# Contents

Acknowledgments ......................................................................................................................... 9

## 1 Introduction .......................................................................................................................... 11

1.1 Introduction to SAP NetWeaver BW ......................................................................................... 11
1.2 SAP's Data Model in Business Content ..................................................................................... 13
1.3 Structure of the Book .................................................................................................................. 16
1.4 Target Groups ............................................................................................................................ 17

## 2 The Enterprise Data Warehouse (EDW) ............................................................................... 19

2.1 Introduction to the EDW ......................................................................................................... 19
2.1.1 The Corporate Information Factory (CIF) ............................................................................. 19
2.1.2 From the CIF to the EDW ....................................................................................................... 22
2.1.3 Denominations of the Layers ................................................................................................. 24
2.1.4 Costs for Setting Up an EDW ................................................................................................. 25
2.2 Layer 1 — Data Acquisition Layer .......................................................................................... 27
2.2.1 Number of Extracted Data ..................................................................................................... 27
2.2.2 Handling Business Content .................................................................................................. 28
2.2.3 Cleansing of Data .................................................................................................................... 28
2.2.4 Data Storage .......................................................................................................................... 29
2.3 Layer 2 — EDW Layer ............................................................................................................. 30
2.3.1 Storage of Data ........................................................................................................................ 30
2.3.2 Transformation in the EDW Layer ......................................................................................... 31
2.4 Layer 3 — ODS Layer .............................................................................................................. 34
2.5 Layer 4 — ADM Layer .............................................................................................................. 35
2.5.1 Objects in the Layer ................................................................................................................. 35
2.5.2 Type of Data Transfer Processes (DTPs) .............................................................................. 36
2.5.3 Storage of Data ....................................................................................................................... 36
2.5.4 Reporting at the ADM Layer ................................................................................................. 37
2.6 New Requirements in the EDW ............................................................................................... 38
2.6.1 Update of Characteristics ....................................................................................................... 39
2.6.2 Creation of Key Figures ......................................................................................................... 40
2.6.3 Extension of the EDW Model ................................................................................................. 40
2.7 Advantages and Disadvantages of an EDW ............................................................................ 42
## Contents

2.8 Variations in the EDW Concept ........................................................ 44  
   2.8.1 Variation 1 — Decrease of the Data Acquisition Layer .... 44  
   2.8.2 Variation 2 — Performance Increase of the ETL Process .... 46  
   2.8.3 Variation 3 — Prevention of Redundancies .................... 47  
   2.8.4 Variation 4 — Increase of the Reporting Performance ..... 48  
   2.8.5 Variation 5 — Increase of the Flexibility for New  
        Requirements ................................................................. 49  
2.9 Summary .................................................................................... 49

3 Development of an Enterprise Data Warehouse (EDW)  
   Based on Examples .................................................................. 51

   3.1 Introduction of the Example .................................................. 51  
       3.1.1 HCM Infotypes .......................................................... 52  
       3.1.2 Requirements on the Application .............................. 54  
   3.2 Concept of the EDW .............................................................. 55  
       3.2.1 Determining Required Characteristics .................... 55  
       3.2.2 Determining the Data Origin .................................. 55  
       3.2.3 Reference to Business Content ................................. 56  
       3.2.4 Designing the Data Model ...................................... 57  
       3.2.5 Designing the Extraction Transformation Loading  
            (ETL) Process ............................................................ 58  
       3.2.6 Other Special Features of the SAP ERP HCM System .. 60  
   3.3 Implementing the DataSources ............................................... 62  
       3.3.1 Activating the DataSources from Business Content .... 63  
       3.3.2 Creating Views ......................................................... 64  
       3.3.3 Creating Customer-Specific DataSources .................. 66  
   3.4 Designing the Basic Structure in the BW System .................... 69  
       3.4.1 Replicating the DataSources .................................. 69  
       3.4.2 Creating InfoAreas ................................................. 71  
   3.5 Creating InfoObjects ........................................................... 71  
   3.6 Modeling the Data Acquisition Layer .................................... 73  
       3.6.1 Modeling the Master Data ..................................... 74  
       3.6.2 Modeling the Transaction Data ............................... 76  
   3.7 Modeling the EDW Layer ...................................................... 78  
       3.7.1 Modeling the Master Data ..................................... 79  
       3.7.2 Modeling the Transaction Data ............................... 81
3.8 Modeling the ADM Layer ........................................................... 84
3.8.1 Modeling the Headcount (GP3_PB) .......................................... 85
3.8.2 Modeling the Personnel Actions (GP3_PM) ............................ 88
3.8.3 Modeling the Payment Analysis (GP3_PY) ............................. 90
3.8.4 Modeling the Reporting Layer ............................................. 91
3.9 Implementation of the Master Data Transformations .............. 93
3.9.1 Transformations of the DataSources to the A Layer .......... 94
3.9.2 Transformations from the A Layer to the EDW Layer .......... 96
3.9.3 Transformations from the EDW Layer to the A Layer ........ 102
3.10 Implementation of the Transaction Data Transformations ....... 104
3.10.1 Transformations of the DataSources to the A Layer .......... 104
3.10.2 Transformations from the A Layer to the EDW Layer .......... 104
3.10.3 Transformations from the EDW Layer to the ADM Layer ... 108
3.11 Summary .............................................................................. 112

4 Extension of the Enterprise Data Warehouse (EDW) .......... 113
4.1 Description of the New Requirements ...................................... 113
4.1.1 New Requirements of the Business Department ............... 113
4.1.2 Classification of the Requirements ..................................... 113
4.2 Update of Characteristics .......................................................... 114
4.3 Extension of the EDW Model ..................................................... 116
4.3.1 Extensions in the Source System ....................................... 116
4.3.2 Extension of the Data Model ............................................ 119
4.3.3 Extension of the ETL Process ............................................ 123
4.3.4 Extension of the Process Chains ....................................... 126
4.4 Summary .............................................................................. 128

5 Usual Requirements on Business Warehouse (BW) Systems .............................................................................. 129
5.1 Historization of Objects ......................................................... 130
5.1.1 Options for the Historization in SAP NetWeaver BW ......... 130
5.1.2 Type and Impact of the Modeling on the Historization .... 134
5.2 Number of Objects to be Included ......................................... 135
5.2.1 Impact on the InfoCube ................................................ 135
5.2.2 Methods for Modeling Many Objects ............................... 136
Contents

