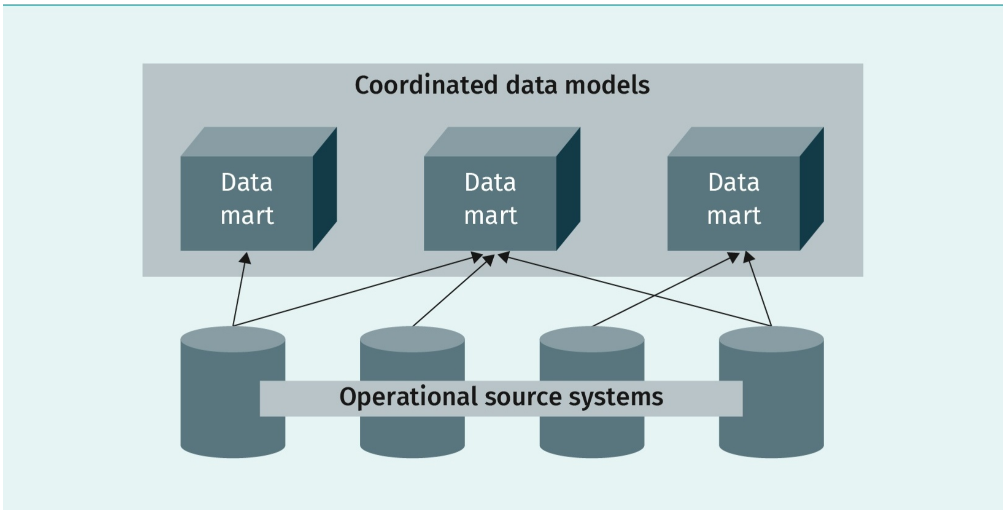
When using independent data marts, a central database (core data warehouse) is not required. This approach reduces the complexity of the entire DWH, making it easier and more manageable. As a result, usable results for the departments can be achieved in a relatively short period of time. However, the development of a company-wide data warehouse becomes much more difficult due to the subsequent isolation of applications (Kemper et al., 2010).

Data Marts with Coordinated Data Models

As in the previous variant, source data is prepared several times for different data management systems. However, the individual **data marts** coordinate with each other with regard to a common data model, as seen in the following figure.

Data marts
When data marts have coordinated data models, there are several data marts using a common data model.

Figure 6: Data Marts with Coordinated Data Models



Source: Kemper et al., 2010, p. 22.

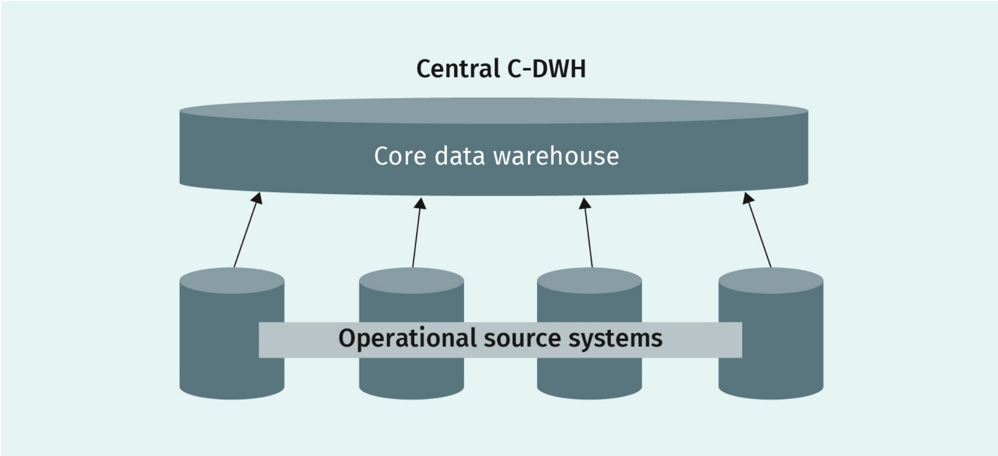
The use of data marts with conceptually coordinated data models ensures the consistency and integrity of the dispositive data model. Compared to the first variant, the establishment of a company-wide data warehouse will involve less effort (Kemper et al., 2010).

Central C-DWH (No Data Marts)

For smaller BI solutions, it may make sense to dispense with data marts, e.g., if the number of end users and data volumes are small. In this case, we recommend the variant **Central C-DWH**.

Central C-DWH
A central C-DWH places the evaluation function of the C-DWH in the foreground.

Figure 7: Central C-DWH



Source: Kemper et al., 2010, p. 22.

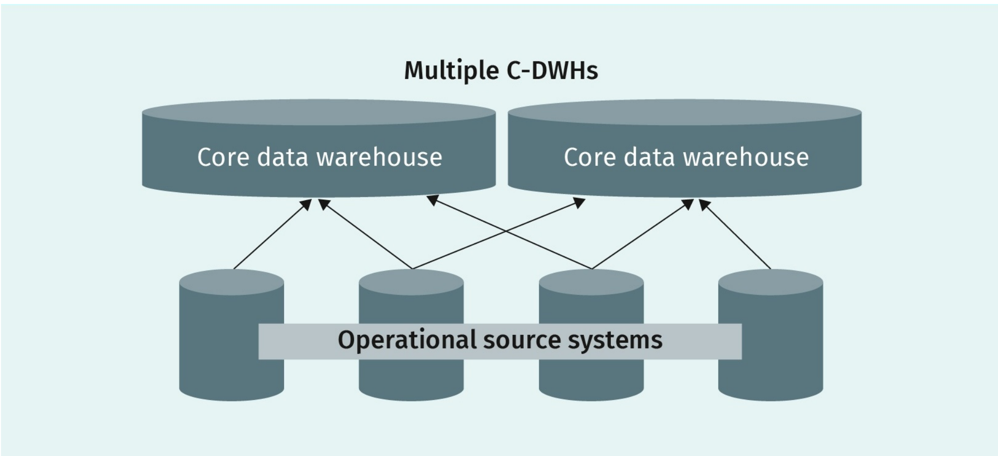
With this monolithic approach, the evaluation function of the core data warehouse is placed in the foreground. However, this approach can have considerable disadvantages (e.g., performance, administration effort) in complex solutions (Kemper et al., 2010).

Multiple C-DWHs

Under certain business conditions, such as in the case of different product or market structures, it is possible to set up several core data warehouses, creating a variant known as **multiple C-DWHs**.

Multiple C-DWHs
This is a variant that is useful for large, division-oriented companies.

Figure 8: Multiple C-DWHs



Source: Kemper et al., 2010, p. 22.

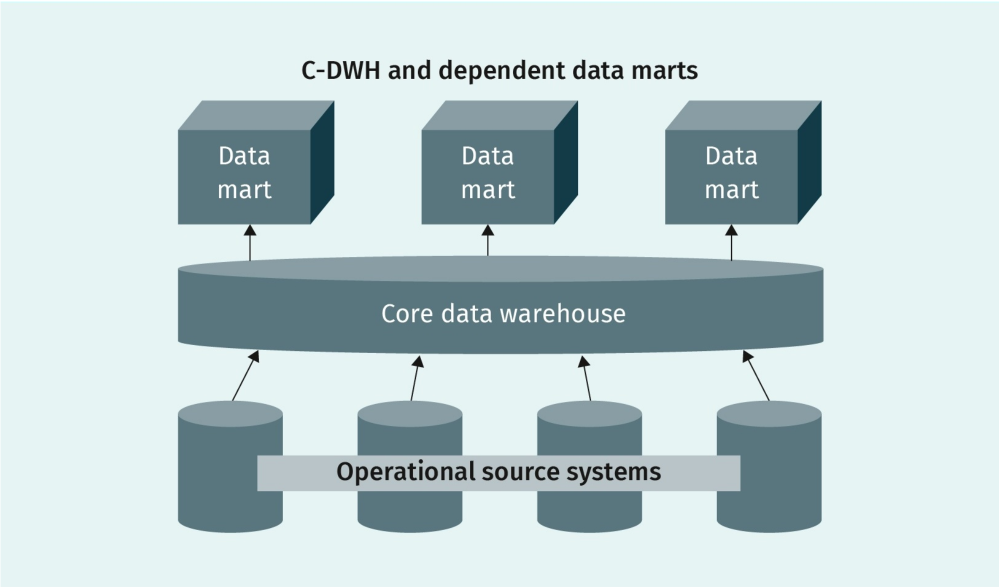
This framework is particularly prevalent in sector-oriented companies and large corporations that produce a diverse range of products and services (Kemper et al., 2010).

C-DWH and Dependent Data Marts

Extending the core data warehouse with data marts is the architecture variant most frequently presented in textbooks. The data marts are supplied with the help of transformation processes and data from the core data warehouse. The following figure shows the **C-DWH and dependent data marts** architectural variant.

C-DWH and dependent data marts
This is the most frequently presented architectural variant in literature. Its major advantage is short response times.

Figure 9: C-DWH and Dependent Data Marts



Source: Kemper et al., 2010, p. 22.

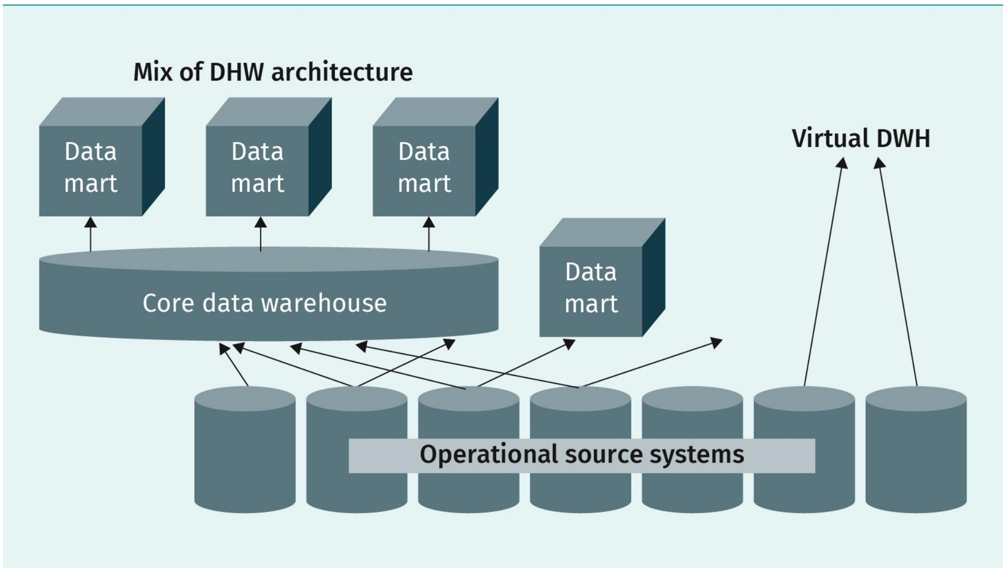
With dependent data marts, data are periodically extracted from the C-DWH and stored in data marts. The extracted data are small, department-specific data extracts from the core data warehouse. By creating these extracts, the data volume of the data marts is considerably smaller. As a result, faster response times can be achieved for queries to this data stock (Kemper et al., 2010). The structure used for dependent data marts is often referred to as hub-spoke architecture (Bauer & Günzel, 2008).

DWH Architecture Mix

In practice, a common architectural variant is the **DWH architecture mix** that consists of C-DWHs, dependent and independent data marts, and direct data access (i.e., virtual DWH with its own data transformation).

DWH architecture mix
A DWH architecture mix consists of C-DWHs, dependent and independent data marts, and direct data access.

Figure 10: Mix of DHW Architecture



Source: Kemper et al., 2010, p. 22.

The approach taken when developing BI architecture can be an iterative, organic process where the architecture evolves over time in keeping with the BI needs of the organization. However, the approach can also be the result of a conscious design process to ensure optimal support for value-adding primary processes and adjunct cross-sectional processes (Kemper et al., 2010).

In practice, there are many data warehouse systems that modify standard architecture variants and adapt them to the respective requirements of the specific organization. For example, several data marts and an operational data store (ODS) can be created to cooperate with the company-wide core data warehouse.

SUMMARY

Using Erich Gutenberg's criteria, application systems can be divided into operational and dispositive systems according to the type of work they support. Operational systems serve to store and manage everyday information for a company. The data warehouse is considered a dispositive system. DWHs are used to extract information from operational data. The corresponding data from these systems can also be classified as operational or dispositive.

From extracting operational data to managerial decision-making, BI (or use of a DWH) can be distinguished according to the process phases: (1) data provision, (2) information generation, storage, distribution, and (3) information access.

BI architecture consists of various components, e.g., source systems, staging area, ODS, C-DWH, data mart, ETL, aggregation, and front end combined together in a number of different architectural variants. In practice, there are a large number of DWH or data mart architecture variants, some of which have emerged directly from the BI field.

UNIT 3

DATA WAREHOUSE

STUDY GOALS

On completion of this unit, you will have learned …

- how data from different operational systems are integrated company-wide.
- what transformation steps are necessary to achieve this.
- what distinguishes a C-DWH from data mart architecture.
- which functions are offered by an operational data store.
- the extent to which metadata can support business intelligence.

3. DATA WAREHOUSE

Introduction

Before business intelligence (BI)-relevant data is made available in the data warehouse, a number of activities need to take place. BI applications require integrated data that is organized in a subject-specific manner, e.g., according to customer, product, or organizational unit. Using the extracting, transforming, and loading (ETL) process, data from operational systems are transformed into data that can be interpreted from a business management perspective. This requires that data stored over long periods of time are made available to management in aggregated form. It also requires that large amounts of data from several operational databases are consolidated and stored in the data warehouse.

