

Zweitveröffentlichung



Sinz, Elmar J.; Ulbrich-vom Ende, Achim

Architektur von Data-Warehouse-Systemen

Datum der Zweitveröffentlichung: 10.07.2024

Akzeptiertes Manuskript (Postprint), Beitrag in Sammelwerk

Persistenter Identifikator: urn:nbn:de:bvb:473-irb-964162

Erstveröffentlichung

Sinz, Elmar J.; Ulbrich-vom Ende, Achim (2010): „Architektur von Data-Warehouse-Systemen“. In: Peter Chamoni (Hrsg.), Analytische Informationssysteme : Business Intelligence-Technologien und -Anwendungen, 4. Aufl., Berlin: Springer, S. 175-196.

Rechtehinweis

Dieses Werk ist durch das Urheberrecht und/oder die Angabe einer Lizenz geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die für Sie geltende Gesetzgebung zum Urheberrecht und/oder durch die Lizenz erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

Für dieses Dokument gilt das deutsche Urheberrecht.

Architektur von Data-Warehouse-Systemen

ELMAR J. SINZ; ACHIM ULBRICH-VOM ENDE

Abstract

In den letzten Jahren haben sich Data-Warehouse-Systeme als Hilfsmittel zur Unterstützung des Managements betrieblicher Systeme etabliert. Bei der Wahl der geeigneten Architektur für ein Data-Warehouse-System ist dessen Einsatzform im Unternehmen zu berücksichtigen. Hierzu gehören aus fachlicher Sicht insbesondere der Aufbau der zu unterstützenden Organisations- und Führungsstruktur sowie spezifische Anforderungen an die Datenbereitstellung.

Im vorliegenden Beitrag werden zunächst die Funktion des Data-Warehouse als Hilfsregelstrecke im Lenkungssystem des Unternehmens und die daraus resultierenden Gestaltungsmerkmale für die Systemarchitektur herausgearbeitet. Anschließend wird auf der Grundlage einer allgemeinen Darstellung des Aufbaus und der Funktionsweise eines Data-Warehouse-Systems ein generischer Rahmen zur Beschreibung von Architekturformen für Data-Warehouse-Systeme entwickelt.

Im Gegensatz zu verbreiteten Ansätzen zur Beschreibung von Architekturformen für Data-Warehouse-Systeme, welche die einzelnen Datenbestände und deren Beziehungen in den Mittelpunkt stellen, ist der generische Rahmen als Hierarchie von Softwareschichten konzipiert. Dadurch kann der von den einzelnen Softwareschichten bereitgestellte Funktionsumfang unabhängig von der Realisierungsform der zugrunde liegenden Datenhaltung beschrieben werden. Anhand des generischen Rahmens werden ausgewählte Architekturformen für Data-Warehouse-Systeme mit Bezug zu den eingangs genannten Anforderungen vorgestellt. Hierbei wird aufgezeigt, dass die Unterschiede zwischen den einzelnen Architekturen aus der Konfiguration der Schichten sowie aus der Realisierungsform der Datenhaltung innerhalb der Schichten resultieren.

Inhalt

1	Das Data-Warehouse im Lenkungssystem der Unternehmung	3
2	Aufbau und Funktionsweise eines Data-Warehouse-Systems	6
3	Ein generischer Rahmen für die Beschreibung der Architektur von Data-Warehouse-Systemen	7
4	Persistenzaspekte in Data-Warehouse-Systemen	10
5	Verteilungsaspekte von Data-Warehouse-Systemen	14
	5.1 Zentrales Data-Warehouse-System	14
	5.2 Unabhängige Data-Marts	15
	5.3 Hub-and-Spoke-Architektur	17
	5.4 Data-Mart-Bus-Architektur	18
	5.5 Föderierte und hierarchische Data-Warehouse-Systeme	19
6	Zusammenfassung	21

1 Das Data-Warehouse im Lenkungssystem der Unternehmung

Die Anforderungen an die Gestaltung der Architektur eines Data-Warehouse-Systems für ein Unternehmen (bzw. Unternehmensbereich, Unternehmensverbund) ergeben sich primär aus der Funktion, welche ein Data-Warehouse im Lenkungssystem dieses betrieblichen Systems erfüllt.