5.3 Admission of Future Data ........................................................... 146
5.4 Summary .................................................................................... 148

6 Load Control ............................................................................. 151

6.1 Principles of Data Currency ......................................................... 151
  6.1.1 Historical Scenarios .......................................................... 152
  6.1.2 Frequency of Data Update ................................................ 154
6.2 Load Control with SAP NetWeaver BW .................................... 155
6.3 Designing a Simple Load Control ............................................. 156
6.4 Designing a Complex Load Control ........................................... 162
  6.4.1 Concept of the Load Control ............................................ 163
  6.4.2 Implementation of the Load Control ................................. 164
6.5 Summary .................................................................................... 176

Appendices ..................................................................................... 177

A Literature ............................................................................................. 179
B List of Abbreviations ............................................................................ 181
C Glossary ............................................................................................... 183
D InfoObjects from Business Content ....................................................... 185
E The Author ........................................................................................... 187

Index ......................................................................................................... 189
1 Introduction

This book begins with a brief outline of SAP NetWeaver Business Warehouse (BW) and the data model provided in SAP Business Content. The aim is to illustrate how SAP NetWeaver BW and the data model of Business Content work. You will see that Business Content is well suited for prototyping purposes in the form of InfoCubes; for the project flow, however, it is no longer used. In the Enterprise Data Warehouse (EDW), which is presented in Chapter 2, only the extractors are still relevant.

This chapter concludes with a description of the book’s structure and the intended target groups.

1.1 Introduction to SAP NetWeaver BW

If you consider the system landscape of your enterprise, you usually find more than one system that is used for corporate management. The larger your enterprise is, the more systems that usually exist. There are two reasons for this: data protection and enterprise history. Small enterprises generally require only one system for management and control. If the enterprise expands and integrates further subsidiaries, the number of systems grows. These new systems typically differ from the systems currently used (see Figure 1.1).

![Diagram of heterogeneous system landscape in large enterprises]

**Figure 1.1** Heterogeneous System Landscape in Large Enterprises
The heterogeneous systems and system landscapes are problematic if you want to implement consolidated reporting across the entire enterprise. Due to the differences of the individual systems, you need to consolidate to reduce the data to a common denominator and make it available throughout the enterprise. One solution could be to consolidate your system landscape. However, this is only possible to a certain degree. Another solution is the *data warehouse*, which you can consider as an intermediate system of your enterprise (see Figure 1.2). The name of SAP’s data warehouse solution is SAP NetWeaver BW and it is an integral part of the SAP NetWeaver platform.

![Diagram](image)

**Figure 1.2** Using SAP NetWeaver BW for Company-wide, Consolidated Reporting

The central task of a data warehouse is to enable company-wide reporting by connecting the ERP systems that are referred to as “source systems” to the data warehouse. The data of the source system is then cleaned (reduced to a common denominator), consolidated, enriched, and saved in special data models optimized for reporting. This is the multidimensional data model that is mapped with SAP’s extended star schema in SAP NetWeaver BW.

The data model is the core of a data warehouse. The application, and therefore the acceptance of the system, completely depends on it. Therefore, good data models should at least have the following properties:

- high-performance reporting
- quick response to new requirements and changes
- flexible evaluation options
There are many ways to design data models that fulfill these criteria. One approach is the EDW approach, which is described in the following chapters. The EDW consists of different tiers (or layers) to provide flexible reporting with high performance and efficiency in changes to the data model.

Therefore, EDW features the basic prerequisites of a good data model. However, these very positive properties also involve some disadvantages:

- **Increased memory consumption**
  The different tiers, which are described in the course of this book (see Chapters 2 and 3), cause redundancy in data and result in higher storage space requirements.

- **Higher implementation effort in the beginning**
  The structuring of the different tiers is associated with higher implementation effort. The effort relativizes with the increasing age of the data warehouse, but initially it is higher than for conventional data warehouses that emerged as project solutions (standalone solutions).

- **Data protection aspects**
  The first tier of EDW extracts more data than is required for evaluation. This can result in problems in data protection. The data protection aspects are discussed in more detail in Section 2.2, “Layer 1—Data Acquisition Layer.”

Section 2.7, “Advantages and Disadvantages of an EDW,” details the pros and cons of the EDW approach.

### 1.2 SAP’s Data Model in Business Content

For the design of data models, you always follow the same path: You design a data model based on the reporting requirements, develop the extractors from the source systems, and consolidate and clean the data for evaluation from the source system to the evaluation level. The design of extractors, in particular, is very complex and should not be underestimated.

Here, an SAP NetWeaver BW system sets itself apart from other data warehouse solutions. SAP NetWeaver BW provides you with various preconfigured objects in the form of Business Content, which you can use for your data model. Business Content covers all facets of a reporting system. It contains data models in the form of InfoCubes, extractors from SAP source systems, and reports on the data
model. The extractors from the logistics department or payroll, in particular, are very complex and provide you with a real added value for the development of a data warehouse.

The data model of Business Content, however, which is briefly outlined in the following text, is used more for prototyping than for production use. Andreas Worch, trainer of SAP Course BW360, BW Performance and Administration, expressed this very aptly: “The InfoCubes from Business Content were modeled ‘nicely,’ but don’t have a high performance.” Therefore, the InfoCubes from Business Content are hardly ever used in production environments and are primarily used in prototyping. They are well suited for this purpose because you provide the users with a system that can be operated intuitively and give them a first impression of the BW system.

The majority of the objects in Business Content are structured according to the same schema: There is one or more DataSources (extractors) that extract the data from an SAP source system and forward it to an InfoCube via an InfoSource (see Figure 1.3). Furthermore, the major part of the data flow is still based on the 3.x data flow concept, that is, the new functions of SAP NetWeaver BW 7.0, such as data transfer processes and transformations, are not (yet) used. For more information on the new and old data flow and the concept migration, refer to the SAP PRESS book, SAP NetWeaver BW 7.0 Migration Guide (see Appendix A).

![Figure 1.3 Typical Dataflow in Business Content](image-url)
The InfoCube is modeled in such a way that related business objects are located in one dimension, that is, the Material characteristic can be found in the Material dimension, the Employee characteristic in the Employee dimension, and so on. As a result, users can handle the application more quickly; however, this type of data modeling doesn’t always give the best performance. Section 5.2, “Number of Objects to be Included,” and SAP Training BW360, BW Performance and Administration, provide more information on the performance in data models. Figure 1.4 shows the Employee-specific payroll data InfoCube, including its dimensions and the subordinate characteristics.