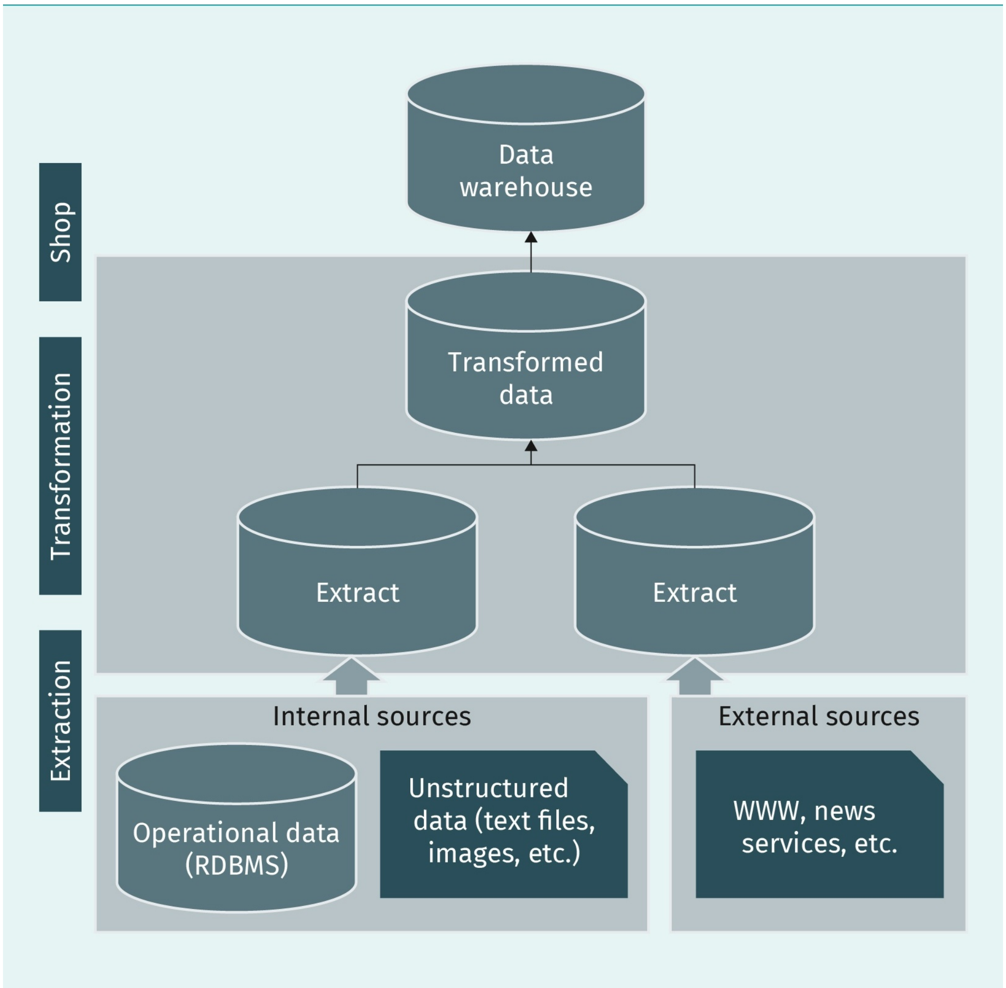
The ETL process cleanses and transforms operational data, which are then stored in the data warehouse for further analysis. After the extraction of operational data from the source systems, the transformation process prepares the data for use. Preparation takes place via four sub-processes: filtering, harmonization, aggregation, and enrichment. Data are then loaded into the evaluation level of the data warehouse (DWH).

3.1 ETL Process

In order to merge and prepare data from several operational data sources, it is converted into management-relevant information via a process of targeted conversion. This is carried out using three steps (extract, transform, and load) which are collectively known as the **ETL process**. This process is illustrated in the following figure.

ETL process
This process is used to convert operational data into management-relevant information.

Figure 11: ETL Process



Source: Glasker, 2017.

Large amounts of data are extracted from source systems, processed according to the requirements of the DWH, and then inserted into, or written to, the DWH.

The process of transferring data from operational sources to the DWH typically takes place at periodic intervals and consists of the following three steps:

1. Extraction of relevant data from various sources
2. Transformation of data into a uniform multidimensional format
3. Loading of data into the data warehouse to be available for analysis

Establishing the ETL process is the most complex step in data warehouse development. The ETL process is of central importance as the creation of a solid DWH is only possible if it contains high-quality data.

In principle, ETL processes can be individually programmed or developed with the help of various tools. Due to the complexity of ETL processes, the use of a tool is recommended in most cases (Kimball & Caserta, 2004). The following sections describe in detail the transformation process in detail

Components of the Transformation Process

Transformation
This consists of filtering, harmonization, aggregation, and enrichment.

The transformation step is the most elaborate and complex part of the integration process. According to Kemper, **transformation** consists of four sub-processes—filtering, harmonization, aggregation, and enrichment—which are outlined below.

Table 2: Sub-Processes of Transformation

Components of the transformation process	
Filtering	Extraction and correction of technical and content defects in the data
Harmonization	Business reconciliation of the filtered data
Aggregation	Aggregation of the filtered and harmonized data
Enrichment	Calculation and storage of key business figures

Source: Kemper et al., 2010, p. 28.

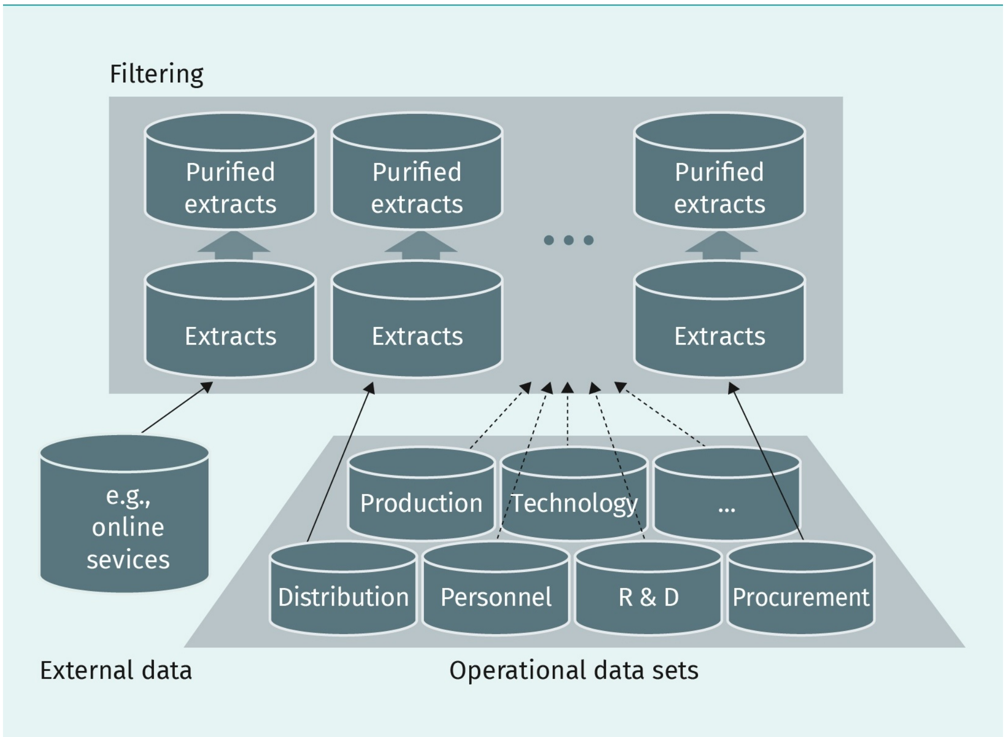
The individual components are described in detail below. The first two transformation steps—filtering and harmonization—are responsible for cleansing and preparing data, e.g., aligning different codes and currencies. Afterwards, the data are, in principle. ready for BI analyses. The next two steps—aggregation and enrichment—summarize data according to topic. Business key figures are also added to the data. The data generated in this way and loaded into the data warehouse are thus already oriented to the needs of individual user groups and their analysis purposes.

Transformation 1: Filtering

Filtering
The filtering sub-process includes the intermediate storage of extracts and data cleansing.

With the help of **filtering**, the data required for the DWH are selected, temporarily stored, and freed from defects. The filtering process is divided into extraction and cleansing. During extraction, the data is placed in the extraction areas (staging areas) specially provided for this purpose. The purpose of cleansing is to remove syntactic and semantic defects. In the following figure, filtering is represented as a sub-process of the transformation process.

Figure 12: Transformation 1: Filtering



Source: Kemper et al., 2010, p. 28.

The purpose of data cleansing is to correct defects and achieve a specified level of data quality. Cleansing is necessary because operational systems do not always contain correct data. There are many reasons for incorrect data, e.g., incorrect entries by users, system errors, system updates.

The types of defects to be remedied can be divided into syntactic (technical) and semantic (content) defects. Syntactic defects are formal errors such as incorrect control characters, alphanumeric values in numeric fields, NULL values in a NOT NULL field, or values outside the value range. Semantic defects are errors of a business nature, such as obviously incorrect sales figures.

Defect classes

In the literature, a distinction is made between first, second, and third class defects. Defects of the first class can be detected and corrected automatically during the extraction process. For second class defects, defect recognition is automatic but the correction must be made manually after the extraction process. Defects of the third class can only be detected and corrected manually.

Table 3: Classification of Defects in the Framework of the Correction

	Class 1	Class 2	Class 3
Adjustment	Automatic detection and correction	Automatic detection and manual correction	Manual detection and manual correction
Syntactic defects	Known format adjustment	Recognizable format incompatibilities	–
Semantic defects	Missing data values	Outlier values/inconsistent value constellations	Undetected semantic errors in source data

Source: Chamoni & Gluchowski, 2015, p. 135.

The basic defects of the first class that are automatically recognized can be corrected using certain algorithms. For example, internal format, control, and special characters can be identified at the syntactic level during extraction and processed in the extracted data using assignment tables (mapping tables). The same applies to semantic errors. If, for example, data from individual stores were omitted when transferring sales data, these can be supplemented using equivalent values, such as monthly planned values or actual values from the previous month.

Defects of the second class can also be detected automatically but must be corrected manually by technicians or business economists. In the case of syntactic defects, an example of these would be syntax variants in the operational data sources that have not yet been taken into account. Once detected and corrected, these can be handled automatically in the future. On a semantic level, automated plausibility checks and value range checks can detect invalid data fields, e.g., by comparing balance sheet and control totals. Depending on the severity of the error, the operational sources may also need to be corrected.

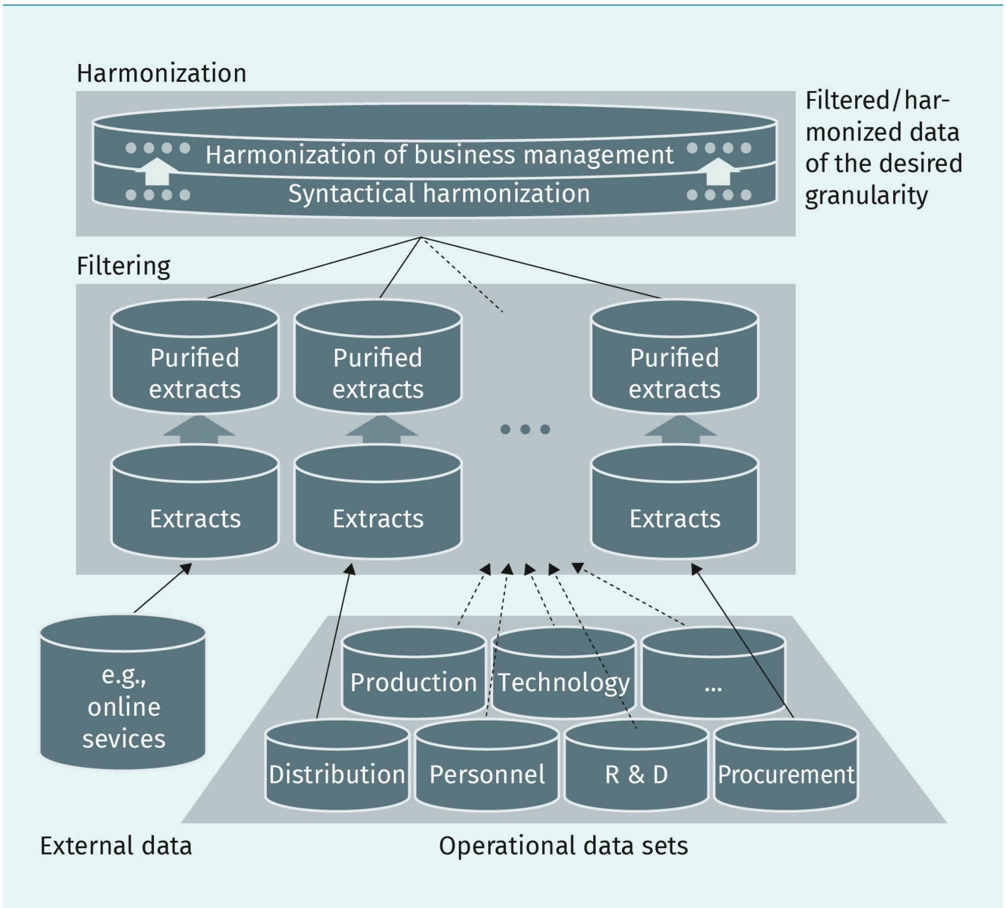
While syntactic defects can always be detected automatically, this does not apply to semantic defects. Defects of the third class only concern semantic errors. These are all defects that cannot be detected by the test procedures for second class errors, i.e., neither plausibility nor value range checks. Rather, these defects can only be identified by business experts. Here too, the operational sources may need to be corrected as well.

Transformation 2: Harmonization

Harmonization
The harmonization subprocess includes the business reconciliation of filtered data.

The second transformation step after filtering deals with the **harmonization** of the data. Harmonization, also known as normalization, refers to the process of reconciling filtered data. Harmonization is necessary if data from different source systems are integrated. In source systems that have grown heterogeneously, different keys or characteristics are often used for the same facts or properties. The classic example is the key for gender, which could be represented differently in three different systems, e.g., male/female, M/W, 0/1. The goal of harmonization is therefore to combine the same facts and characteristics into a common key. Harmonization can also involve transferring different measures into a common measure, e.g., different currencies are converted into a single currency so that they are comparable (Kimball & Caserta, 2004).

Figure 13: Transformation 2: Harmonization



Source: Kemper et al., 2010, p. 32.

In the harmonization sub-process, the filtered and cleansed data are merged. A distinction is made between the following types of syntactic and business harmonization.

Syntactic harmonization

Syntactic harmonization includes key harmonies as well as codes, synonyms, and homonyms, which we will now discuss in more detail. Key harmonies: Harmonization must include the dissolution of key harmonies. In principle, a common key is necessary when data from several databases are merged. The problem is usually solved with the help of a mapping table, which generates a new, artificial primary key, e.g., for each customer.

The primary keys of the operational systems are then carried along as foreign keys, so that evaluations can be carried out on them. The keys of data records must be unique within the basic database, DWH, and data mart. The keys available in source systems do not usually fulfill this requirement due to their heterogeneity and also the distribution of the data. During the transformation phase, global, unique keys must therefore be assigned. These global keys are called “surrogate keys.” Modern ETL tools have standardized transformations that generate unique surrogate keys.

The mapping of local keys to global surrogate keys must be documented in order to be able to react flexibly to changes. Besides the standardization of data, surrogate keys play an important role in historicization (Kimball & Caserta, 2004). For example, take two data sources (e.g., CRM, ERP) that contain customer data that must be integrated into a table. - Both source data records have a customer key. A global key for customer data must therefore be introduced for the target database. To do this, the keys of the source tables are removed during transformation and replaced by surrogates.