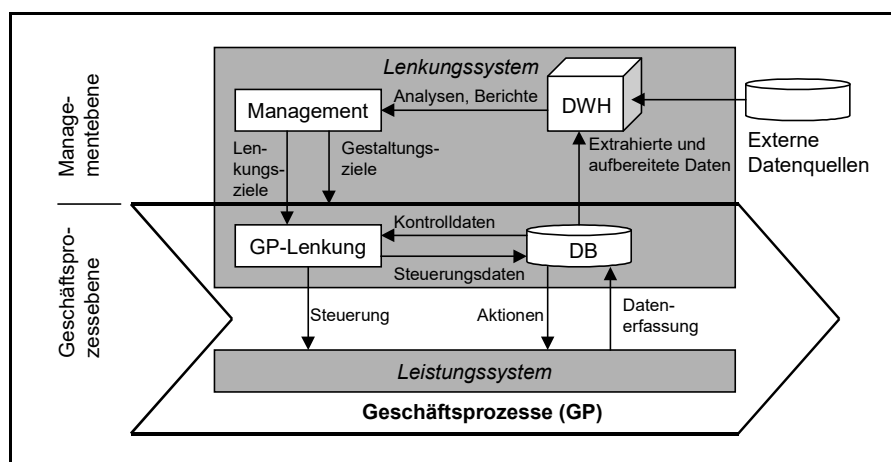


Abb. 1: Data-Warehouse im Regelkreis eines betrieblichen Systems [Sinz01, 309]

Hierzu zeigt Abb. 1 die regelkreisorientierte Interpretation eines betrieblichen Systems. Danach wird ein betriebliches System in zwei Teilsysteme gegliedert, ein Lenkungssystem und ein Leistungssystem [FeSi08, 5 f.]. Das Lenkungssystem (Regler), auch als Führungssystem [Küpp97, 15] bezeichnet, dient der Planung, Steuerung und Kontrolle der betrieblichen Leistungserstellung. Das Leistungssystem (Regelstrecke) dient der Durchführung der Leistungserstellung. Innerhalb des Lenkungssystems lassen sich weiter zwei Ebenen von Reglern unterscheiden:

- Aufgabe der GP-Lenkung (Regler) ist die unmittelbare Planung, Steuerung und Kontrolle des Leistungssystems (Regelstrecke) im Rahmen der laufenden Geschäftsprozesse. GP-Lenkung und Leistungssystem gehören zur Geschäftsprozessebene eines betrieblichen Systems.
- Aufgabe des Managements (Regler) ist die Lenkung und Gestaltung des gesamten betrieblichen Systems. Hierzu gehören insbesondere die Lenkung der Geschäftsprozesse sowie die Gestaltung der Gesamtheit der Geschäftsprozesse (Regelstrecke) über Lenkungsziele bzw. Gestaltungsziele.

Auf beiden Regelkreisebenen erfassen die Regler die Zustände und Strukturen der jeweiligen Regelstrecke nicht direkt, sondern indirekt über eine Hilfsregelstrecke in Form eines Modells der jeweiligen Regelstrecke. Diese Hilfsregelstrecke wird

auf der Geschäftsprozessebene in Form der operativen Datenbanken (DB) gebildet. Auf der Managementebene stellt das Data-Warehouse (DWH) die Hilfsregelstrecke dar. Das DWH enthält extrahierte und aufbereitete Daten aus den operativen Datenbanken sowie aus externen Datenquellen.