![Dimensions and Characteristics of the InfoCube Employee-specific Payroll Data](image)

To ensure high-performance and flexible data modeling, it is necessary to remodel the InfoCubes from Business Content. But because this is almost as complex as designing a new data model, you usually don’t use the InfoCubes from Business Content and design your own objects instead.
1.3 Structure of the Book

The book is structured as follows:

- **Chapter 2: The Enterprise Data Warehouse (EDW)**
  Chapter 2 provides you with the fundamentals of the EDW. First, you learn that the EDW evolved from the Corporate Information Factory (CIF) and has become a synonym for a tier model. You are provided with a detailed description of the structure of the individual layers and how you can implement them in SAP NetWeaver BW. There are numerous variations in addition to the classic EDW approach, which are also presented in this chapter.

- **Chapter 3: Development of an Enterprise Data Warehouse (EDW) based on Examples**
  After you’ve gotten to know the origin and structure of an EDW, this chapter introduces you to the modeling and implementation of an EDW based on a sample application from Human Resources (HR) management. You get to know the structure and implementation from the development in the source system to the modeling and implementation of the individual tiers and their transformations.

- **Chapter 4: Extension of the Enterprise Data Warehouse (EDW)**
  In this chapter, the sample application presented in Chapter 3 is extended with new requirements. The goal is to describe the structure of the process and the necessary steps. The extension is made in the EDW from HR management (from Chapter 3).

- **Chapter 5: Frequent Requirements on Business Intelligence (BW) Systems**
  For the introduction of a data warehouse, there are requirements that are frequently made by business departments. For example, a frequent request is that as many objects as possible are included in historical form—which negatively influences the system’s performance. Furthermore, a data warehouse is supposed to consider future data. These requirements can also be found in the EDW environment; this chapter details how you can implement these requirements there.

- **Chapter 6: Load Control**
  In data warehouses, data can be loaded and saved in many different ways. For example, scenarios are possible that are supposed to store historically stable data, but also historically correct data. Therefore, this chapter begins with a description of the problem of historically stable and correct data. You then learn
how you can implement these requirements via process chains. The implement-
tation is based on one simple and one complex example and the sample appli-
cation of Chapter 3 is used as the basis.

- **Appendix**
  The appendix includes a bibliography, a list of abbreviations, and a glossary. So
as not to interrupt the reading flow in Chapter 3, the appendix also includes a
summary of the InfoObjects that originate from Business Content.

### 1.4 Target Groups

The following target groups will find useful information on EDW or practical data
modeling:

- **Developers** who work in the BW environment and will be challenged with the
  implementation or consolidation of data warehouses, can find extensive infor-
mation about the EDW and comprehensive examples to support their
  implementation.

- **Consultants** who work in the area of BW and carry out data warehouse projects
  for customers can find useful practical tips and procedures for the successful
  implementation of an EDW and the relating load controls.

- **Administrators** who must manage and maintain an EDW can obtain an overview
  of the tasks to be carried out.

- **Users with a general interest in the topic, students, and other users** who require
  comprehensive information on the concepts of EDW, load controls, and practi-
cal information can find extensive descriptions in Chapters 2, 5, and 6.

- **Business departments** that are about to implement SAP NetWeaver BW can obtain
  an overview of the flexibility and the properties of an EDW to enable them to
  better assess an implementation.
2 The Enterprise Data Warehouse (EDW)

In the context of implementing Business Intelligence (BW) systems, the term EDW has been used more and more often in the last few years. The definitions, and particularly the interpretations, of the term vary greatly. Therefore, this chapter aims to detail the origin of the term and the underlying data model. This chapter discusses the theory of the EDW and the implementation within SAP NetWeaver BW.

2.1 Introduction to the EDW

Most users and developers know the term data warehouse. The basic idea of a data warehouse derives from its name: It is a warehouse in which you can collect information from different departments. From a linguistic point of view, the EDW is a warehouse that is restricted to business information, that is, it includes information on the management of materials and employee master records.

In the course of development, however, the term EDW has changed. It not only describes a data warehouse in which you store business information, it is also a tier architecture that enables optimal data storage and optimal reporting.

2.1.1 The Corporate Information Factory (CIF)

W. H. Inmon made a considerable contribution to the proliferation and acceptance of the EDW when he presented his CIF in the early 1990s. The basic idea of a CIF is that each enterprise should only have one single data warehouse that functions according to the principle: Extract Once, Deploy Many. This principle means that every date is extracted only once, but can then distributed within the CIF across many application areas or components. To ensure this, and to cover all business areas of different systems, you need a special architecture (see Figure 2.1).
As you can see in Figure 2.1, the CIF consists of five layers (also referred to as tiers) that are provided with numbers for a better overview. In the CIF, there are five layers:

1. External world
2. Data acquisition
3. Primary storage management
4. Data delivery
5. Metadata management

The **external world** is the external world of the CIF, it consists of the Internet, ERP systems, and transactional processes outside the CIF.

The **data acquisition** layer cleanses, integrates, and transforms data from applications to turn them into evaluable data, that is, *corporate data*.

The **primary storage management** layer contains the data warehouse and the alternative storage media, that is, the archive. Here, you store all data at detail level in a historicized manner for a long time.
The data delivery layer includes the analysis applications and objects for evaluation. Here, you can find the data marts, the Decision Support Systems (DSS) applications, and the data mining warehouse for statistic analyses.

The metadata layer includes the last three layers and determines how data is supposed to be stored within the CIF. It contains the “construction plan” of the CIF (see Haupt, Lehmann, Preuschoff, p. 3).

The layers shown in Figure 2.1 are divided into different components. The following technical terms are used in the literature for the CIF:

- **External world**
  The external world describes processes and data flows outside your enterprise. The external world is an essential tier because it provides the CIF with transactional data.

- **Applications**
  Applications are the applications within your enterprise in which you create, process, and store data. They are used to control your enterprise and were usually developed for special application areas (for example, product management or invoicing). They are divided into integrated and non-integrated applications. While the integrated applications were developed specifically for the CIF, the nonintegrated applications are applications of day-to-day activities that have grown historically.