In addition to key harmonies, codes, synonyms, and homonyms must also be resolved. Here are some examples:

- Codes. Individual data sets can be coded differently. For example, attributes such as gender can be coded as M/W in data source one and as 0/1 in data source two.
- Synonyms. Different attribute names can have the same meaning. For example, in data source one, the attribute "personnel" may be provided for the name of company employees but in data source two, it may be "employees".
- Homonyms. The same attribute names can have different meanings. For example, in data source one, "partner" can mean the name of customers, while in data source two, "partner" can refer to the name of suppliers.

In all three cases, the data must be harmonized. In the first case, the attribute value must be uniformly set, e.g., to 0/1 values; in the second case, a common attribute name must be chosen; in the third case, a different attribute name must be chosen for the two categories. Mapping tables are usually utilized for the matching process, which merge the filtered data into subject-oriented data collections via name and code matching.

Semantic harmonization

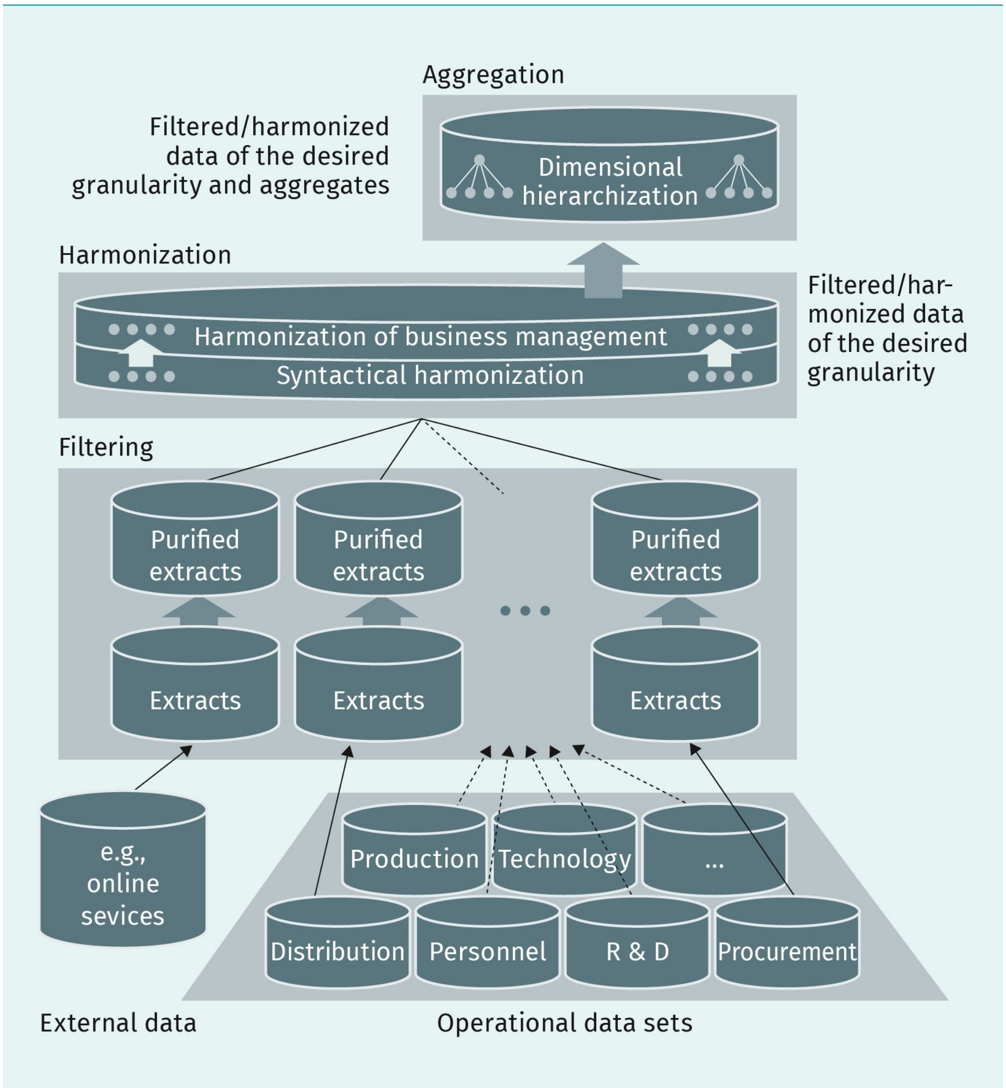
In addition to syntactic alignment, the standardization of business terms is also carried out. This is also known as semantic harmonization. The normalization of business terms is not so much a technical problem as a business and organizational one. The operational data (e.g., currency) must be converted into uniform values, i.e., monetary values of different currencies must exist in a uniform currency system. For the corresponding activities, transformation rules can be implemented. After completion of the harmonization sub-process, cleansed and consistent data are available in the data warehouse for analysis purposes.

Transformation 3: Aggregation

Aggregation
The aggregation sub-process involves the compression of filtered and harmonized data.

With the help of the **aggregation** sub-process, filtered and harmonized data are condensed and converted to the desired granularity.

Figure 14: Transformation 3: Aggregation



Source: Kemper et al., 2010, p. 37.

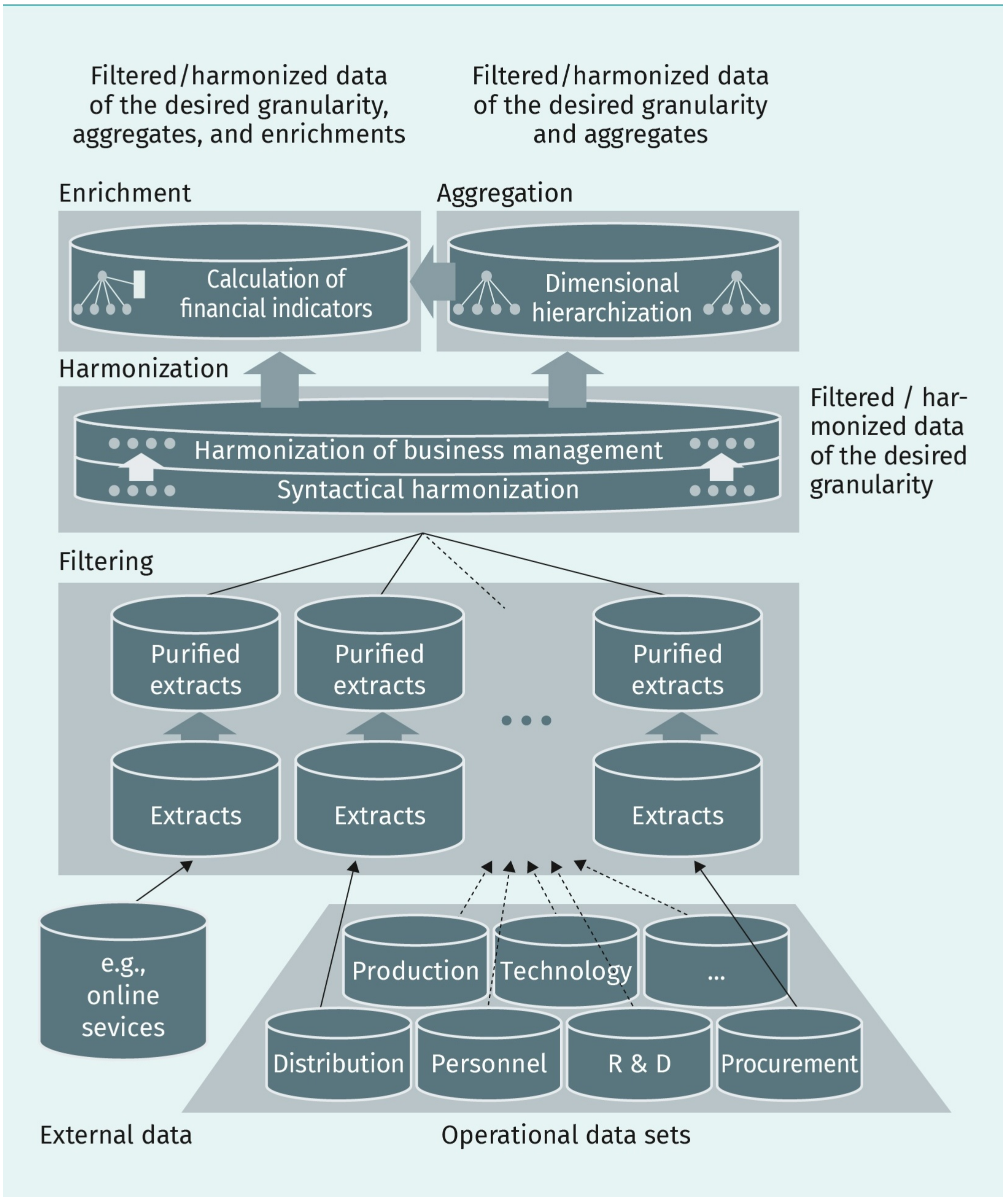
Let's consider the aggregation sub-process in practice. To create daily updated data for product and customer groups, all individual data must be summarized via aggregation algorithms to produce daily product and customer group specific values. In addition, running totals are performed for business key figures. The aim of aggregation is to generate total values that are stored in the data warehouse in pre-calculated form for later use.

Transformation 4: Enrichment

The creation and storage of key business figures from filtered and harmonized data is called enrichment. The **enrichment** sub-process is the final step in the transformation process.

Enrichment
The enrichment sub-process involves the creation of business key figures after harmonization or aggregation.

Figure 15: Transformation 4: Enrichment



Source: Kemper et al., 2010, p. 38.

As previously described, the existing data is enriched with business indicators. Calculations are performed and results are added to the remaining data. In particular, key figures that are relevant for several users are stored. For example, weekly contribution margins at-product level or annual contribution margins at store level can be calculated and integrated. The former may be of interest to the product manager, while the latter is of interest to store managers and executive management. There are several advantages to including these key figures in the dataset. Due to the pre-calculation of these figures, queries can be performed more efficiently. In addition, the pre-calculated values are consistent due to the one-time calculation. Furthermore, these figures facilitate coordinated business management.

The main activities of the ETL process are completed through the extraction process but particularly the transformation process. The transformation sub-processes are carried out in the staging area. After transformation, the data are written to the target system during the loading phase.

3.2 DWH and Data-Mart Concepts

The DWH in a narrower sense includes data storage. Individual components of the DWH including the staging area, basic database, data mart, ODS, and metadata, that all facilitate data storage are now described in detail.

Staging Area

According to Inmon (2005), the staging area or landing zone is a workspace in which data are temporarily stored. Extraction into the **staging area** usually takes place periodically. Transformations are performed within the staging area. In principle, the data is deleted from the staging area after it has been loaded into downstream systems. The purpose of the staging area is to relieve downstream systems (e.g., the basic database). This separate work area is of particular importance, especially when processing large amounts of data and performing complex transformations (Inmon, 2005).

Staging area
This is where extracts are stored temporarily in order to relieve downstream systems.

C-DWH

Functionality

The **basic database** (known as C-DWH) is a data store or repository located between the staging area and the evaluation database. The basis database differs from the staging area in particular in the way the data are stored. The database provides detailed, historicized, consistent, and normalized data for downstream systems. The data are transferred from the staging area when they are in a specific state, whereupon they are integrated into the basic database and stored in an adjusted form. With regard to the evaluation database, the basic database is primarily defined by the data model. In the basic database, data are stored in a normalized and query neutral state. In addition, the data are stored at the smallest required granularity. Depending on requirements, data are also historicized, that is, they are kept track of over time (Bauer & Günzel, 2008). The C-DWH performs the following functions:

Basic database
This contains integrated data for downstream systems.

- Collection and integration. This refers to the collection of company-wide data required for later analysis.
- Distribution. This refers to the data supply of downstream systems (e.g., data marts).
- Quality assurance. This refers to how transformed data ensures the syntactic and semantic coherence of the dispositive database.

Update strategy

The core data warehouse is updated according to the respective requirements of the BI system. A distinction is made between the following types of updates:

- Updates based on changes. The accumulated changes from the source systems are collected. The data are loaded into the C-DWH as soon as a defined number of changes are reached.
- Updates at periodic intervals. Updating is carried out according to the requirements of the respective application, e.g., hourly, daily, weekly, or monthly.
- Real-time updates. Data can be loaded into the C-DWH in a transaction-synchronous manner. This complex procedure usually uses special loading systems that follow event-controlled logic. Data are loaded into the data warehouse in a "push mode."

Data Mart

Definition

Evaluation database
This contains sections of the C-DWH.

In order to simplify data handling in the case of complex C-DWHs, **evaluation databases** are introduced. Data marts are sections of a C-DWH, smaller data pools for applications that serve specific user groups such as certain departments or defined areas of responsibility. The data for data marts are usually extracted from the core data warehouse into small, manageable units using special transformation processes, e.g., all relevant data for a region or a specific product group. The advantage is that the complete data basis of a company does not have to be mapped, rather only the data needed for queries related to a specific area or department. For example, the sales department is usually only interested in key figures such as sales figures, turnover, and commissions, while the production department is interested in production quantities or production times.

The evaluation database thus forms the basis for downstream analysis tools. In the data mart, data are stored fully integrated and cleansed. The data are stored in an analysis-oriented manner. Technically, evaluation databases are usually based on relational databases. Multidimensional storage models are used less frequently.

Basically, the data are structured in a multidimensional model. The data required for analyses are extracted from the basic database by ETL processes and loaded into the evaluation database. Since the basic database already contains integrated and cleansed data, the data only need to be transformed into the target schema and possibly aggregated before loading.

In practice, several evaluation databases (data marts) are often used. Data are divided up according to analysis requirements or organizational units. Security or data protection aspects may make it necessary to distribute the data over several data marts. If a concretely-implemented architecture does not operate a basic database (C-DWH), integration and cleansing must be carried out during the transfer to the evaluation database (Bauer & Günzel, 2008).

Characteristics

The following table summarizes the characteristics of data marts and C-DWH.

Table 4: Data Marts and Core Data Warehouse

Characteristics	Data mart	Core data warehouse
Business management goal	Efficient support for the decision makers of a department, focused solely on their analysis requirements	Efficient management support through strategic, tactical, and operational information available to all decision makers in a company
Alignment	Departmental	Central, company-wide
Granularity of the data	Mostly highly aggregated data	Lowest level of detail
Semantic data model	Semantic model is fixed to premodeled analysis requirements	Semantic model is also open for future analysis requirements
Modeling conventions	• Heterogeneous (proprietary data marts, each department has its own conventions); • Uniform (derived data marts, conventions of the core data warehouse are adopted)	Uniform
OLAP technology used	M-OLAP (proprietary data marts); R-OLAP or H-OLAP (derived data marts)	R-OLAP
Direct access by end users	Usually possible	Often not allowed; central operation of the C-DWH by the IT department; serves as source data system for data marts
Degrees of freedom of the analyses	Rather low (user cannot see beyond departmental boundaries)	Flexible; all accessible (secure) information can be included in analyses
Influence of external data sources	Mostly not included; if so, then only specific extracts	High; all available external data sources will be integrated to improve the quality of analysis
Data volume	Low to moderate	Moderate to very extensive (up to the petabyte range)

Source: Kurz, 1999, p. 110.