Die Unterschiede zwischen DB und DWH ergeben sich aus den unterschiedlichen Aufgabenschwerpunkten von GP-Lenkung und Management:

- Die DB stellt ein aktuelles Abbild der laufenden Geschäftsprozesse dar und beinhaltet die Stammdaten des Unternehmens (Kunden, Artikel, Lieferanten usw.), die Transaktionsdaten (Aufträge, Bestellungen, Rechnungen usw.) sowie weitere geschäftsprozessbezogene Daten. Im Regelkreis der Geschäftsprozessebene stellen die Inhalte der DB Kontrolldaten dar, auf deren Basis die GP-Lenkung ihre Planung durchführt und Steuerungsdaten generiert. Die Steuerungsdaten werden entweder in die DB übertragen und führen von dort zu Aktionen im Leistungssystem, oder es erfolgt eine direkte Steuerung des Leistungssystems (z. B. über Anweisungen). Der Abgleich zwischen dem Zustand des Leistungssystems und dem Inhalt der DB wird über Aktoren und Sensoren durchgeführt. Aktoren lösen Aktionen auf dem Leistungssystem aus (z. B. löst eine Versandanweisung die Kommissionierung einer Warensendung im Lager aus). Sensoren erfassen Veränderungen im Leistungssystem und übermitteln diese in die DB (z. B. wird ein Zahlungseingang eines Kunden in Form einer entsprechenden Finanzbuchung in der DB erfasst). Die Nutzung der DB im Rahmen der GP-Lenkung korrespondiert mit dem Prinzip des **Online Transaction Processing (OLTP)**.
- Im Gegensatz dazu beinhaltet das DWH entscheidungsrelevante Daten zur Unterstützung der Aufgaben des Managements. Diese Daten werden aus der DB extrahiert und aufbereitet; häufig werden zusätzlich auch Daten aus externen Datenquellen einbezogen. Die Daten sind geschäftsprozessübergreifend integriert und können ausgehend von den elementaren Daten einzelner Geschäftsvorfälle entlang unterschiedlicher Merkmalsdimensionen verdichtet werden. Insbesondere weisen die Inhalte des DWH einen Zeitbezug auf, anhand dessen einzelne Merkmalsausprägungen in ihrem zeitlichen Verlauf analysiert werden können. Aus dem DWH werden diese Daten dem Management in Form von Analysen und Berichten zur Verfügung gestellt. Ergebnis der Planungs- und Entscheidungsaufgaben des Managements sind Lenkungsziele als Führungsgrößen für die GP-Lenkung sowie Gestaltungsziele für die gesamte Geschäftsprozessebene. Die Nutzung des DWH im Rahmen des Managements korrespondiert mit dem Prinzip des **Online Analytical Processing (OLAP)**.

Eine der ersten und meist zitierten Charakterisierungen der Merkmale eines DWH stammt von Inmon: „A data warehouse is a subject oriented, integrated, non-volatile, and time variant collection of data in support of management’s decisions“¹ [Inmo96, 33]. Diese Merkmale korrespondieren unmittelbar mit der Funktion des DWH im Lenkungssystem der Unternehmung:

- **Subject oriented:** Aufbau und Inhalt des DWH orientieren sich an den für das Management relevanten Planungs- und Entscheidungsfeldern.
- **Integrated:** Ein DWH integriert im Allgemeinen Daten aus mehreren Geschäftsprozessen und bezieht externe Datenquellen mit ein.
- **Non-volatile:** Daten, die einmal in das DWH eingestellt wurden, werden nicht mehr verändert oder gelöscht. Vielmehr werden die Inhalte des DWH zu festgelegten Aktualisierungszeitpunkten sukzessive erweitert.
- **Time variant:** Um Entwicklungen im Zeitablauf analysieren zu können, enthält ein DWH historisierte Daten.

Neben diesen allgemeinen Merkmalen eines DWH-Systems lassen sich folgende spezifische Gestaltungsmerkmale aus der Interpretation des DWH als Hilfsregelstrecke ableiten [SBPU01]:

- **Reichweite des DWH-Systems:** Dieses Merkmal betrifft den Grad der Vollständigkeit der Hilfsregelstrecke bezüglich des zugrunde liegenden Leistungssystems. Die Reichweite muss auf die zu unterstützenden Planungs- und Entscheidungsaufgaben ausgerichtet sein.
- **Abstimmung des DWH-Systems mit der Organisations- und Führungsstruktur des Unternehmens:** Die Organisations- und Führungsstruktur des Unternehmens (z. B. Spartenorganisation, funktionale Organisation) muss sich in der DWH-Architektur widerspiegeln. Besonders gilt dies für verteilte Organisationen (z. B. Netzwerkorganisationen) mit autonomen oder teilautonomen Organisationseinheiten, die ggf. untereinander im Wettbewerb stehen können. Um das verteilte Lenkungssystem dieser Unternehmen adäquat mit Informationen zu versorgen, kann es sinnvoll sein, das DWH-System als verteiltes System zu gestalten.