- **Operational Data Store (ODS)**
  The ODS component is used to store detail data from many different systems. Therefore, the ODS is an integrated component of the CIF.

---

**Remarks on the Operational Data Store**

Do not confuse the ODS component with the identically named term used in earlier generations of SAP Business Warehouse (BW) (up to SAP BW 3.5). This was usually used for storing operational data at the document level. With SAP NetWeaver 7.0, the ODS object became the Data Store Object (DSO).

---

- **Integration and transformation layer (I&T layer)**
  The I&T layer is used for the transformation of external data of the applications or the external world. The inconsistent format of external data is transformed into an integrated format within the CIF, the corporate data.
The Enterprise Data Warehouse (EDW)

- **Data warehouse**
  The data warehouse is the core of the CIF. Here, you cleanse data, store it in a historicized manner over many years, and make it available to analytical applications. This data is stored at the details level.

- **Data mart(s)**
  A data mart is a subset of data from the data warehouse for special application areas. A data mart only contains aggregated data that is stored in special multidimensional structures in a historicized manner. Key words in this context are the (extended) star schema and the InfoCube as a multidimensional storage object.

- **Metadata management**
  Metadata is “data above data” without which the data model cannot exist. It describes the structures in a data model and the data flow between structures.

- **Exploration and data mining warehouse**
  In addition to the data mart, a data mining warehouse is also an analytic application of the CIF. Like a data mart, it is used for the analysis of corporate data; however, it deploys special data mining procedures. The data mining determines the customer’s buying behavior and responds accordingly, for example.

- **Alternative storage**
  The alternative storage is better known as the *archive*. It is used to store historical data that is no longer stored in the data warehouse. However, the data mining warehouses and data marts still work with the archive, although via aggregated views.

- **DSS**
  DSSs constitute a collection of analytical applications that operate on the data basis of the data warehouse. They are used for the analysis of data and usually provide aggregated view for decision support — hence the name.

Furthermore, the original CIF contains the Internet/intranet component. This component is not listed separately here because it is integrated in the external world.

### 2.1.2 From the CIF to the EDW

The tiers of the CIF can also be found in an EDW, whereas the definition and characteristics of the individual tiers have considerably changed in the 1990s. In
addition, there are many different denominations for each layer, which makes it difficult to compare the concepts.

Today’s EDW concept, as it is presented in most technical books, is strongly based on the CIF and consists of four layers:

- **Layer 1**
  The first layer describes the data inbound layer in which you store the data in its raw format at the details level. This data is stored for each source system.

- **Layer 2**
  The second layer is used to transform, cleanse, and enrich the data. This layer stores the data for a long time and in detail format. It is often referred to as the *corporate memory*. This layer is optimized for data storage.

- **Layer 3**
  The third layer is used for the evaluation of operational data and is therefore also referred to as *operational data storage*. This layer is optional and many enterprises omit it. The descriptions for the arrangement of this layer diverge substantially in the technical literature.

- **Layer 4**
  The fourth layer contains the actual objects for data evaluation. Usually, the data is available in aggregated format and with fewer fields than in Layer 2. It is optimized for reporting and consists of special, multidimensional objects.

Here, the layers don’t have any names intentionally, because the denominations vary strongly in the technical literature. The arrangement of the layers is also very different — although no specific arrangement can claim to be the only one that’s correct. Instead, it is important to consider the context and the integration with the application case in the technical books. The individual applications necessitate deviations — even though the basic idea of the EDW is the same in all cases. Only the different denominations make it difficult to recognize this context.

The definition and arrangement of layers used in this book strongly orientate on the CIF; this approach is sometimes referred to as the *classic EDW*. Figure 2.2 illustrates the arrangement of layers in the classic EDW approach.
You can establish the relation to the CIF as follows:

- Layer 1: data acquisition
- Layer 2: primary storage management
- Layer 3: ODS
- Layer 4: data delivery

The layer of metadata management from the CIF is not omitted in the EDW — in fact, it is an integral part of the EDW without which a data warehouse could not exist (see Staade, Schüler 2007).

### 2.1.3 Denominations of the Layers

As already mentioned, the denominations of the layers vary greatly. The author prefers the following denominations and therefore uses them throughout the book:

- Layer 1: data acquisition layer
- Layer 2: EDW layer
- Layer 3: ODS
- Layer 4: architected data mart (ADM) layer

Figure 2.3 shows the model with the preceding denominations.
This concept is identical to the common denominations except for the first layer. This layer is frequently referred to as the *staging layer*. In the context of SAP NetWeaver BW, however, this term can be confusing because you can easily mix it up with the *Persistent Staging Area* (PSA).

### Further Common Denominations for the Individual Tiers

For the sake of completeness, the following is a list of additional terms used for the individual tiers in the literature. Only layer 3 (ODS) has a uniform denomination.

- Layer 1: staging layer, data inbound layer, data provision tier
- Layer 2: primary storage layer, EDW and harmonization layer (EDW/H layer), harmonization tier
- Layer 4: data delivery layer, data mart layer

### 2.1.4 Costs for Setting Up an EDW

Setting up an EDW is not an easy process because most enterprises already have one or more data warehouses that have emerged as project solutions. Frequently, data protection plays a significant role that prohibits a combination of HCM and Financial Accounting (FI)/Controlling (CO) data in a system.

Therefore, the implementation of an EDW is associated with high costs if data warehouses already exist in an enterprise. Generally, the data models are created in different ways and for specific applications, which complicates the company-wide harmonization and transformation to the EDW concept. Therefore, the decision to
introduce an EDW can be extensive and political (see Staade, Schüler 2007) and is frequently rejected for these reasons.

Provided that no data warehouse exists in your enterprise, it is highly recommended to implement an EDW. In all other cases, the implementation of an EDW is a very tedious process that provides you with a data warehouse that is consistent and harmonized throughout the enterprise.

The implementation of an EDW usually involves an increased basic cost (see Figure 2.4). The reason for this basic cost is that the different layers must be modeled and set up consistently. Only when the number of new requirements increases does the initial additional effort pay off and balance the initial costs. It is an essential strength of the EDW to respond to new requirements, as you will see in the sample application in Chapter 3.