In conclusion, evaluation databases are generally introduced to simplify handling with complex C-DWH.

3.3 ODS and Meta-Data

ODS Definition and Characteristics

Operational data store (ODS)
This is a preliminary stage of the DWH that contains current transaction data for evaluation purposes.

In newer approaches to data warehouse design, an additional data pool is often integrated, known as an **operational data store (ODS)**. From an architectural point of view, the ODS can basically be regarded as a preliminary stage of a DWH. The ODS contains current transaction-based data that originates from various operational source systems. The data are provided for application and evaluation services. Usually, the data are extracted from the core data warehouse using additional transformation processes. A very small and up-to-date section of decision-relevant data are transferred to the ODS. These data are often already adapted to the requirements of the analysis systems. The ODS is defined by the following features:

- subject-oriented,
- integrated,
- time referenced,
- volatile, and
- high level of detail.

The ODS is designed from a decision-oriented perspective. For example, dimensions could be products and regions (subject-oriented). Company-wide data are transferred uniformly to the ODS using appropriate transformations. The transformation process in the ODS primarily involves filtering and harmonization (integrated).

In principle, no historicization is carried out in the ODS. For this reason, no period-related evaluations are possible. For recovery reasons, however, data are retained over a period of several days or weeks (time referenced). However, regular updating is carried out and this data are overwritten (volatile).

The data are maintained in a high level of detail, i.e., not aggregated, since analyses in the ODS are usually operation-related. Detailed storage means that data are stored at the transaction level (high level of detail).

Metadata

For the analysis of data to occur, it is important that users know what lays behind respective data fields. The information about these data is provided in the form of metadata.

Differentiation

Metadata contain information about the data stored in the data warehouse including how they have been processed. Metadata supports the construction, administration, and operation of DWH systems.

Metadata can be divided into passive and active metadata, which differ in a variety of ways, including their use. The key features of each type of metadata are as follows:

- **Passive metadata** primarily document the data on which they are based and their relationship to the environment. They serve to define and store information about structure, development process, and data usage. Users of passive metadata are all users active in the BI environment, i.e., users, administrators, developers (Kemper et al., 2010).
- **Active metadata** are metadata upon which methods are executed, i.e., metadata used for operational purposes. Transformation rules for ETL operations can be understood as active metadata. Active metadata can be interpreted at runtime and are used to influence transformation and analysis processes (Kemper et al., 2010).

Passive metadata
These document data and their relationship to the environment.

Active metadata
These represent methods that are executed on data.

The following distinctions can be made between technical and business metadata:

- **Technical metadata** focus on the first layer of transformation (filtering). They describe where data is sourced from and the structure of data in its native environment.
- **Business metadata** focus on the subsequent layers of transformation (harmonization, aggregation, and enrichment) and authorization management (Kemper et al., 2010). They provide additional information such as keywords and notes about the meta objects.

Technical metadata
These focus on filtering.

Business metadata
These focus on harmonization, aggregation, enrichment, and authorization management.

Advantages

With the help of metadata, the efficient design of development and operating processes can be ensured and the effectiveness of BI systems can be increased. In addition to their task of providing information, metadata also serve the data warehouse manager as a control element. For example, fully executable specifications of data processing steps are stored as metadata and interpreted and executed by the corresponding tool at the time of execution. The advantages of metadata, especially for development and operation, include:

- adaptation of source systems,
- harmonization of data from heterogeneous source systems,
- maintenance and reuse,
- authorization management,
- data quality, and
- understanding of terms.

In the first transformation layer (filtering), operational data are transferred to the DWH. The extraction and cleansing processes performed can be documented with the help of metadata. This makes it easier to modify or extend processes (adaptation of source systems).

In the second transformation layer (harmonization), data are integrated syntactically and semantically. Transformation activities can be facilitated by information about the structure and meaning of the source systems. This ensures efficient further development of the BI system (harmonization of data from heterogeneous source systems).

By storing metadata, the maintenance and further development of BI systems is simplified. Business and technical changes can be carried out quickly and without contradiction due to consistent metadata. For example, the reusability of data models and transformation processes can be supported (maintenance and reuse).

Authorization management is a central component of planning data management. Metadata are used to describe user roles that allow consistent administration access. This enables simple administration of relationships between BI users, applications, and data authorizations (authorization management).

In principle, metadata can be provided throughout the entire transformation process to create the highest possible transparency for the user. Responsibilities, quality of source systems, harmonization processes, and enrichments can all be documented in metadata. In this way, data quality can be ensured, particularly consistency, temporal proximity, accuracy, and completeness (data quality). Key business figures can be described in terms of their designation, differentiation, origin, and use with the help of metadata. The metadata for dispositive data storage thus represents a “single point of truth” in the company context (understanding of terms).

Architecture variants

The complex, individualized approaches taken by companies regarding their BI systems use a multitude of special software components. A basic distinction can be made between end-to-end and best-of-breed approaches.

In end-to-end approaches, software manufacturers offer tools that are coordinated with one another. The tools support all development and operating processes, from ETL design to report generation or portal integration of reports. In contrast, software vendors of best-of-breed solutions offer specialized tools that are used to develop powerful components of a company-specific BI concept. The consistent metadata management of all components of an integrated BI concept is a complex undertaking.

The three architecture variants are described below:

Central metadata management
This is where the metadata of all components and authorization structures are stored in a database.

1. **Central metadata management**. Here, a central database is used for metadata management. Metadata of all components and authorization structures are stored in the database. This solution is used especially for end-to-end approaches. In practice, there are few companies using central metadata management.

Table 5: Central Metadata Management

Advantages	Disadvantages
Redundancy-free, consistent metadata management	Dependence on the central data storage component
Global access to all metadata	Complex, central maintenance of component-specific metadata

Advantages	Disadvantages
Renunciation of exchange mechanisms (metadata)	Poor performance of large, centralized solutions

Source: Gluchowski et al., 2008, p. 156.

2. **Decentralized metadata management**. This is the opposite approach to the previous concept. In principle, all components have their own metadata repository and communicate with each other to exchange metadata. In practice, BI concepts are usually implemented with decentralized metadata management.

Decentralized metadata management
This is where all components have their own metadata repository.

Table 6: Decentralized Metadata Management

Advantages	Disadvantages
Autonomy of the applications	Various interfaces
Fast, local access	Redundant data management

Source: Gluchowski et al., 2008, p. 156.

3. **Federated metadata management**. This is a combination of the previously presented approaches. The components each manage their own metadata. In addition, there is a central repository in which shared metadata are managed. With the help of a standardized interface, metadata are exchanged between the individual components and the central repository. Advantages of utilizing a federated metadata management approach are:
   - uniform presentation of shared metadata,
   - autonomy of the local repository,
   - reduced number of interfaces between repositories, and
   - controlled redundancy.

Federated metadata management
This is where, in addition to the individual metadata, there is a central repository with shared metadata.

Using standard interfaces, you can enable the exchange of metadata between BI tools and the metadata repository.

Authorization Structures

The regulation of access authorizations is mostly done within the individual systems. In contrast, BI concepts with integrated data storage allow central authorization management. This eliminates the need for the separate authorization management of different systems, such as ETL, C-DWH, and analysis systems.

In practice, role-based access controls are increasingly used due to their high degree of flexibility. Here, users or user groups receive access based on their corresponding roles. In roles-based access, rights that are necessary to fulfill defined tasks are summarized and functions available to users according to the need-to-know principle. By assigning data views to the roles, it is ensured that users are only allowed to access certain data fields, e.g., sales of their own store.

Administration Interfaces

With the help of administration interfaces, technical and business management specialists can maintain all areas of dispositive data management. The relevant people can generate, modify, and delete

- transformation rules,
- dispositive data, and
- role-based access authorizations.

Technical administration interface
This serves to modify data and the first transformation layer (filtering).

A distinction can be made between the technical and the business administration interface. You can use the **technical administration interface** to modify data and the first transformation layer (filtering). In addition to data manipulation, this also includes processing all structures for extracting and cleansing data.

Business administration interface
This serves to modify the subsequent transformation layers (harmonization, aggregation, enrichment) and authorization structures.

The **business administration interface**, on the other hand, is used to maintain the additional three transformation layers (harmonization, aggregation, enrichment) and manage the authorization system. For example, business specialists use the business administration interface to intuitively edit syntactic and semantic harmonization processes, hierarchy trees, aggregations, and key business figures. Within the framework of authorization management, the functional administration interface is used to intuitively assign roles and employees or employee groups.

SUMMARY

The ETL process cleanses and transforms operational data, which are then made available in the data warehouse for further analysis. The preparation takes place via the four sub-processes: filtering, harmonization, aggregation, and enrichment.

The DWH in a narrower sense involves data storage. Individual components of the DWH are the staging area, basic database, data mart, ODS, and metadata.

The staging area is a work area in which data are temporarily stored in order to relieve downstream systems. The basic database is a data storage mechanism located between the staging area and the evaluation database. The C-DWH provides company-wide, integrated data from downstream systems. Data marts are sections of a C-DWH that provide data for specific user groups (e.g., departments).

The ODS contains current transaction-based data that originates from various operational source systems. The classic characteristics of an ODS are subject-orientation, integration, time reference, volatility, and a high level of detail. Metadata contain information about the data stored in the data warehouse and how they are processed. They provide sup-

port for setting up, managing, and operating DWH systems. A distinction can be made between active and passive as well as technical and business metadata.

UNIT 4

MODELING MULTIDIMENSIONAL DATASPACES

STUDY GOALS

On completion of this unit, you will have learned …

- what basic modeling techniques exist.
- which analysis possibilities are offered by OLAP cubes.
- how multidimensional models are physically stored.
- which options are available for historicizing dimensions.

4. MODELING MULTIDIMENSIONAL DATASPACES

Introduction

A data warehouse (DWH) system supports decision makers throughout a company in their work. Depending on their area of responsibility, they usually have different interests in the data. For example, an employee from controlling would typically be interested in key business figures, whereas a doctor might be interested in the success of a particular therapy. For this reason, it is necessary to provide flexible views of data.

For DWH applications, a multidimensional data model is more flexible than a relational data model. A relational data model is where data is located in two dimensions whereas a multidimensional model is where data is located in across multiple dimensions. In multidimensional models, enterprise data is arranged in a multidimensional data space.

4.1 Data Modeling

Relational and Multidimensional Models

In addition to the relational data model, multidimensional data spaces play a particularly important role in business intelligence. Star and snowflake schemas in particular allow performance-oriented modeling of multidimensional spaces.

In principle, data models can be semantically, logically, or physically oriented. Physical data models are technically oriented and specify how data are physically stored. Logical and semantic views are more interesting for users. Logical data models describe all data on a logical level, regardless of how it is stored. Semantic models, on the other hand, are the closest to reality. They depict the data on a completely technology-neutral level.

One of the best known semantic data models is the entity relationship model (ERM). It was developed in the seventies by Peter Chen and has been modified and extended over time (1977). ERM is used in the conceptual phase of application development to structure the communication between users and developers as well as in the implementation phase as a basis for database design. With ERM, operational data structures can be modeled well. Even multidimensional data structures, which are presented below, can be modeled with it.

Redundancies and Normal Forms

In practice, relational databases are not always created using a model, such as ERM. The major disadvantage resulting from this is that redundant information can be stored. Redundancy refers to the duplicated storage of identical attribute values for a single object characteristic. For example, a redundancy would occur if an employee's name was stored in an employee database together with both their department and their department number. Redundancies compromise the consistency of the database and can lead to anomalies. For example, if the department number changes, several tuples must be changed at the same time. This is not only costly, but also carries the risk of inconsistency (update anomaly). If all employees leave a department, the information about which department number is assigned to the department is also lost (deletion anomaly). To prevent anomalies, a number of rules and principles apply. A widely used procedure to avoid redundancies and inconsistencies is normalization.

Primary Key and First Normal Form

Normalized data are data that are free of redundancies and inconsistencies and can therefore be managed more efficiently. However, a strong normalization can also have a negative effect on performance. The theory of normal forms goes back to Edgar F. Codd, inventor of the relational database. He developed mathematical rules that transform relational databases into data structures that are free of redundancy or at least minimize redundancy. In essence, this process consists of three normalization steps.

The definition of the first normal form (1NF) is as follows: a table row may only contain one attribute value. The technical definition is that the attribute value must be atomic. If repeating groups occur in a table, a separate table line must be created for each value of the repeating group.

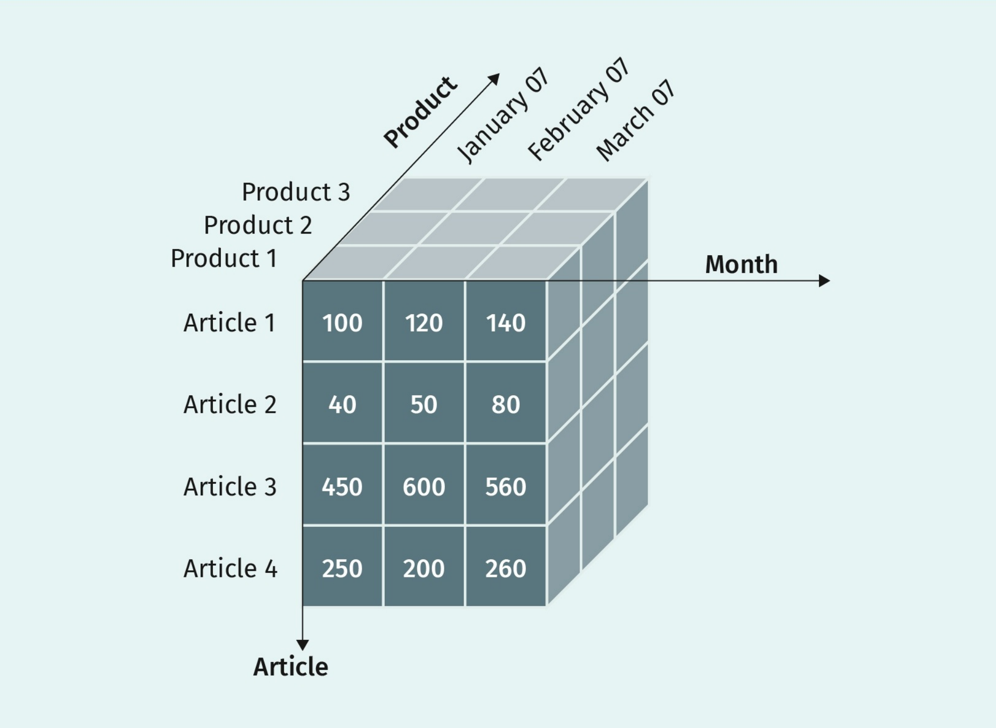
Second and Third Normal Form

The definition of the second normal form is as follows: a relation is in the second normal form if it is in the first normal form and all non-key attributes depend functionally on the entire key. This means that it must be impossible for key parts to identify certain attributes of the relation. A new table must therefore be created from those attributes that only depend on a part of the compound primary key. A relation of the third normal form exists if the second normal form exists and no functional dependencies exist between non-key attributes.