Weitere Merkmale eines DWH-Systems resultieren aus speziellen Anforderungen an die Aktualität der bereitzustellenden Daten (z. B. Einbeziehung eines Operational-Data-Store; siehe Abschnitt 4) sowie an die Entkopplung des DWH-Systems von den internen Datenquellen des Unternehmens.

Der weitere Inhalt des Buchkapitels ist wie folgt: In Abschnitt 2 werden zunächst der Aufbau und die Funktionsweise eines DWH-Systems allgemein beschrieben. In Abschnitt 3 wird dann ein generischer Rahmen für die Beschreibung spezieller Architekturformen von DWH-Systemen eingeführt. Abschnitt 4 ergänzt den Beschreibungsrahmen um Persistenzaspekte in DWH-Systemen. In Abschnitt 5 werden schließlich verschiedene Architekturformen für DWH-Systeme vorgestellt, die sich u. a. in ihrer Verteilung, im Integrationsgrad und in der Integrationsform der Teil-DWH-Systeme unterscheiden.

2 Aufbau und Funktionsweise eines Data-Warehouse-Systems

Der idealtypische Aufbau eines DWH-Systems ist in Abb. 2 dargestellt (siehe auch [BaGü08, 33 ff.], [GlCh06, 154 ff.]). Dieser folgt der Metapher eines zweistufigen industriellen Produktionssystems mit den Produktionsstufen Datenerfassung und Datenbereitstellung sowie einem dazwischen geschalteten Lager (Datenhaltung). Die Funktionsbausteine Datenerfassung, Datenhaltung und Datenbereitstellung werden im Folgenden kurz erläutert:

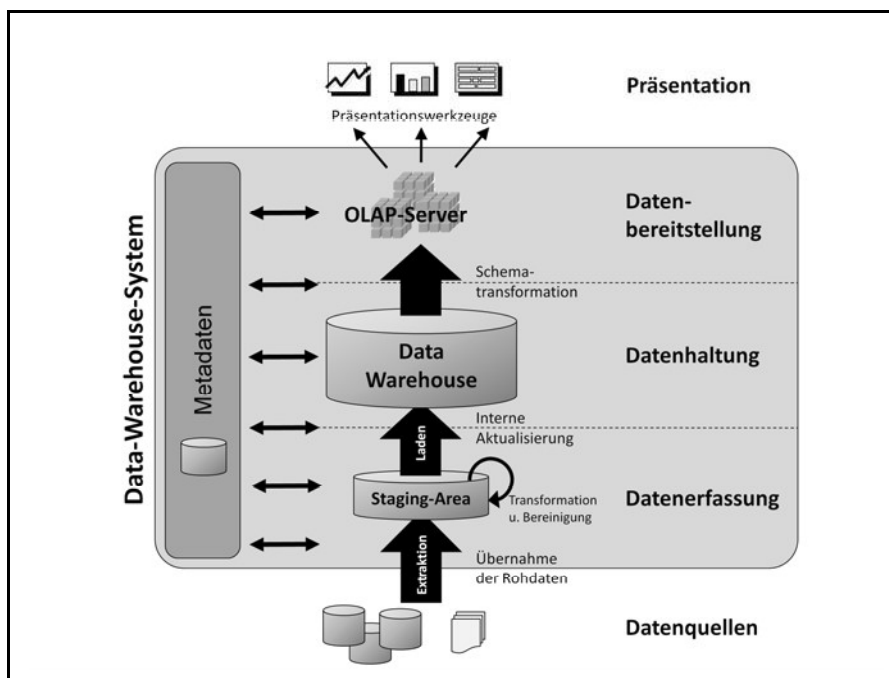


Abb. 2: Idealtypischer Aufbau eines Data-Warehouse-Systems

- **Datenerfassung:** Dieser Funktionsbaustein führt den ETL-Prozess durch, bestehend aus den Schritten Extraktion (selektive Übernahme und Filterung der Rohdaten aus den internen und externen Datenquellen), Transformation (syntaktische und semantische Bereinigung der Daten) und Laden (Aufbereitung der Daten und Übernahme in das DWH) (vgl. u. a. [KeFi06]). Ein als „Staging-Area“ bezeichneter Arbeitsbereich dient der Speicherung von Zwischenergebnissen im Rahmen des ETL-Prozesses.

- **Datenhaltung:** Der Funktionsbaustein verwaltet das eigentliche DWH, d. h. die Hilfsregelstrecke auf der Managementebene des Lenkungssystems der Unternehmung. Das DWH wird dabei im Allgemeinen als persistenter Datenbestand realisiert, kann aber auch in virtualisierter, nicht-persistenter Form bereitgestellt werden (siehe Abschnitt 4).
- **Datenbereitstellung:** Diese Schicht stellt die Datenbestände für die nutzende Umgebung des DWH-Systems (Präsentationswerkzeuge) in Form von multidimensionalen Datenstrukturen bereit. Der Funktionsbaustein wird im Allgemeinen durch einen OLAP-Server realisiert. Bezüglich der Funktionsweise von OLAP-Servern lassen sich als Grundformen das relationale OLAP (ROLAP) und das multidimensionale OLAP (MOLAP) unterscheiden (siehe z. B. [GlCh06]).

Die Umgebung des DWH-Systems (Abb. 2) umfasst auf der „Beschaffungsseite“ die **Datenquellen** für das DWH. Hierzu gehören insbesondere die Datenbanken der operativen Anwendungssysteme sowie externe Datenquellen (siehe z. B. [Fisc01, 107 ff.]), wie etwa öffentliche Datenbanken.

Auf der „Absatzseite“ erfolgt die **Präsentation** der Daten durch Darstellung und Auswertung der Anfrageergebnisse in Form von interaktiven Berichten und Diagrammen. Werkzeuge hierfür sind insbesondere OLAP-Standardberichte für bestimmte betriebliche Funktionsbereiche, OLAP-Berichtsgeneratoren sowie Erweiterungen von Standard-Bürosoftware (siehe z. B. [ChGl00, 364 ff.]). Querschnittlich zu den beschriebenen Schichten liegt die Verwaltung der Metadaten des DWH-Systems (siehe z. B. [BaGü08, 72 ff.]). Die Metadaten beschreiben insbesondere die Datenstrukturen des DWH sowie die verschiedenen Transformationen von der Datenerfassung bis zur Datenbereitstellung. Die Verwaltung der Metadaten kann in einem zentralen Repository erfolgen oder auf die einzelnen Funktionsbausteine verteilt sein.

3 Ein generischer Rahmen für die Beschreibung der Architektur von Data-Warehouse-Systemen

Wie in Abschnitt 1 dargelegt, ist die Architektur eines DWH-Systems mit der Organisations- und Führungsstruktur des Unternehmens abzustimmen. Die Heterogenität der Organisations- und Führungsstrukturen macht dabei unterschiedliche Architekturformen von DWH-Systemen notwendig. Auf der Basis des in Abschnitt 2 erläuterten allgemeinen Aufbaus eines DWH-Systems wird nun ein generischer Rahmen eingeführt, der es erlaubt, diese unterschiedlichen Architekturformen von DWH-Systemen in vereinheitlichter Form zu beschreiben:

- Der generische Rahmen gliedert die Architektur eines DWH-Systems in eine Hierarchie von logischen Softwareschichten. Jede Schicht kapselt bestimmte Funktionen oder Datenbestände und weist Schnittstellen zu benachbarten Softwareschichten auf.
- Die Architektur eines konkreten DWH-Systems entsteht durch Instanziierung des generischen Rahmens, d. h. durch Festlegung, in welcher Form und in wie vielen Exemplaren eine bestimmte Schicht in der jeweiligen Architektur vorkommt und wie die einzelnen Exemplare konkret konfiguriert sind.
- Grundsätzlich gilt, dass jede Schicht in Form von mindestens einer Instanz notwendig ist, um den gesamten Funktionsumfang eines DWH-Systems (siehe Abschnitt 2) bereitzustellen. Freiheitsgrade bestehen allerdings hinsichtlich der Implementierung der logischen Softwareschichten als physische Softwarekomponenten. Zum Beispiel könnten in einer konkreten Architektur benachbarte Schichten, falls diese nur jeweils in einem einzigen Exemplar auftreten, zu einer Softwarekomponente zusammengefasst sein.

Im Vergleich zu Architekturmodellen von DWH-Systemen, welche die Beziehungen zwischen den beteiligten Datenbeständen in den Mittelpunkt stellen (siehe z. B. [SeSi05], [LaSt08]), besitzt eine Spezifikation der Softwareschichten den Vorteil, dass deren Funktionen unabhängig von der Realisierungsform (persistiert, virtualisiert) der zugrunde liegenden Datenbestände beschrieben werden können.

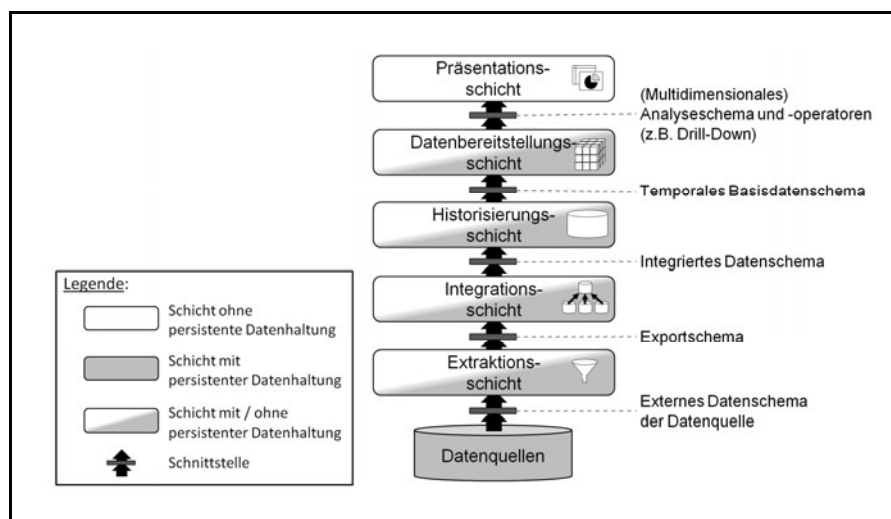


Abb. 3: Schichtenmodell als generischer Rahmen für die Architektur von Data-Warehouse-Systemen

Der generische Rahmen ist in Form eines Schichtenmodells (Abb. 3) dargestellt. Die einzelnen Schichten sind sachzielbezogen anhand ihres Funktionsschwerpunkts abgegrenzt. Der Sachzielbezug kommt in der Bezeichnung der

jeweiligen Schicht zum Ausdruck. Schnittstellen zwischen benachbarten Schichten werden in Form von Datenschemata mit zugehörigen Operatoren beschrieben. Die einzelnen Schichten werden nun kurz dargestellt und mit den Funktionsbausteinen aus Abschnitt 2 in Beziehung gesetzt:

- **Datenquellen:** Die unterste Schicht des generischen Rahmens bilden die unternehmensinternen und die externen Datenquellen (vgl. Abb. 2). Die Datenquellen sind kein Bestandteil des DWH-Systems, sie stellen die Datengrundlage für den Aufbau des DWH bereit. Die Anbindung von Datenquellen an ein DWH-System schränkt deren Entwurfs- und Änderungsautonomie ein, da Schemaänderungen der Datenquellen mit dem DWH-System abgestimmt und inhaltliche Änderungen an das DWH-System weitergeleitet werden müssen (vgl. auch [LeNa07, 58]). Formale Anforderungen an Datenquellen betreffen u. a. die Verfügbarkeit und die Qualität der Daten sowie die rechtliche Zulässigkeit ihrer Verwendung [BaGü08, 41]. Die von den Datenquellen angebotenen Schnittstellen entsprechen jeweils einem **externen Schema** der für das DWH-System bereitgestellten Datenstrukturen.
- **Extraktionsschicht:** Die Extraktionsschicht führt den Zugriff auf die Datenquellen durch und stellt die extrahierten Daten für die weitere Verarbeitung im DWH-System bereit. Externe Datenquellen verfügen im Allgemeinen über vordefinierte Schnittstellen für die Datenextraktion. Interne Datenquellen, insbesondere operative Anwendungssysteme, realisieren das externe Schema für die Datenbereitstellung alternativ auf Basis der Datenhaltungsschicht, der Anwendungsschicht oder der Präsentationsschicht (siehe auch [Lehn03, 142]). In Abhängigkeit von der jeweiligen Realisierungsform sind die zugehörigen Extraktionsfunktionen zu gestalten. Weitere Gestaltungsmerkmale der Extraktionsschicht betreffen den Umfang der Datenextraktion (vollständiger Datenbestand oder Veränderungen seit der letzten Extraktion) [Lehn03, 140], die Identifikation der Datenänderungen in den Datenquellen [BaGü08, 49 f.] sowie den Zeitpunkt der Datenextraktion [BaGü08, 51 f.]. Die Extraktionsschicht stellt als Schnittstelle ein **Exportschema** der zur Aufnahme in das DWH vorgesehenen Roh-Datenstrukturen zur Verfügung.
- **Integrationsschicht:** Die Integrationsschicht führt die syntaktische und semantische Bereinigung der über das Exportschema zur Verfügung gestellten Daten durch und führt die ggf. aus verschiedenen Datenquellen stammenden Daten zusammen. Als Schnittstelle stellt die Integrationsschicht ein **integriertes Datenschema** der zur Aufnahme in das DWH vorgesehenen Datenstrukturen zur Verfügung.

- **Historisierungsschicht:** Gegenstand der Historisierungsschicht ist der Aufbau des Zeitbezugs der Datenbestände (Historisierung). Die zeitbezogenen Datenbestände werden mit dem bestehenden Datenvorrat des DWH zusammengeführt. Die Historisierungsschicht stellt diesen Datenvorrat über ein **temporales Basisdatenschema** zur Verfügung. Das Basisdatenschema ist unabhängig von der Realisierungsform der Datenbereitstellung (siehe z. B. [BaGü08, 53 ff.]). Abhängig von der Datenhaltung in der Historisierungsschicht bietet diese Schnittstelle entweder einen Zugriff auf den vollständig historisierten Datenbestand oder lediglich auf die Änderungen zur Historisierung des Datenbestands in der Datenbereitstellungsschicht (siehe Abschnitt 4).
- **Datenbereitstellungsschicht:** Die Datenbereitstellungsschicht dient dem Aufbau und der Bereitstellung der multidimensionalen Sichten auf den Datenvorrat des DWH. Die Realisierung erfolgt auf der Basis von ROLAP oder MOLAP [GlCh06] bzw. direkt auf der Basis von SQL (siehe Abschnitt 4). Die Datenbereitstellungsschicht bietet an ihrer Schnittstelle ein **multidimensionales Analyseschema** gemäß einem multidimensionalen Datenmodell mit zugehörigen Operatoren (Drill-down, Roll-up, Slice, Dice usw.) zur Durchführung von Analysen an. Als spezielle Anfragesprache kann z. B. MDX (Multidimensional Expressions [Micr09]) eingesetzt werden.
- **Präsentationsschicht:** Die Präsentationsschicht wird im Allgemeinen nicht als Bestandteil des DWH-Systems angesehen. Hier werden verschiedene Präsentationswerkzeuge genutzt, welche die GUI für die Nutzer des DWH realisieren (siehe Abschnitt 2).