The following sections describe the individual layers in more detail and present possible implementations of the layers with SAP NetWeaver BW.
2.2 Layer 1 — Data Acquisition Layer

The *data acquisition layer*, the A layer or A tier for short, is used as the data recipient in the EDW. It receives data from all source systems connected to the EDW in raw format, that is, uncleansed. The A tier must strictly follow the *Extract Once, Deploy Many* principle, in other words, each piece of data must only be extracted once. In SAP NetWeaver BW, this requires that a DataSource be available (created or replicated) for each object of a source system to be connected. Therefore, the A tier contains the PSA that is connected to every DataSource.

2.2.1 Number of Extracted Data

In the EDW, the general rule is that you not only extract characteristics that are required for the reporting requirements, but you also determine a multitude of additional fields from the source system that may be required in future. Therefore, the A tier includes more fields than required, but they serve to cover future requirements. Frequently, it is difficult to identify these fields; therefore, users hardly ever take the trouble to find the fields, but extract all fields that belong to the appropriate group of themes (for example, all HCM infotypes used in an enterprise or all fields of required tables).

From a technical viewpoint, this is usually implemented via DataSources that are based on a database table or a view. This is also a relief for an enterprise that wants to connect a lot of source systems and doesn’t want to take on the tasks of development of extractors. Often, the extractors are also implemented by other companies that are not interested in drawing up complex projects just to provide data to other areas.

<table>
<thead>
<tr>
<th>View or Table?</th>
</tr>
</thead>
<tbody>
<tr>
<td>When creating a DataSource, you have the option to directly address the table or use a view as a basis. It is recommended to use the view because this enables you to hide fields of the database table that are not required in the BW system. Furthermore, this enables you to combine multiple tables and reduce the number of DataSources.</td>
</tr>
</tbody>
</table>

The benefit is that requirement changes only result in an adaptation of a view to transfer the corresponding field to the EDW. In other cases, you would have to extend the function module DataSources — a process that can be very time-consuming depending on the complexity of the function modules.
2.2.2 Handling Business Content

SAP NetWeaver BW provides Business Content that also offers extractors. These extractors are very complex in some subject areas and drop out of the EDW because they were usually implemented via function modules. To still work in compliance with the EDW, the Business Content DataSources are loaded on a one-to-one basis into the EDW and are further processed there. However, no field extension is carried out in the source system any longer.

For example, if you extract payroll data from SAP ERP HCM, you should use the 0HR_PY_1 DataSource from Business Content and enrich this data in the EDW first. For master data of employees you wouldn’t extend the 0EMPLOYEE_ATTR extractor in the source system, but implement this extension on the way from the A layer to the EDW layer.

<table>
<thead>
<tr>
<th>Handling Data Protection</th>
</tr>
</thead>
<tbody>
<tr>
<td>In many projects, the procedure for extracting more fields than required proved to be very problematic for data protection. Therefore, include the responsible contact person in the project from the outset and explain why it is necessary to add the data. An important reason in this context is that no reporting is implemented for the data at the lower layers.</td>
</tr>
</tbody>
</table>

2.2.3 Cleansing of Data

Furthermore, cleansings are implemented within the A layer that are required for further data processing. This means, for example, that objects that exist in multiple source systems and can have identical keys must obtain one shared, global key.

<table>
<thead>
<tr>
<th>Example: Cleansing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material 47110000 exists in both System S01 and System S02; however, the material number describes a different material in each of the two systems. This necessitates a shared key for the material.</td>
</tr>
</tbody>
</table>

The system ID of the source system is usually ideal as the shared key, that is, the ten-digit object becomes a 13-digit object.

Furthermore, you implement plausibility checks of the inbound data. This way, you can develop routines that check the absence of the appropriate fields (for example, missing material number, missing cost center) and cancel the processing.

28
### 2.2.4 Data Storage

As soon as the data has been loaded at the technical data inbound layer PSA of the EDW, the data is updated in the DSOs that consist of the same fields of the Data-Sources. As of SAP NetWeaver 7.0, special DSOs exist that considerably improve the Extraction Transformation Loading (ETL) performance: the *write-optimized* DSOs. These DSOs only contain the active data table and are only one relational database table. All data received in the objects is active automatically.

The data at the data acquisition layer is always deleted upon receipt or stored for a maximum of one month. The same applies to the content of the PSA tables. Figure 2.5 summarizes the lowest tier of the EDW layer model.

![Schematic Illustration of the Data Acquisition Layer](image)

You can view the DataSources implemented as views that are loaded on a one-to-one basis to the A tier (PSA). Then, you can (optionally) extend the key to make the data available for the next layer. The data itself is stored in write-optimized DSOs and is removed again during the next load process at the latest.

Table 2.1 summarizes the properties of the data acquisition layer.
The Enterprise Data Warehouse (EDW)

<table>
<thead>
<tr>
<th>Layer</th>
<th>Data acquisition layer (A layer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>Is used in the EDW for data receipt, the data is stored in raw format</td>
</tr>
<tr>
<td>Transformation</td>
<td>Cleansing, creation of global keys, plausibility checks</td>
</tr>
<tr>
<td>Data storage</td>
<td>Max. 1 - 2 months</td>
</tr>
<tr>
<td>Type of the objects</td>
<td>DSOs (write-optimized)</td>
</tr>
<tr>
<td>Contents</td>
<td>Master data, transaction data (separated from one another)</td>
</tr>
<tr>
<td>Reporting</td>
<td>Not permitted</td>
</tr>
</tbody>
</table>

Table 2.1  Data Acquisition Layer Summary

2.3  Layer 2 — EDW Layer

The EDW layer represents the memory of your enterprise. It is often also referred to as corporate memory. This means that you cleanse, transform, and store all detail data of all connected systems regardless of the application in this layer for a long time. All further layers make use of this layer; so the EDW layer is the “single point of truth” of reporting.

2.3.1  Storage of Data

Because the data is stored for a long time in the EDW layer, the system maps all objects for which storage is required as standard DSOs. This way, you can develop a delta procedure because, in principle, the data is received as complete data deliveries in the EDW. Write-optimized DSOs would be a possible alternative here provided that you don't require any delta procedures.

The objects of the EDW layer are stored application-neutral, that is, for no specific application. The data is stored in such a way that the application can optimally use the data inventory in the EDW layer. Therefore, the EDW layer is optimized for long-term storage, but not for reporting.