4.2 OLAP Cubes

The focus of the online analytical processing (OLAP) cube model is usually on business ratios as carriers of quantitative information, which are described by corresponding sets of dimensions. Each dimension is explained by a set of attributes. For example, the dimension "time" is described by the attributes "year", "month", "quarter", "week", and "day". The attributes can be related to each other within a dimension and form a relationship hierarchy. The hierarchy facilitates both the aggregation of data and navigation through the data. In common parlance, the resulting multidimensional data space is also referred to as an OLAP cube, which is illustrated below:

Figure 16: Cube and Dimensions



Source: Gluchowski et al., 2008, p. 156.

Facts
These are corporate key figures, such as turnover, costs, and unit numbers.

Dimensions
These describe key business figures, e.g., by specifying time, region, or product.

A cube consists of dimensions (edges of the cube) and key figures or facts (values at the coordinates inside the cube). **Facts** are operation-related, quantitative variables such as sales, costs, and unit numbers. **Dimensions** are descriptors for key figures such as time, customer, and product. The elements of a dimension can be grouped into hierarchies based on functional dependencies. For example, months can be grouped into quarters along the time dimension or products into product groups along the product dimension. Key figures can be consolidated across the dimension hierarchy (e.g., calculated totals). There are several operations available for navigating through a multidimensional data model: roll-up and drill-down, slice and dice, drill-across, pivoting, and rotation.

Roll-Up and Drill-Down

Users can use the roll-up and drill-down operations to navigate along the relationship hierarchy.

Figure 17: Roll-Up & Drill-Down

	Product A	Product B	Product C	Product D
1st quarter	140,000	100,000	200,000	120,000

Drill-down ↓ Roll-up ↑

	Product A	Product B	Product C	Product D
January	40,000	30,000	70,000	40,000
February	45,000	35,000	60,000	35,000
March	55,000	35,000	70,000	45,000

Source: Kemper et al., 2010, p. 103.

The **drill-down** operation increases the level of detail, e.g., the user has the possibility to view the monthly key figures for individual products at a daily level. The **roll-up** operation is the inverse function to the drill-down operation, i.e., the level of detail is reduced so that data are viewed in aggregated form. For example, the user can view the monthly key figures at a quarterly level.

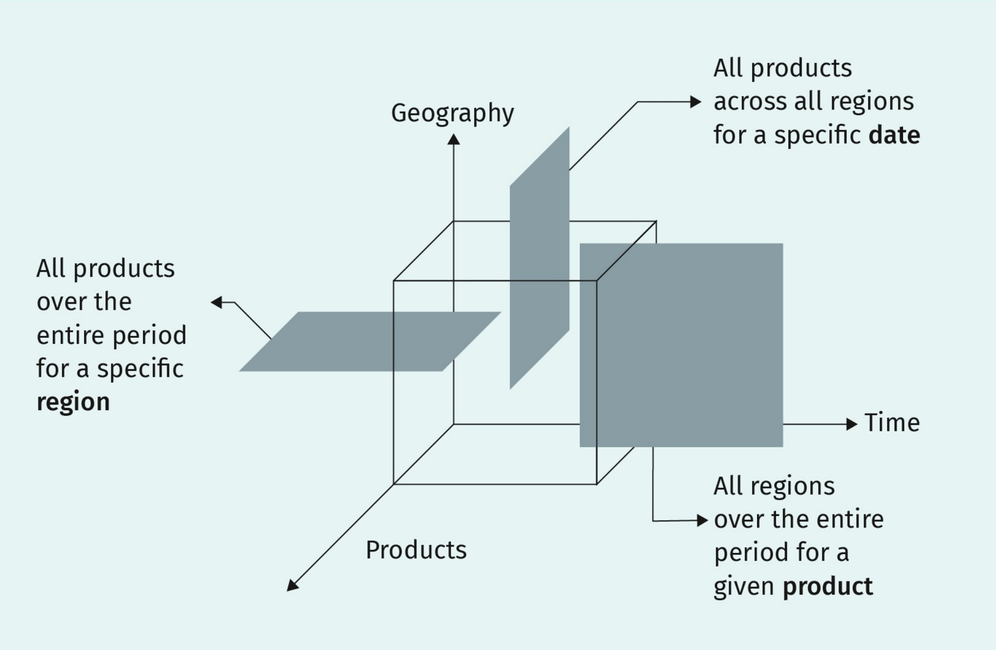
Drill-down
Using drill-down, you can jump to a deeper level of detail in a report.

Roll-up
Using roll-up, you can jump to a higher level of aggregation.

Slice and Dice

Individual views of the data model are generated using the slice and dice operation. An example of the slice operation is provided below.

Figure 18: Slice Operator

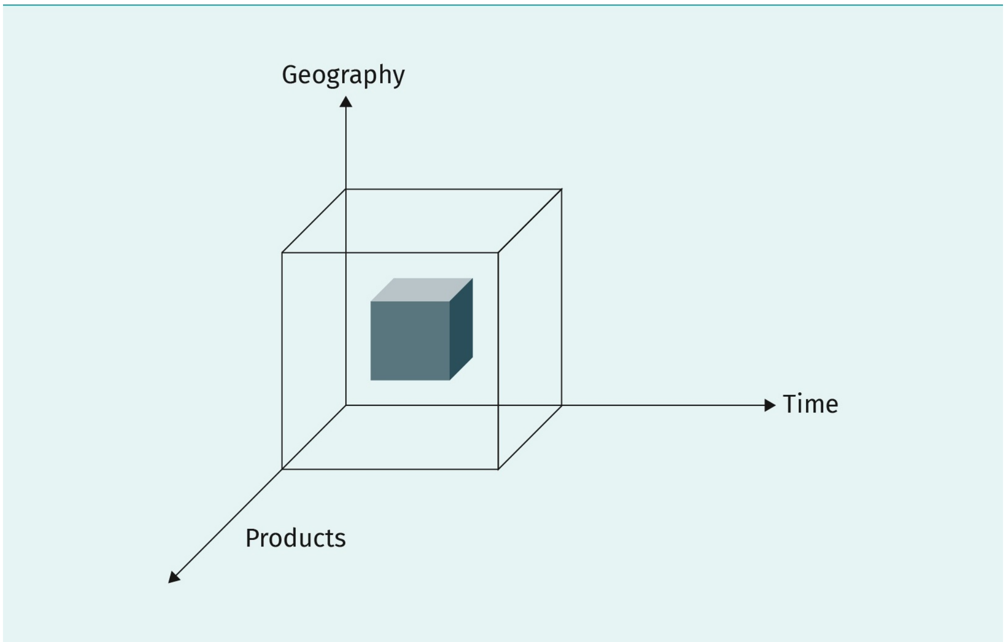


Source: Kemper et al., 2010, p. 105.

Slice
The slice operation allows you to view data across a single plane.

The **slice** operation cuts individual slices out of the data cube and thus allows a single plane of the cube to be viewed. For example, using this operation, the user can limit their analysis to viewing the turnover for a single region by “cutting out” the slice of the region.

Figure 19: Dice Operator



Source: Kemper et al., 2010, p. 105.

The **dice** operation cuts a partial cube out of the total cube. The operation allows the user to view the key figures for a concrete combination of dimension elements. Using the **drill-across** operation, you can switch between different dice.

Dice
The dice operation allows you to view individual sub-cubes.

Drill-across
The drill-across operation allows you to switch to other cubes.

Pivoting

Pivoting allows you to virtually rotate data cubes to view data from different perspectives. The order of the displayed dimensions is reversed when pivoting occurs.

Pivoting
This involves rotating the cube to view data from different perspectives.

4.3 Physical Storage Concepts

There are a number of different concepts that exist for the storage of data in a DWH. The use of the multidimensional data model does not necessarily require multidimensional data management. The **relational storage model (ROLAP)**, an alternative to the **multidimensional storage model (MOLAP)**, is actually used more often. Another option is the **hybrid storage model (HOLAP)**, which combines multidimensional and relational storage.

Relational storage model (ROLAP)
This is where multidimensional data models are converted into relational storage concepts.

Multidimensional storage model (MOLAP)
This is where the physical storage of data takes place in a multidimensional database management system.

Hybrid storage model (HOLAP)
This model combines the strengths of the relational and multidimensional concepts.

Relational Storage (ROLAP)

The most widely used DWH data storage model is the ROLAP. In this model, data from the multidimensional data model are stored in two-dimensional tables. However, the multidimensional interface must be retained for the overall system. Thus, the data from the multidimensional data model are mapped onto a relational database system (Bauer & Günzel, 2008).

Multidimensional Storage (MOLAP)

With the MOLAP, the physical storage of data takes place in a multidimensional database management system. Storage is made possible by transferring the model elements directly into physical objects. The data elements are stored in arrays.

Due to the multidimensional structure of data storage, a very high query speed can be achieved. However, the problem with multidimensional database management systems is that these systems cannot manage very large data sets. For this reason, a comprehensive data warehouse with low granularity should be implemented using relational storage, whereas a small data warehouse (e.g., data mart), which is formed using already aggregated values, should be implemented using multidimensional storage.

Hybrid Storage (HOLAP)

The HOLAP attempts to combine the strengths of the relational and multidimensional concepts. Both a relational and a multidimensional database are used for storage. The relational database stores data that are detailed and available in large quantities, while aggregated data are stored in the multidimensional database. Data access is achieved by means of a multidimensional query tool. By using both technologies, it is possible to exploit the respective advantages and overcome the inherent disadvantages of each of the types of data storage. However, comprehensive knowledge of both storage models as well as additional implementation efforts is necessary for a hybrid storage model to be effective.

4.4 Star Schema and Snowflake Schema

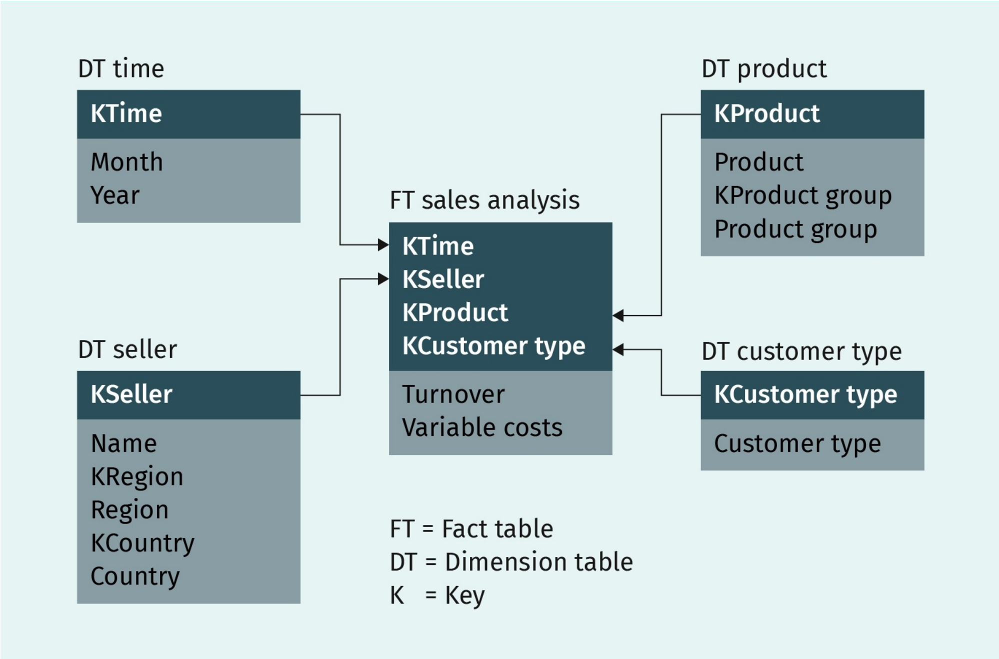
In the physical implementation of a multidimensional data model, a distinction can be made between two different database structures: star schema and snowflake schema.

Star Schema

Star schema
This describes how dimensions are arranged in a star shape around the fact table.

The **star schema** consists of a fact table in which the keys for a data cube are managed.

Figure 20: Star Schema



Source: Kemper et al., 2010, p. 68.

The figure shows that the dimension tables are arranged in a star shape around the fact table. There is a separate table for each dimension. Relationships only exist between the fact table and the dimension table, not between the dimension tables. The link between the fact table and the dimension table is achieved by including the primary key of the fact table as a foreign key in the primary keys of the dimension tables.

Snowflake Schema

Another multidimensional schema is the **snowflake schema**. In this schema, a fact table is surrounded by dimension tables, which are in turn surrounded by dimension tables; in this manner, hierarchies are split into additional dimension tables rather than contained to a single dimension table. In contrast to the star schema, the snowflake schema is normalized with regard to functional dependencies, which results in a large number of tables that are joined together in a query (SQL join). As data are split into different dimension tables, the additional joins mean that data fetching can take longer, resulting in reduced performance for queries. For this reason, the star schema is used more often. A star schema can be created from a snowflake schema by de-normalizing the tables belonging to each dimension (Bauer & Günzel, 2008).

Snowflake schema
This is normalized with respect to functional dependencies.

The modeling of multidimensional data is done within the relational data model with star and snowflake models. The schemas allow a performance-optimized modeling of multidimensional data spaces. Thereby, the strict theoretical requirements for relational systems are (knowingly) violated in some cases.

4.5 Historicization

In a multidimensional data model, a distinction is made between fact tables and dimension tables. Facts are usually historicized using the foreign key relationships of the data to the time dimension. The historicization of dimensions, on the other hand, is more complex. A common challenge in the DWH context—the problem of historicizing dimensions—is described in detail below.

In the data warehouse context, changes to dimensions must be taken into account in addition to the historicization of facts. Changes of dimensions occur, for example, when products are renamed. Data warehouse expert Ralph Kimball introduced the term **"slowly changing dimensions (SCD)"** to describe the historicization of dimensions. The term slowly changing dimensions acknowledges that changes to dimensions occur less frequently, less consistently, and more unexpectedly compared to changes in facts.

Slowly changing dimensions (SCD)
These describe concepts for the historicization of dimensions.

Slowly changing dimensions can be processed in three ways according to Kimball. The following example illustrates the three types of responses to SCD in their treatment of data. A product table is given, which describes a dimension in the data warehouse. The key attributes are underlined in the representations.

Table 7: Dimension Table Products

Product number	Product name	Product group
1	Phone1	Corded tel.
2	Phone2	Cordless tel.
3	Phone3	Cordless tel.

Source: Glasker, 2017.

From the corresponding data sources, the following data set is extracted on 02.06.2015, where the product group of a product has changed.

Table 8: Extracted Data Products

Product number	Product name	Product group
3	Phone3	Cordless ISDN tel.

Source: Glasker, 2017.

We can see in this table that the product group for Product No.3 has been modified to "Cordless ISDN Tel." According to Kimball, there are three basic responses to documenting this change in the DWH which he terms SCD type 1, SCD type 2, and SCD type 3.

SCD Type 1

With **SCD type 1**, the corresponding data record is overwritten. Information is lost as a result. When analyzing data over a period of time, it is no longer possible to trace the change in the dimension. In the case of a dimension historicization according to SCD type 1, the previous data record “Phone3” is overwritten by the new data record. In this case, the information about the previous product group is lost.

SCD type 1
This overwrites the data record.

Table 9: Historicization SCD Type 1

Product number	Product name	Product group
1	Phone1	Corded tel.
2	Phone2	Cordless tel.
3	Phone3	Cordless ISDN tel.

Source: Glasker, 2017.

As seen in the above product dimension table, when the SCD type 1 response is implemented, no historicization takes place. The corresponding data record is simply overwritten.

SCD Type 2

For a dimension historicization according to **SCD type 2**, two additional attributes are required that define the validity interval of the dimension. When a dimension is changed, the upper limit of the validity of the previous dimension is set to the date of the change. The new dimension is inserted as a new data record with a validity interval from the date of the change to infinity.

SCD type 2
This is where a validity interval is added to each data record.

With SCD type 2, a validity interval is added to each data record, which is defined by the attributes ValidFrom and ValidTo. When the new record arrives, the previous entry for Phone3 is recorded by setting the ValidTo attribute to the day before the change. The new record for Phone3 is inserted in the table with the validity interval from the date of the change to infinity. In this way, the original product group remains traceable through the history management as well as the date of the change.

In order to distinguish the current record from the historicized one, the validity period must be included together with the product number in a new composite primary key. Alternatively, the distinction can also be made by replacing the key with a surrogate key. The new key must then be propagated to the fact tables so that new facts reference the correct dimension (Kimball & Caserta, 2004, p. 185). The mapping from the previous product to the changed one must be controlled via the metadata of the ETL process.

Table 10: Historicization SCD Type 2

Product number	Product name	Product group	ValidFrom	ValidTo
1	Phone1	Corded tel.	01/01/1999	31/12/9999
2	Phone2	Cordless tel.	01/01/1999	31/12/9999
3	Phone3	Cordless tel.	01/01/1999	01/06/2015
4	Phone4	Cordless ISDN tel.	02/06/2015	31/12/9999

Source: Glasker, 2017.

In the above table, a historicization according to SCD type 2 was carried out using a validity interval and surrogate key.

SCD Type 3

SCD type 3
This is where the data record is extended by an attribute with the new name.

SCD type 3 extends the existing data record by an attribute containing the new designation, i.e., a new field is added. The original state and the current state are not completely historicized, rather only the original state and the current state are saved. With SCD type 3, the product table is extended by the additional attribut "NewProdGroup" in which the new product group is stored.

Table 11: Historicization SCD Type 3

Product number	Product name	Product group	New product group
1	Phone1	Corded tel.	Corded tel.
2	Phone2	Cordless tel.	Cordless tel.
3	Phone3	Cordless tel.	Cordless ISDN tel.

Source: Glasker, 2017.

In the above table, a history (SCD type 3) was created using the additional attribute "NewProdGroup." The additional attribute allows old and new product groups to be traced. Multiple changes cannot be saved. The date of the change is not saved.

Comparison of SCD Types 1—3

SCD type 1 is the easiest to implement and keeps the data volume low. However, there is no historicization. Important information may be lost when data is overwritten. Given the analytical nature of most applications utilizing the data warehouse, this loss of information will be unacceptable in most cases.

SCD type 2 offers full historicization of dimensional changes. The implementation of SCD type 2 is complex. The necessary recognition of the change is also difficult. However, with SCD type 2, no information is lost.

SCD type 3 only historicizes the original and current value of an attribute. Information is lost if an attribute is changed several times. This type of dimension change is also complex to implement. Therefore, SCD type 3 is only to be used for very special cases. An example would be the conversion of postal codes. Since it can be assumed that such a change occurs rarely and the change does not have to be fully traceable, the loss of information can be accepted here.

SUMMARY

Users usually have different interests in the data generated via business intelligence. Flexible views are therefore necessary. Multidimensional data models can be used to arrange data in a multidimensional data space.

The focus of the OLAP cube model is usually on business ratios as carriers of quantitative information, which are described by corresponding sets of dimensions. There are a number of different concepts (ROLAP, MOLAP, and HOLAP) that exist for the storage of data in a DWH.

In the physical implementation of a multidimensional data model, a distinction can be made between two different database structures: the star schema and the snowflake schema. The historicization of dimensions, also known as “slowly changing dimensions (SCD),” is a regularly occurring challenge in the DWH context. Slowly changing dimensions can be processed via three different ways of recording changes in dimension tables.

UNIT 5

ANALYTICAL SYSTEMS

STUDY GOALS

On completion of this unit, you will have learned …

- how analysis systems or front-ends can be classified.
- what is meant by free data research and ad hoc analysis systems.
- which reporting systems are used in business intelligence.
- which model-based and concept-oriented analysis systems exist.

5. ANALYTICAL SYSTEMS

Introduction

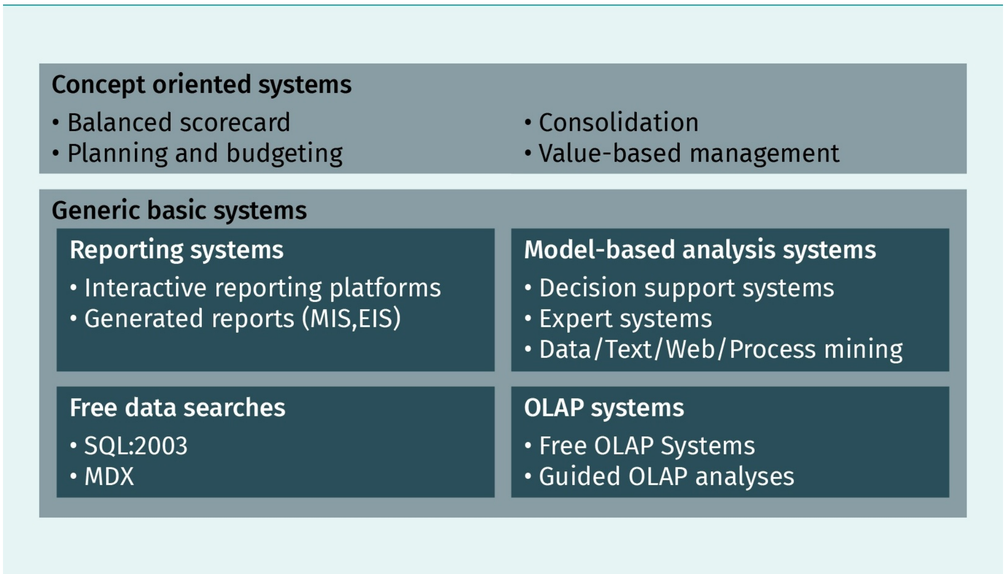
Based on the data stock of an implemented data warehouse architecture, special IT systems can be used to perform analyses for the purpose of information retrieval. Accessing a data warehouse (DWH), or the underlying data marts, can be done in different ways and with different applications—this often depends on the IT knowledge of the respective user.

In order to satisfy the various demands of users, an array of procedures and tools, each with their own distinct features, were developed, making classifying them difficult. One possible classification differentiates between

- free data research,
- ad-hoc analysis systems (e.g., OLAP),
- report systems,
- model-based analysis systems, and
- concept-oriented systems,

wherein the first four classes are (simply) categorized as generic systems. The following figure illustrates the facts and also shows some of the common aspects of these five procedures.

Figure 21: Analysis Systems for Management



Source: Kemper et al., 2010, p. 90.

5.1 Free Data Research and OLAP

Free Data Research

Since both a data warehouse and data marts are data objects, a respective data query can be performed with a data manipulation language (DML) that underlies a database management system (DMBS), very often with structured query language (SQL). One such query is named free data research as it can generally be very flexibly formulated, particularly regardless of the underlying models (Kemper et al., 2010). Such SQL queries can, for example, be integrated directly into programming environments and executed there.

A second variant of free data research is Multidimensional Expressions (MDX), which was developed by Microsoft specifically for data marts and can be seen as an extension of SQL in the context of a data warehouse, since MDX queries inquire about the dimensions of a cube rather than the attributes of tables (as in SQL). Both variations—SQL and MDX—are directed toward technically adept users, who should be well-acquainted with, among other things, the structure and the underlying data model. Hence, they are not aimed toward a wide user base.

OLAP—Online Analytical Processing

The termonline analytical processing (OLAP) includes methods and technologies, which allow for ad-hoc analysis on the basis of multidimensional information. The user of such a procedure has the opportunity to analyze data under different "aspects"—these are called dimensions.

The term OLAP goes back to the British mathematician Edgar F. Codd, who came up with 12 rules to characterize the capability of an IT system (Codd et. al., 1993). These rules can be summarized into four groups and are listed briefly below.

General requests

The general requests are:

- Multidimensional conceptual view. A multidimensional conceptual data model is used as a basis, which allows the user to perform intuitive analysis.
- Transparency for data stocks. Data is compiled from heterogeneous sources and made available.
- Intuitive data processing. Slice and dice functions, as well as drill-down operations, are provided for analysis over dimensions.
- Accessibility. Access to data should be consistently and uniformly possible.

Requirements for report generation

The requirements for report generation are:

- Consistent response times for report generation. Independent of the underlying data models, the amount of data or number of dimensions, consistent and quick response times should be possible.
- Flexible generation. Ensuring of the comparability of different dimensions.

Dimensional aspects

The dimensional aspects are:

- Generic dimensionality. The structure of dimensions should be uniformly operationally available.
- Unrestricted cross-dimensional operations. Support of calculations and aggregation functions for dimensions.
- Unlimited amount of dimensions and classification levels. Between 15 and 20 dimensions of a data model should be supported.

Technological requests

The technological requests are:

- Dynamic handling of sparsely populated matrices. Dynamic memory.
- Multi-user support. According to the requirements of a client/server architecture.
- Client/server architecture. Optimized load balancing through different front-ends having access to various back-end servers.

Since Codd et al. published these rules with various companies, he was accused of being dependent on a manufacturer. Because of this, but also because of technical requirements, the importance of these rules for evaluating OLAP systems has lessened somewhat. Two years later, Nigel Pendse and Richard Creeth developed five rules under the acronym FASMI, which stands for fast analysis of shared multidimensional information. This defined OLAP systems based upon user-specific requests, such as (Pendse & Creeth, 1995):

- Fast. Queries should be answered within a time window of five to twenty seconds.
- Analysis. An OLAP system should be able to handle any required logic. The definition of a more complex analysis query should be realizable without much programming effort.
- Shared. An OLAP system should allow for multi-user operation, which implies the availability of suitable access protection mechanisms.
- Multidimensional. A multidimensional structuring of the data with full support of dimensional hierarchies is the main criteria.
- Information. During the analysis, all required data should be transparently available to the user. Analyses cannot be influenced by the restrictions of the OLAP system.

It follows, then, that OLAP systems are high-performance and easy to use. Their queries from the data source are compiled in a multi-dimensional data cube and they present their reports in the form of graphics and tables. The user can select individual criteria and combine them.

5.2 Reporting Systems

Reporting systems present users with a clear and simple evaluation of company data. There are many software solutions on the market that create and design reports. Graphic-based user interfaces as well as drag-and-drop operations support the analyst to create reports.

Scorecards and Dashboards

A number of useful presentation tools are used in the preparation of reports. Common visual presentation tools include "scorecards" and "dashboards". These tools typically include **key performance indicators (KPI)**. KPIs are key business indicators that represent the achievement of certain strategic goals. In the case of corporate websites, an example of an indicator would be the average length of time visitors spend on the website or the turnover generated via the website (Liberty, 2018).

Key performance indicators (KPI)
These are key business indicators.

Scorecards provide users with a snapshot of decision-relevant data that can be consumed at a glance. They summarize the current performance of the company (typically in the form of KPIs) so that the current performance of the company can be quickly compared with its target performance. Scorecards are usually comprised of large amounts of distributed information (e.g., KPIs) displayed in condensed form. The degree to which information is condensed and the visual form in which it is presented depends on the context. Key figures can be displayed in various visual forms such as charts, graphs, traffic lights, speedometers, and thermometer displays.

Scorecards
With scorecards, data are offered at a glance in visualized form.

The **dashboard** visualizes the key performance indicators, metrics, and other key data points from different areas of the company in a uniform screen presentation using simple business graphics and tables. It provides users with multiple data points that describe the current performance of the company in real-time. In addition, an alarm can be automatically triggered in the form of emails or SMS as soon as a certain threshold value is undercut or exceeded.

Dashboard
A dashboard visualizes key figures by means of graphics and tables.

In practice, the terms "dashboard" and "scorecard" are often used synonymously. However, there are a number of significant differences between the two. A scorecard is a type of report that displays a collection of KPIs (key performance indicators) and the performance targets for each KPI. A dashboard is a container for a related group of scorecard and report views that are displayed together. Thus, a dashboard contains a collection of other elements such as scorecards, reports, and filters. Another difference between a scorecard and a dashboard is that a scorecard is a report summarizing company performance at a set point in time whereas a dashboard reports company performance in real-time. Two types of reporting systems are described below. The difference between MIS and EIS is also shown.

Generated Reports

Management information systems (MIS)
These are report-oriented analysis systems.

Executive information systems (EIS)
These are cross-divisional information systems designed for upper management.

Management information systems (MIS) are report-oriented analysis systems that focus on the planning, management, and control of the operational value chain. MIS are used to help management gain an overview of business processes and make informed business decisions. MIS systems offer powerful visualization and analysis tools as well as provide attractive graphics and planning tools. Using MIS systems, trends can be visualized and "what-if" analyses can be carried out. **Executive information systems (EIS)** are company-specific and cross-divisional integrative and dynamic information systems that provide information support to upper management. They are characterized by a high degree of flexibility and ease of use (Kemper et al., 2010).