The DSOs of the EDW tier now contain both transaction and master data for the first time. This way, all transaction data is enriched with possible master data that is supposed to be stored in a historicized manner.
On the way to the EDW layer, it may be necessary to integrate intermediate layers that are responsible for the harmonization. These intermediate layers don’t conflict with the EDW concept because they can always be assigned to one layer.

### 2.3.2 Transformation in the EDW Layer

The transformations in the EDW layer from the data acquisition layer are usually very complex because they include the entire harmonization and enrichment. There is no ideal solution to implement the transformations and the harmonization. The optimal procedure depends on the application. Experience has shown, however, that the expert routine, which has been available since SAP NetWeaver 7.0, is very helpful.

---

**Expert Routine**

The expert routine has been available in the BW system since SAP NetWeaver 7.0. In addition to the known start routine, which had accessed the source structure and was therefore able to process data packages *prior* to the actual transformation, you are now provided with two new routines: the end routine and the expert routine. In contrast to the start routine, the end routine has access to the target structure, that is, it accesses the data *after* the transformation. The expert routine replaces the automatic transformation; here, the developer must make the assignments within a method in ABAP Objects himself. If you select the expert routine as the transformation, the start and end routine are no longer available. In addition, no monitor entries are written; these must be done by the developer. Figure 2.6 illustrates the connections of the different routines.

---

**Figure 2.6** Options for the Transformation in SAP NetWeaver BW
By means of the expert routine, you have access to both the source and the target structure and you can add all required data during the transformation. For example, it is possible to implement a transformation of the master data enrichment or the harmonization of master data in an expert routine. Section 3.9, “Implementation of the Master Data Transformations,” discusses this procedure in detail. Figure 2.7 illustrates this procedure. (The SAP Course BW350, BW Data Acquisition, provides more detailed information on the various transformations.)

![Diagram](image)

**Figure 2.7** Harmonization with Expert Routines

As you can see, the number of transformations is very low although a lot of objects are read. The benefits of using expert routines are the reduction of individual transformations and a decreased sparsity (gap formation) in the target object.

### Example: Sparsity

A target object has 2 key fields and 18 data fields; there are also 2 source objects that have the same key fields and 9 data fields each (see Figure 2.8).

Two independent transformations result in gaps in the target object because the fields not known in the source structures cannot be filled by the transformation. As a result, the number of data records in the target object increases, which negatively affects the performance.
Another benefit of the expert routine is its reusability — provided that you implement these routines with function modules or methods. Ultimately, an expert routine can work faster than standard transformations depending on the experience of the developer, for example, because the ABAP code snippets that are generated if formulas are used in transformations are no longer applied.

A disadvantage of the expert routine becomes apparent when you execute the Show Data Flow option for the target object of the EDW layer in the Data Warehousing Workbench (DWWB). Here, only the DSO of the A tier linked through the expert routine is displayed as the origin, not all of the other linked objects that are read in the expert routine.

However, this problem can be solved either with dummy transformations or corresponding documentation. Dummy transformations are created between all objects that are used in the expert routine, but are not displayed in the data flow. They are only created as a transformation without field assignment so that at least the data flow is displayed correctly. Dummy transformations are not frequently used, so a clear documentation is recommended.

Table 2.2 summarizes the properties of the EDW layer.
The Enterprise Data Warehouse (EDW)

<table>
<thead>
<tr>
<th>Layer</th>
<th>EDW layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>Corporate memory, long-term storage of all data at detail level</td>
</tr>
<tr>
<td>Transformation</td>
<td>Comprehensive harmonization and transformation of data, application-neutral storage of data</td>
</tr>
<tr>
<td>Data storage</td>
<td>Approx. 2 to 10 years</td>
</tr>
<tr>
<td>Type of the objects</td>
<td>DSOs, optionally of the “standard” or “write-optimized” type</td>
</tr>
<tr>
<td>Contents</td>
<td>Master and transaction data</td>
</tr>
<tr>
<td>Reporting</td>
<td>Not permitted (exceptions possible)</td>
</tr>
</tbody>
</table>

Table 2.2  EDW Layer Summary

### 2.4 Layer 3 — ODS Layer

As mentioned previously, the ODS layer is optional. It is only used if you want to permit the evaluation of operational data in the BW system. The evaluation of operational data can be useful, for example, if you want to evaluate data in the BW system that is updated on a daily basis to decrease the load on the transactional system.

The objects of the ODS layer are mapped via the standard DSOs, which consist of several fields depending on the application case. Normally, the data is not stored for a long time because they quickly lose their informative value. Figure 2.9 illustrates the typical structure of an ODS layer.

![Structure of the ODS Layer](image-url)

Figure 2.9  Structure of the ODS Layer
Table 2.3 summarizes the properties of the ODS layer.

<table>
<thead>
<tr>
<th>Layer</th>
<th>ODS layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>Evaluation of transactional data</td>
</tr>
<tr>
<td>Transformation</td>
<td>Minor harmonization, preparation for the evaluation, combination of data to evaluable objects</td>
</tr>
<tr>
<td>Data storage</td>
<td>Maximum six months, data that is updated on a daily or hourly basis</td>
</tr>
<tr>
<td>Type of the objects</td>
<td>Standard DSOs</td>
</tr>
<tr>
<td>Contents</td>
<td>Master and transaction data</td>
</tr>
<tr>
<td>Reporting</td>
<td>Permitted</td>
</tr>
</tbody>
</table>

**2.5 Layer 4 — ADM Layer**

The ADM layer contains the objects available for reporting in aggregated form and a subset of the fields that are present in the EDW layer. It therefore contains objects that are implemented for specific applications (data marts).

**2.5.1 Objects in the Layer**

In principle, the objects of the ADM layer are modeled via InfoCubes provided that the data is supposed to be historicized and formatted for reporting. The appropriate transformations are relatively simple because the data is already available in cleansed and harmonized form in the EDW layer. The evaluation of different subject areas or the calculation of key figures, however, may require some additional work. Here, the type of data marts that is supposed to be developed is an important factor. Figure 2.10 illustrates this structure.
2.5.2 Type of Data Transfer Processes (DTPs)

Because the underlying EDW layer contains historicized data that is not deleted after each load run as in the A layer, a standard DTP is no longer sufficient. The associated DTP must be supplemented with a corresponding routine that only extracts data that is current according to the load control. Chapter 6 provides further information on load control. Section 6.4, "Designing a Complex Load Control," illustrates an implementation example.