Modern MIS and EIS often resemble one another in their external appearance, presentation, and user interface. However, there are two differences between the two systems: (1) EIS primarily present highly condensed, controlling-relevant internal data, while MIS will present all relevant operational data, and (2) EIS often integrate external information that is often unstructured but nevertheless relevant to the company, whereas MIS typically only present internal company data. For these reasons, EIS users are usually higher in a company management hierarchy (Kemper et al., 2010).

5.3 Model-Based Analysis Systems

Model-based analysis systems
These are based on business models.

While free data queries and OLAP systems usually involve minor calculations, complex evaluations require **model-based analysis systems** that have a strong algorithmic or rule-based orientation. Decision support systems, expert systems, and data mining belong to this category (Kemper et al., 2010). Model-supported analysis systems use business models for calculations and analyses. These are therefore evaluations with a higher degree of complexity and abstraction.

Decision Support Systems

Decision support system (DSS)
This is an interactive, model- and formula-based system.

A **decision support system (DSS)** is an interactive model- and formula-based system (Kemper et al., 2010). A decision support system uses existing models and formulas to support management in more or less structured decision situations. As the name suggests, decision support systems provide recommendations for action, taking into account the available data. The focus is on planning in the narrower sense, investigating possible alternatives for action using mathematical methods and models (Hansen & Neumann, 2001).

Expert Systems (XPS)

An expert system is an information system that makes specialist knowledge available in a limited area of application. It essentially seeks to emulate the decision-making ability of experts using artificial intelligence technology. Its main components are a knowledge base (i.e., database with expert knowledge) and a problem-solving component (Hansen & Neumann, 2001). It should be noted that in addition to the specialist knowledge of the experts, their problem-solving techniques also flow into the expert system. This allows the experience of experts to be used in a holistic way (Kemper et al., 2010).

Expert systems combine learned heuristics with expert knowledge and are therefore able to a certain extent to solve novel problems independently. Due to the integrated knowledge acquisition component, knowledge within the system can be expanded or made obsolete and incorrect information can be corrected (Hansen & Neumann, 2001). The explanatory component serves to make the procedures of the expert system transparent. For this reason, XPS are assigned to the research area of "artificial intelligence". Today, expert systems are mostly used in the banking and insurance industry for credit assessments and risk analysis. XPS are often part of integrated applications, e.g., in the form of interactive help systems.

Data Mining

Data mining is the software-supported determination of previously unknown correlations, patterns, and trends evident in the data stock of very large databases or DWHs. In contrast to standard query tools, the analyst does not need to know what they are looking for from the outset; instead, they are guided towards interesting information.

Data mining
This process identifies previously unknown correlations, patterns, and trends.

Typical data mining methods include the following:

- Outlier detection. This is where unusual data records are identified (e.g., outliers, errors, changes).
- Cluster analysis. This groups objects based on similarities.
- Classification. Here, previously unassigned elements are assigned to corresponding classes.
- Association analysis. This is used to identify relationships and dependencies in the data in the form of rules such as "A and B normally follow C."
- Regression analysis. This deals with the identification of relationships between several dependent and independent variables.
- Aggregation. This is where data sets can be reduced to more compact descriptions without a significant loss of information.

There are many examples of data mining in practice. Banks use data mining to detect credit card fraud and profile customers who are likely to be unable to meet their credit obligations. In marketing, data mining is used to make sales forecasts, conduct customer segmentation, perform shopping cart analyses, and detect abuse. In human resources, data mining can be used to support personnel recruitment and detect employee errors or oversights. There has been a strong upswing in the use of data mining with the increase in web applications. The practice of web mining is also growing as data mining techniques are applied to the internet in order to generate information from online data.

5.4 Concept-Oriented Systems

Concept-oriented systems
These include tools that implement comprehensive business concepts in BI analyses.

Concept-oriented systems are business intelligence (BI) tools that provide analyses and data based on specific business concepts or procedures. The balanced scorecard is a well-known example of a concept-oriented system utilized in modern BI systems. There are many BI solutions on the market that provide tools for planning and consolidation based on specific business concepts such as value-based management (Kemper et al., 2010).

Balanced scorecard (BSC)
This type of scorecard directs the attention of management to all relevant parts of the company.

The **balanced scorecard (BSC)** is a specific form of business scorecard. The BSC originated from a research project conducted in the early 1990s by Robert S. Kaplan and David P. Norton (1996). The original impetus for the project was dissatisfaction felt at the time with the one-dimensional, financially-oriented criteria used to measure the performance of companies. The use of a single dimension to measure success was deemed insufficient to really measure the success of a company.

As a strategic planning and management system, the BSC focuses not only on the financial activities of the company but also on human aspects. In creating and using a BSC, the attention of management is directed to all relevant parts of the company and leads to a more balanced picture of the company's activities. Traditionally, the BSC framework views the company from four different perspectives: financial, internal process, customer, and learning and growth. However, these perspectives are not prescribed; they constitute a basic framework that can be supplemented with company-specific perspectives, as is done in many cases.

SUMMARY

In order to generate business-relevant information from company data, it must be analyzed accordingly. In principle, analysis systems can be differentiated into free data queries, ad-hoc analyses, and reporting systems as well as model-based and concept-oriented systems. With free data queries, data can be retrieved relatively easily using data manipulation languages such as SQL or MDX.

With the help of ad-hoc analyses or OLAP, the analyst has the possibility to adapt corresponding evaluations "live" or "online" in order to obtain different views of the data depending on specific business requirements. The reporting systems (e.g., scorecards, dashboards, MIS, EIS) allow a simple, clear evaluation and presentation of company data.

While free data queries and OLAP systems usually involve minor calculations, complex evaluations require model-based systems that have a strong algorithmic or rule-based orientation. Decision support systems,

expert systems, and data mining belong to this category. Concept-oriented systems are tools that provide analyses and data based on extensive business concepts or procedures.

UNIT 6

DISTRIBUTION AND ACCESS

STUDY GOALS

On completion of this unit, you will have learned …

– how knowledge is generated from business intelligence content.
– which support systems and formats are used for the distribution of information.
– how portals can be used to facilitate access to information.

6. DISTRIBUTION AND ACCESS

Introduction

Business intelligence (BI) analyses can provide a company with important insights that support informed and strategic decision-making. However, the role of employees in enacting subsequent decisions is not always sufficiently understood. In practice, it can be that results of BI analyses are only applied by a few users or not adequately distributed so that some decision makers and departments are left without important analysis results.

There are many reasons for a suboptimal supply of information. It is often the case that only a limited group of users can access the analysis results of data mining systems. It can also be that the selection of recipients for the manual transfer of analysis results can be subjective. However, the most common reason for suboptimal sharing of information is that BI analyses are often performed by specific departments without the knowledge of others.

The success of any BI analysis in the long-term is ultimately determined by the adequate supply of information to relevant parties within the company. Accordingly, corporate communication must be optimized and BI must be linked to the existing knowledge management structures within companies. The results of BI analyses must be documented and made available in appropriate form so the information can be used throughout the company. Content and document management systems, as well as central, customized BI portals, can be useful tools to ensure that this occurs.

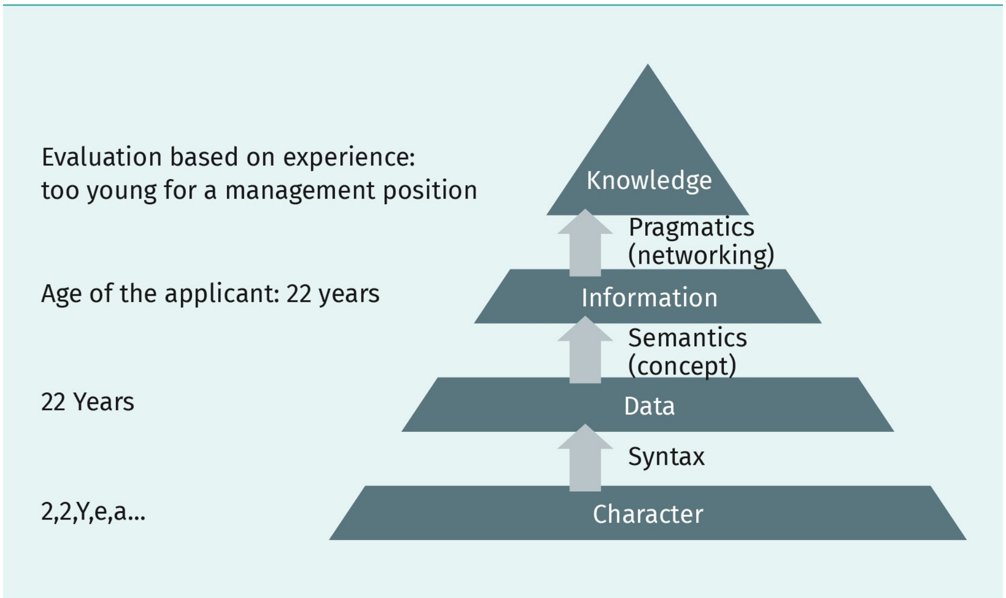
6.1 Distribution of Information

Valuable knowledge can be gained by implementing BI solutions. The targeted deployment of knowledge created via BI solutions and the effective distribution of knowledge that already exists within the company promise enormous gains in efficiency. BI users can benefit immensely from accessing knowledge that already exists within the company. This helps to avoid the duplication of work. Developers can also access existing or consolidated applications. For the systematic dissemination of knowledge, operational management tools can be used. In this context, content and document management systems can play a particularly important role. Content and document management systems, collectively referred to as knowledge management tools, are described below.

Knowledge Management

To aid the classification of knowledge management tools in the context of BI, the concept of knowledge is described in detail in the following figure.

Figure 22: Delineation of the Concept of Knowledge



Source: Bodendorf, 2006, p. 1.

DEFINITION: KNOWLEDGE
Knowledge is defined as the sum of facts, information, and skills acquired through experience or education used to solve problems.

When considering the above figure, we can see a hierarchy of components that constitute knowledge. We can see that knowledge is not just the accumulation of information; distinguishing which information is justified relative to the specific context (pragmatics) is the process by which information is transformed into knowledge. This property is not present in the information itself. Although information has context-dependent meaning (semantics), it only focuses on discrete aspects of a subject area. Data, comprised of alphanumeric characters that have been put together according to predefined rules (syntax), only becomes information when combined within the structure of meaningful syntax. These components of knowledge become relevant when considering how to manage the distribution and access to it.

Knowledge management enables companies to document operational knowledge and make it available to relevant employees. On a technical level, knowledge management systems provide IT-based support for operational knowledge management. The degree to which knowledge can be codified will determine how it can support operations; that is, knowledge that can be stored or documented in a structured form is the knowledge that can be made accessible to corresponding user groups via knowledge management systems. In contrast to this form of knowledge, implicit knowledge refers to the knowledge stored in the minds of employees which cannot be stored in electronic form due to its complexity. The exchange of implicit knowledge is primarily possible through interpersonal communication.

Content management systems (CMS) and document management systems (DMS) are designed for handling unstructured data and are used to manage codified knowledge. These systems are critical when building integrated BI and knowledge management solutions. Integration is facilitated in part by the fact that leading end-to-end providers offer corresponding BI and CMS tools as product suites that are designed to work together.

Content and Document Management Systems

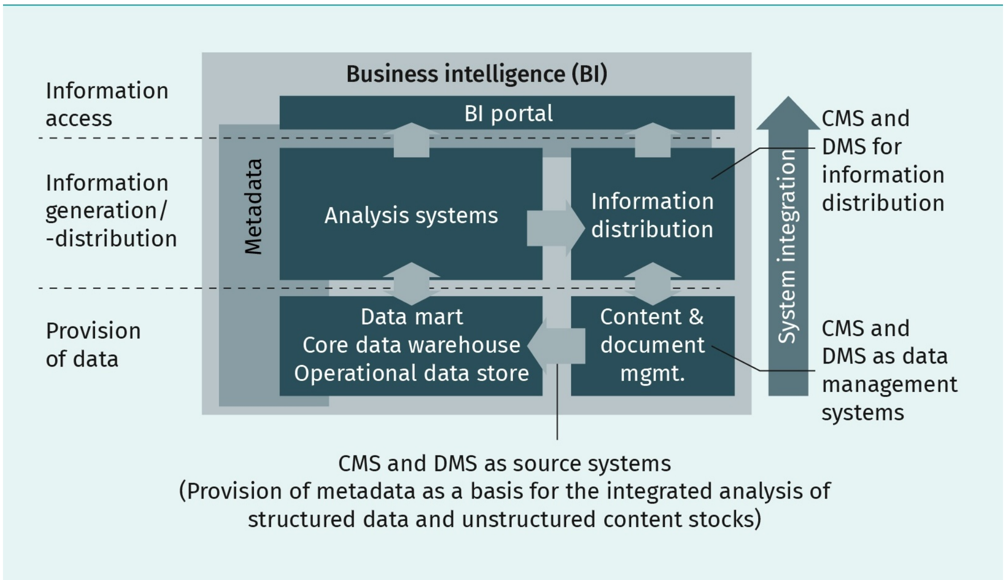
Document management systems
These are used for the effective management of paper-based documents.

Content management systems
A content management system focuses on handling different media formats.

Both content management and document management systems support unstructured data. **Document management systems** were originally designed to efficiently manage paper-based documents. DMS functions include the capture, archiving, version management, and provision of documents in electronic form. **Content management systems** focus on handling heterogeneous media formats such as numerical data, text, graphics, images, audio, or video sequences.

With CMS, content, structure, and layout are strictly separated to allow media to be used multiple times. CMS support the insertion, updating, and archiving of articles as well as content preparation and compilation in case of use. A CMS is supported by procedures for version control, authorization assignment, and quality assurance. The different systems grow together due to increasing web-based infrastructures. DMS are extended by CMS functions and vice versa. The following figure shows that CMS and DMS can be used for information distribution, data storage systems, and source systems within a BI approach.

Figure 23: Possible Roles of CMS and DMS in a BI Approach



Source: Kemper et al., 2010, p. 146.

CMS and DMS have various functions for the controlled management and distribution of content. The different functions and their potential for use in the context of distributing BI knowledge are

- dedicated access control,
- workflow management,
- life cycle concepts for knowledge,
- check-in/check-out: controlling concurrent access,
- document collation,
- version management, and
- information retrieval.

With the help of corresponding DMS or CMS functions, role assignments can be differentiated and different rights assigned. In this way, access to distributed BI knowledge can be controlled centrally (i.e., via dedicated access control). Multi-stage release processes are made possible by workflow management. The problem of "information overhead" (i.e., resources such as processing time or storage space consumed) can be avoided due to the underlying functionality (DMS or CMS).

The distributed BI content does not usually have unlimited validity. Analysis results and the analysis models behind them become unusable at some point, e.g., due to the dynamics of market structures. As a result, there are functions available to provide BI documents with an "expiration date" (life cycle concepts for knowledge units).