The routine in the DTP is used to select only the current data from the DSO. Therefore, the underlying DTPs must be of the "full" type. This enables you to delete overlapping requests from the InfoCube, if required, because the DELETION OF OVERLAPPING REQUESTS function checks for equivalence of the selection conditions.

2.5.3 Storage of Data

The storage of data in the ADM layer depends on the application. For example, some data loses its informative value after two years and is archived; other data, in turn, must be available for more than ten years. However, the minimum storage period is generally six months. Therefore, the ADM layer is another layer that stores data for a long time and has a lot of characteristics of the EDW layer. As
you can see, the layer architecture is redundant and requires larger data storage. Redundant storage, however, is indispensable for a consistent data warehouse.

### 2.5.4 Reporting at the ADM Layer

Basically, you model the data that is made available in data marts for reporting via InfoCubes. However, you should not implement reporting on InfoCubes in an SAP NetWeaver BW system. You should instead create MultiProvider or InfoSets, that is, logical views of the modeled InfoCubes, and only allow reporting on these views. SAP’s training courses provide comprehensive information on the use of MultiProviders.

How is this procedure justified? On the one hand, you are more flexible in modeling if you use MultiProviders or InfoSets because you can combine multiple InfoCubes in one MultiProvider. More important, however, is that InfoCubes have dimension structures that are different to those of the MultiProviders because they are stored in an optimized manner for quick reporting. In an InfoCube, you will hardly ever find all of the objects that are related to each other in terms of business in one dimension. They are instead collected according to technical aspects and grouped in dimensions. Only the MultiProvider contains dimensions of which the user benefits — you must always consider this for modeling! Figure 2.11 shows the use of MultiProviders.

![Figure 2.11 Use of MultiProviders in the ADM Layer](image_url)
The combination of InfoCubes via a MultiProvider has another performance advantage. For very large data quantities, you distribute the data across InfoCubes with identical structures, for example, separated according to the calendar year (partitioning). If you query for a period of two calendar years, the system creates two queries at runtime that access the underlying InfoCubes in parallel. The performance advantage increases if you create month InfoCubes and connect them in a MultiProvider, for example. Section 6.2, “Load Control with SAP NetWeaver BW,” provides additional information on this topic.

Table 2.4 summarizes the properties of the ADM layer.

<table>
<thead>
<tr>
<th>Layer</th>
<th>ADM layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage</td>
<td>Evaluation of aggregated data that is harmonized and stored in an optimized form for reporting</td>
</tr>
<tr>
<td>Transformation</td>
<td>Calculation of key figures, connection of logical units for creating reports</td>
</tr>
<tr>
<td>Data storage</td>
<td>Application-dependent, 6 months to 10 years</td>
</tr>
<tr>
<td>Type of the objects</td>
<td>InfoCubes, MultiProvider, InfoSets</td>
</tr>
<tr>
<td>Contents</td>
<td>Master and transaction data</td>
</tr>
<tr>
<td>Reporting</td>
<td>Permitted</td>
</tr>
</tbody>
</table>

Table 2.4  ADM Layer Summary

2.6  New Requirements in the EDW

This section discusses the handling of new requirements that can be implemented in the EDW model. In the EDW layer model, you can divide new requirements to different classes that require a different reaction and implementation:

- **Update of characteristics**
  A characteristic is supposed to be added that already exists at the EDW level.

- **Creation of key figures**
  A new key figure is supposed to be calculated based on existing characteristics. The characteristics are already at the EDW layer.
Extension of the EDW model
A characteristic is supposed to be added that is not available in any of the layers and cannot be derived from other characteristics.

The following sections describe these classes in more detail.

2.6.1 Update of Characteristics
New requirements usually require the inclusion of a new characteristic. Here, the EDW is well suited because the EDW layer already has more objects than are available in reporting at the ADM layer later on.

To provide a new characteristic, you only update the already-harmonized EDW layer to the corresponding InfoCubes of the ADM layer. Here, the initial extra costs for the implementation of an EDW pay off.

Depending on the variation of the EDW approach, the update can be done as follows (the individual variations are discussed in more detail in Section 2.8, “Variations in the EDW Concept”):

- A one-to-one update is possible because the EDW layer is already enriched with master data.
- The characteristic needs to be read from the master data for the update into the EDW layer, because the characteristic is not provided in the transaction data DSOs (variation 3, see Section 2.8).
- The new characteristics already exist in the InfoCubes of the ADM layer (variation 5).

After the update in the InfoCubes of the ADM layer, you only need to release the appropriate MultiProviders provided that you didn’t use variation 5 (see Section 2.8.5).

A critical advantage only becomes clear afterward: Legacy data can also be equipped with new characteristics thanks to the consistent storage of characteristics that are not (yet) required for reporting. For project solutions that are not based on the EDW approach and only contain data required for reporting, you must either reload the entire dataset (if the characteristic is to be historically filled) or the new characteristic is not available until the next load run. This problem doesn’t exist in the EDW.
2.6.2 Creation of Key Figures

In some cases, it might be necessary to develop new key figures and add them to the EDW. The type of the new key figures can cause different reactions in the EDW model:

- The new key figure can be created directly as a calculated or restricted key figure in the Query Designer.
- The new key figure can be calculated based on already-existing characteristics in the EDW layer.
- The new key figure is based on characteristics that don’t exist in the EDW yet.

For the definition of calculated or restricted key figures, you don’t need to modify the EDW — the key figures can be created in the Query Designer.

Provided that the key figures available in the InfoCubes (basic key figures) are not sufficient to cover the new requirement, but the existing characteristics in the EDW tier are, you can derive the key figure directly from the EDW layer. The calculation may be complex and require a routine. However, this routine must be developed from the EDW to the ADM layer. Here, the same rules apply as described in Section 2.6.1, “Update of Characteristics.”

- A simple creation of key figures is possible because the EDW layer is already enriched with master data and all required characteristics suffice for the new key figure.
- The characteristic needs to be read from the master data for the update into the EDW layer to calculate the new key figure with characteristics that are not yet provided in the InfoCube (variation 3).
- The new characteristics already exist in the InfoCubes of the ADM layer (variation 5). However, the calculation must also be implemented and does not differ from variation 3.

You must then create and assign the key figure in the associated MultiProviders.

2.6.3 Extension of the EDW Model

In rare cases, you need characteristics that are not yet available in the EDW, for example, for the new implementation or extension of components in the SAP ERP system. An enterprise can decide to implement positive time management (see
New Requirements in the EDW

box), which is supposed to be evaluable in SAP NetWeaver BW, for example. Or restructurings take place in the enterprise that require their own objects that must also be evaluable in SAP NetWeaver BW. So the extension of the EDW model is not the normal case! (Unless you are in the EDW implementation phase and noticed after the test phase that subareas are covered insufficiently or not at all.)

<table>
<thead>
<tr>
<th>Positive and Negative Time Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>The terms, <em>positive</em> and <em>negative</em> time management, originate from Human Resources (HR). For positive time management, all actual times of the employees are stored in the SAP HCM system. If no time is recorded, this is considered an absence. For negative time management, however, the system assumes that the employees work according to their planned working time. Only deviating times (for example, vacation or illness) are recorded. Negative time management is easier to implement than positive time management.</td>
</tr>
</tbody>
</table>

For this type of requirement, you need a new modeling that is performed analogous to the implementation. Here, you should consider whether this new characteristic suffices for the future or whether it is foreseeable that further characteristics must be added.

Initially, you add the requirement in the form of new characteristics or key figures and specify the data origin. Then, you decide whether the specified table is added to the EDW as a whole (which is the normal case) or in parts. Subsequently, you develop a new DataSource for each source system, which is created based on a view, and replicate it to SAP NetWeaver BW.

Now, the most challenging task is the harmonization if the characteristic exists in the same or a similar form in different source systems (for example, material number from different ERP systems). The other characteristics from the new DataSource, which you already added in the data acquisition layer, are also harmonized and stored in the EDW layer to be prepared for new requirements.

Then, you decide in which InfoCubes you need the new characteristics and extend the ADM layer in these InfoCubes and MultiProviders.

Bear in mind that the new fields are usually not available until the next data delivery. If it is a completely new component, this is usually clear to the business department and is not a problem. For characteristics that have been used in the past but are now available in SAP NetWeaver BW, you must consider — together
Index

A
Aggregate, 141, 158
Aggregation, 141
Alternative storage, 22
Application, 21
Application component, 119
Application-neutral, 34
Architected data mart layer, 24
Archive, 22
Attribute
  Time-dependent, 125, 131
  Time-independent, 131

B
Basic cost, 26, 43
Basic key figure, 40
Batch event, 170
BI Accelerator, 143
BI-integrated planning, 146
Bitmap index, 146
Break-even, 26
B-tree index, 146
Business content, 13, 28, 156, 185

C
CIF, 19
Cleansing, 28
Compression, 139
Construction plan, 21
Control program, 167
Corporate data, 21
Corporate Information Factory (CIF), 19
Corporate memory, 30, 44
Customizing table, 162

D
Data acquisition, 20
Data acquisition layer, 24, 27, 44
Database view, 116, 163
Data currency, 151
Data delivery, 20
Data mart, 22
Data mining warehouse, 22
Data model, 12
Data protection, 13, 25, 28
DataSource, 27, 119, 183
  OEMployee_Attr, 172
  OHR_PA_0, 166
  OHR_PA_1, 165
  OHR_PY_1, 28, 157
  OPERSON_ATTRIB, 172
Data store object, 21
  Standard, 34
  Write-optimized, 29
Data target, 183
Data transfer process, 183
Data update, 154
Data warehouse, 12, 22
Data warehousing workbench, 33
Deletion of overlapping requests, 36
Delta procedures, 43
Direct access, 44
Dummy transformation, 33

E
EDW, 19, 22, 159
  Classic, 23
EDW layer, 24, 30
End routine, 31
Enterprise data warehouse, 19
Enterprise data warehouse (EDW), 19
E table, 140
ETL process, 46, 123, 183
Index

Event, 170
Expert routine, 31, 42, 123
External world, 20, 21
Extract once, deploy many, 19, 27
Extractor, 13, 27, 116
Extrapolation, 147

F
Fact table, 139
Field symbol, 183
Flexibility, 49
F table, 140
Function module
  BP_EVENT_RAISE, 164, 170

G
Gap formation, 32

H
Harmonization, 31, 43
Harmonization tier, 25
High cardinality, 146
Historical correctness, 152
Historical stability, 152
Historization, 129
History, 152
HR management, 113

I
InfoCube, 183
  Payroll data, 15
InfoObject, 183
InfoPackage, 183
InfoProvider, 183
Infotype, 113
  0001 (organizational assignment), 113
  0002 (personal data), 116, 128
  0004 (challenge), 113, 128

K
Key date, 162

L
Legacy data, 39
Line item dimension, 145
Load control, 151, 163
Long-term storage, 30, 49

M
Metadata, 22
Metadata management, 20
Migration, 14
MultiProvider, 37

N
New requirement, 26, 38, 113, 115

O
Operational data store, 21, 24

P
Partitioning, 38, 48, 143
Performance, 15, 29, 37, 46, 129, 135
Persistent staging area, 25, 44, 184
Plausibility check, 30
Primary storage management, 20
Process chain, 155
Program
  ZBI_PC_FINISH, 164, 168
  ZBI_PC_START, 164, 175
Prototyping, 14
### Index

**Q**
- Query, 184
- Query designer, 184

**R**
- Redundancy, 44, 47, 49
- Remodeling Toolbox, 42
- Repartitioning, 145
- Report-report interface, 43, 136
- Request, 141
  - Overlapping, 36
- Routine, 36, 160

**S**
- SAP Business Explorer Suite, 184
- SAP Course
  - BW350, 32
  - BW360, 14
- SAP Training
  - BW360, 15, 134
- Selection condition, 123
- Separation, 144
- Show Data Flow, 33
- Single point of truth, 30

- Sparsity, 32
- Staging layer, 25
- Standalone solution, 13
- Start process, 156

**T**
- Target group, 17
- Time-dependent, 125, 131
- Time-independent, 131
- Transaction
  - MM01, 138
  - RSA1, 119, 140, 155
  - RSBBS, 137
  - RSMON, 155
  - RSPC, 155
  - SBIW, 117
  - SM62, 170
- Transformation, 184
- Truth
  - Current, 131
  - Historical, 133
  - Real, 133

**W**
- Web application, 184