With the check-in/check-out mechanism, concurrent access can be avoided if content is required for more than simply "read" access. For example, updates are required to adjust model parameters or update obsolete result reports. With the help of DMS, individual documents can be combined into a collective document. Corresponding contents can be stored together, e.g., balanced scorecard reports can be bundled together to enable different detail views. Version management functions allow you to trace the history of any BI content or documents that may be modified.

Information retrieval refers to functions that find documents for specific information needs. The corresponding functionality is therefore of central importance for information distribution. The integration benefit increases if a corresponding document stock is already stored in a DMS or CMS. In the case of a query, both the existing and supplemented BI content is delivered. Accordingly, such an approach can also be understood as a step towards merging structured and unstructured information.

Distribution

Knowledge content ready for distribution can be used more efficiently and flexibly depending on its preparation. These range from static documents with limited scope for edits to customized templates that can be used as samples for applications. Distribution can be divided into the distribution of analysis results and the distribution of subsystems.

Distribution of analysis results

DMS and CMS can be utilized to reuse electronic documents created in the context of BI throughout the organization. In particular, results that were generated using analysis systems are distributed. These are typically prepared as MS-Excel files, PDF documents, and web pages.

The further use of results is facilitated if the information is kept in a machine-readable form. For example, the CSV format can be imported by almost all programs for data preparation and analysis. Values are stored in simple text files and mapped to a tabular structure using separators. Further flexibility in data exchange can be achieved by using extensible markup language (XML). The XML schema can be used to define extensive structure and format specifications. The power of XML lies in the fact that different dialects can be combined.

In the BI field, a variety of dialects are used. A commonly used format is extensible business reporting language (XBRL). XBRL is primarily designed to be used in the exchange of business data. For example, the XBRL format can be used to define financial statement data in order to meet specific accounting standards and to publish it as required for external accounting purposes. Typical recipients include auditors, shareholders, analysts, and government organizations. The XBRL format is also suitable for exchanges between application systems as well as for internal reporting. It is also increasingly used as an exchange format in the BI area, as a result of its widespread use, simplicity, and flexibility.

Distribution of subsystems

OLAP and data mining tools are designed to solve unique and individual results. For example, an analysis conducted via OLAP might review a one-off sales promotion or identify the cause of a decline in regional sales. The individual results of these analyses have no corresponding validity in any other context. Disseminating these results company-wide is therefore not a valuable exercise.

Moving beyond the actual results of each analysis, the knowledge regarding the preparation and execution of the analysis is actually relevant. The analysis model, analysis procedure, dimensions, parameters, data connection, and, if necessary, the form and layout of the data preparation should be recorded. In a similar way, reusing report definitions can also support future report development. Writing a user manual involves considerable effort. Reusing a manual can be made easier if machine-readable formats for defining application modules are exchanged.

In addition to the manufacturer-specific formats, XML standards can also be used for this purpose. The predictive model mining language (PMML) is designed for the description and exchange of data mining models. Here, data sources, preparatory transformations, and parameters of the model used are specified. Similarly, the report definition language (RDL) from Microsoft is used to record corresponding report definitions. Here data binding, layout, and report metadata are all taken into account.

The professionalism of reuse is further enhanced by using individually configured application templates. Templates are ideally created so that manual saving, loading, and installation activities are not necessary. Leading manufacturers have implemented corresponding concepts and supply standard templates for common analysis applications. In addition to the analysis layer, some of these templates also contain data modeling and data source connection. In larger BI environments, the targeted distribution of templates via DMS or CMS has become increasingly relevant.

6.2 Access to Information

Access via Portals

A user interface for management support must hide the complexity of the BI infrastructure through integration performance to ensure consistent and efficient access. In addition, personalization must be used to meet the requirements of various users. In principle, BI content can reach the user in various ways, e.g., through analysis front-ends, as part of company websites, and integrated into operational applications. Newer channels include feeds, which are designed for distributing discrete content to different front-end applications. Feeds can be used for exception reporting, e.g., via widgets. Widgets are independent program components with their own user interface that are integrated into a graphical user interface. Widgets are often used for the visualization of process states. In addition to stationary devices, internet-enabled mobile devices are increasingly being used. Smartphones are a popular end device used in this context.

The implementation of solutions is simplified by open, internet-compatible formats for the transmission and conversion of data. Implementation is also facilitated by technologies for the web-based provision of functions or individual program components. **Portals** are a particularly effective and convenient way of presenting BI analyses to end users. These are mostly central web applications with which companies can offer structured information.

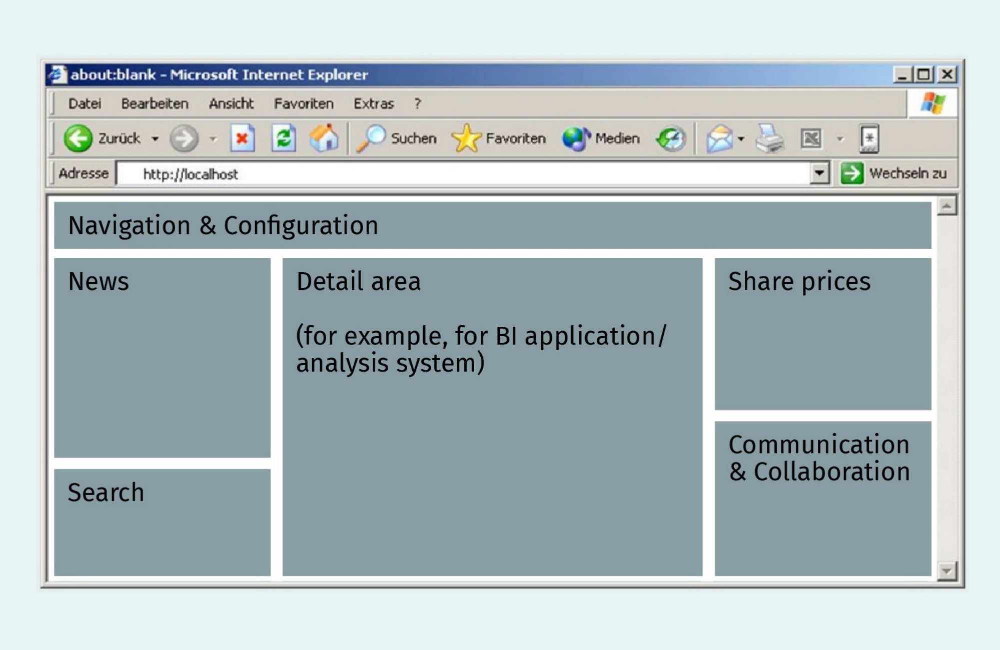
Portals
These are central web applications used by companies to offer structured information.

Content Integration

The most important feature of portals is that different content and applications are combined via a common interface. For example, BI portals integrate web-based analysis systems that are made available to users without them needing to install additional software. This gives management and other users centralized and structured access to the available information. The users have a defined entry point into the company knowledge and are enabled to make informed business decisions. Technically, a BI portal is composed of several parts called **portlets**.

Portlets
These are the component parts of a portal.

Figure 24: Schematic Structure of a Portal



Source: Kemper et al., 2010, p. 158

For example, one portlet can be responsible for news, another for BI applications, and a third for communication and collaboration. The content can come from internal or external company sources. Detailed infrastructure is required to implement BI portals. It should consist of the following components:

- A portal framework with functions for personalization, single sign-on (to give a portal user access to all required content and applications according to their authorization profile), and device-independent access. The latter is intended to allow mobile employees external access from outside the company.
- Integration. Modern portals support the integration of almost any kind of data sources and applications, from any database and simple text files to third-party applications such as ERP/CRM and XML imports via the internet. The use of portlets ensures fast integration, with non-programmers being supported by numerous wizards.
- The following BI tools should be included: end-user-oriented analysis and reporting tools, OLAP and data mining tools, and ETL tools.
- Basic services such as administration, security, directory services (SSO), and other services facilitate the use of the portal and ensure security.
- Data warehouse infrastructure allows access to dispositive data and data evaluation.

In order for the results of BI analyses to be used throughout the company, they must be documented and made available to specific employees in an appropriate form. The BI concepts that operate within companies should therefore be integrated into the company IT.

Content management systems play an important role here. They make it possible to provide analysis results and models to authorized interested parties in a targeted manner. BI portals are ideal for central information access. With these central web-based applications, all BI analyses and information can be made available via a uniform user interface. Through personalization and single-sign-on access, BI portals can also be highly individualized and adapted to the needs of individual users.

Personalization

Personalization can be applied differently depending on the needs of the organization, i.e., role-based, group-based, or per individual. Alongside integration, personalization is another central characteristic of portals. Personalization is used to offer results tailored to the individual needs of users, e.g., sales and research and development departments only receive the data that is interesting or relevant to their area of responsibility.

Personalization
This allows content to be offered in a user-oriented way.

Role-based or group-based personalization is where content is personalized uniformly for specific roles or groups, so that the same settings apply to all users who belong to that specific group. In contrast, individual personalization is user-specific and always tailored to a single person. Individual personalization can be performed explicitly or implicitly. With explicit personalization, the user actively defines settings such as the portal layout and content such as specific channels or applications. Implicit personalization uses user profiles and usage data and independently makes recommendations for relevant portal content. Another way to increase user orientation is the single sign-on. The basis for a single sign-on access system is a directory service such as a lightweight directory access protocol (LDAP).

SUMMARY

BI analyses can provide a company with important insights that support informed and strategic decision-making. Accordingly, corporate communication must be optimized and BI must be linked to the existing knowledge management structures within companies. The results of BI analyses must be documented and made available in appropriate form so that the information can be used throughout the company. Content and document management systems as well as central, customized BI portals can be useful tools to ensure that this occurs.

Document management systems were originally designed to efficiently manage paper-based documents. DMS functions include the capture, archiving, version management, and provision of documents in electronic form. Content management systems focus on handling heterogeneous media formats such as numerical data, text, graphics, images, audio, or video sequences. The further use of results is facilitated if information is kept in a machine-readable form. Portals are a particularly effective and convenient way to present BI analyses to end users. These are mostly central web applications with which companies can offer structured information.

BACKMATTER

LIST OF REFERENCES

Bachmann, R., & Kemper, G. (2011). *Raus aus der BI-Falle. Wie Business Intelligence zum Erfolg wird* [Out of the BI trap: How business intelligence becomes successful] (2nd ed.). mitp.

Bauer, A., & Günzel, H. (2008). *Data Warehouse Systeme. Architektur, Entwicklung, Anwendung* [Data warehouse systems: Architecture, development, application] (3rd ed.). dpunkt.

Bodendorf, F. (2006). *Daten- und Wissensmanagement* [Data and knowledge management] (2nd ed.). Springer.

Chamoni, P., & Gluchowski, P. (2015). *Data Warehouse Systeme. Architektur, Entwicklung, Anwendung* [Data warehouse systems: Architecture, development, application] (4th ed.). dpunkt.

Chen, P. (1977). *The entity-relationship model: Toward a unified view of data*. Creative Media Partners.

Codd, E. F., Codd, S. B., & Salley, C. T. (1993). *Providing OLAP to user-analysts: An IT mandate arbor soft* . E.F. Codd & Associates

Gansor, T., Totok, A., & Stock, S. (2010). *Von der Strategie zum Business Intelligence Competency Center (BICC). Konzeption—Betrieb—Praxis* [From strategy to business intelligence competency center (BICC). Conception—Operation—Practice]. Carl Hanser Publishing House.

Gluchowski, P., Gabriel, R., & Dittmar, C. (2008). *Management Support Systeme und Business Intelligence. Computergestützte Informationssysteme für Fach- und Führungskräfte* [Management support systems and business intelligence: Computer-aided information systems for specialists and managers] (2nd ed.). Springer.

Granularity. (2020, August 20). In *DWH Wiki*. http://en.dwhwiki.info/glossary/g/granularity

Grothe, M. (2000). *Business Intelligence. Aus Informationen Wettbewerbsvorteile gewinnen* [Business intelligence: Gain competitive advantages from information]. Addison-Wesley.

Gutenberg, E. (1983). *Grundlagen der Betriebswirtschaft. Band 1: Die Produktion* [Fundamentals of business administration. Volume 1: Production] (18th ed.). Springer.

Hannig, U. (2002). *Knowledge management and business intelligence*. Springer.

Hansen, H.-R., & Neumann, G. (2001). *Wirtschaftsinformatik I. Grundlagen und Anwendungen* [Business informatics I. Foundations and applications] (8th ed.). UTB.

Humm, B., & Wietek, F. (2005). Architektur von Data Warehouses und Business Intelligence Systemen [Architecture of data warehouses and business intelligence systems]. *Informatik-Spektrum, 28*(1), 3—14.

Inmon, W. (2005). *Building the data warehouse* (4th ed.). Wiley.

Kaplan, R., & Norton, D. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review, 74*(2), 75—85.

Kemper, H.-G., Baars, H., & Mehanna, W. (2010). *Business Intelligence—Grundlagen und praktische Anwendungen. Eine Einführung in die IT-basierte Managementunterstützung* [Business intelligence—Foundations and practical applications. An introduction to IT-based management support] (3rd ed.). Vieweg + Teubner.

Keyes, J. (2006). *Knowledge management, business intelligence, and content management: The IT practitioner's guide.* Taylor & Francis Ltd.

Kimball, R., & Caserta, J. (2004). *The data warehouse ETL toolkit*. Wiley.

Kurz, A. (1999). *Data warehousing: Enabling technology.* mitp.

Liberty, D. (2018, June 4). *Scorecard vs dashboard—What each adds to business intelligence*. Sisense. https://www.sisense.com/blog/scorecard-vs-dashboard-adds-business-intelligence/

Muksch, H., & Behme, W. (1998). *Das Data-Warehouse-Konzept. Architektur – Datenmodelle – Anwendungen* [The data warehouse concept. Architecture—Data models—Applications]. Springer.

Pendse, N., & Creeth, R. (1995). *The OLAP report*: *Succeeding with on-line analytical processing*. Business Intelligence.

Schinzer, H., Bange, C., & Mertens, H. (1999). *Data Warehouse und Data Mining. Marktführende Produkte im Vergleich* [Data warehouse and data mining. Market-leading products in comparison] (2nd ed.). Vahlen.

Schmidt, R. (1998). *Erich Gutenberg und die Theorie der Unternehmung* [Erich Gutenberg and the theorie of the firm] [Working paper]. Goethe Universität Frankfurt.

LIST OF TABLES AND FIGURES

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

Mailing Address
Albert-Proeller-Straße 15-19
D-86675 Buchdorf

media@iu.org
www.iu.org

Help & Contacts (FAQ)
On myCampus you can always find answers
to questions concerning your studies.