Mit Bezug zu Abschnitt 2 realisieren die Extraktionsschicht, die Integrationsschicht und Teile der Historisierungsschicht zusammen den Funktionsbaustein Datenerfassung. Teile der Historisierungsschicht realisieren gemeinsam mit der Datenbereitstellungsschicht den Funktionsbaustein Datenhaltung (vgl. Abb. 2). Wie Abb. 3 zeigt, können die Schichten mit oder ohne persistente Datenhaltung ausgestattet sein. Diese Persistenzaspekte werden im folgenden Abschnitt behandelt.

4 Persistenzaspekte in Data-Warehouse-Systemen

In Abschnitt 3 wurde ein generischer Rahmen für die Beschreibung der Architektur von DWH-Systemen anhand eines Schichtenmodells eingeführt. Alle Schichten mit Ausnahme der stets als persistent angenommenen Datenquellen und der nicht persistenten Präsentationsschicht können dabei mit und ohne persistente Datenhaltung ausgestattet sein. Im Folgenden werden die korrespondierenden Realisierungsformen (virtualisiert bzw. persistiert) für die Datenbestände der einzelnen Softwareschichten diskutiert:

- Die Notwendigkeit einer persistenten Datenhaltung innerhalb der **Extraktionsschicht** ist abhängig von der Fähigkeit der Quellsysteme, die Daten in der erforderlichen Form (siehe Abschnitt 3) für das Laden des DWH zur Verfügung zu stellen. Beispielsweise können die Nettoänderungen einer Datenquelle innerhalb der Extraktionsschicht erfasst und für die Übertragung an das DWH gespeichert werden. Häufig wird hierzu die Datenquelle um entsprechende Tabellen erweitert. Diese Datenhaltung wird beispielsweise in SAP R/3-Quellsystemen als „Delta-Queue“ bezeichnet und dient der Aktualisierung angeschlossener SAP Business Information Warehouses [Mehr05, 215 ff.].
- Die Konsolidierung und Aufbereitung der extrahierten Daten in der **Integrationschicht** erfolgt in der Regel innerhalb einer separaten Datenbasis (u. a. [BaGü08, 36]), dem sogenannten **Arbeitsbereich** bzw. der **Staging-Area** (Abb. 2). Hierfür ist lediglich eine temporäre Datenhaltung erforderlich, da die integrierten Daten anschließend an die Historisierungsschicht weitergegeben werden. Unter Umständen, z. B. zur Rekonstruktion des DWH nach der Anpassung von Transformationsroutinen, kann jedoch eine dauerhafte Speicherung der Datenextrakte sinnvoll sein. Diese Funktionalität wird beispielsweise im SAP Business Information Warehouse in der „Persistent Staging Area“ angeboten ([ChGH05, 49], [Mehr05, 287 ff.]).
Zusätzlich besteht die Möglichkeit, das Ergebnis der Datenintegration (ggf. nur teilweise) in einem sogenannten **Operational-Data-Store**² (ODS) persistent für Auswertungen zur Verfügung zu stellen [Muck06, 136]. Die Merkmale eines ODS beschreibt INMON [Inmo99] folgendermaßen: „An operational data store is a subject-oriented, integrated, volatile, current-valued, detailed-only collection of data in support of an organization's need for up-to-the-second, operational, integrated, collective information“. Damit unterscheidet sich ein ODS von einem DWH im Wesentlichen durch den fehlenden Vergangenheitsbezug, eine häufigere Aktualisierung und ggf. eine feinere Datengranularität. Die Realisierung eines ODS ist zweckmäßig, wenn die Aktualität der Daten im DWH für bestimmte Entscheidungsbereiche nicht ausreichend ist, aber die Aktualisierungsfrequenz und die zeitliche Granularität im DWH nicht erhöht werden sollen. Durch die zeitliche Entkopplung der Prozesse zur Datenintegration und zur Historisierung können im ODS aktuellere Daten für zeitnahe Auswertungen bereitgestellt werden, ohne die Datenhaltung im DWH zu beeinträchtigen. Beinhaltet ein DWH beispielsweise lediglich monatlich historisierte Daten, für einige Anwendungen werden jedoch tagesaktuelle Daten vom laufenden Monat benötigt, dann würde der Datenbestand im ODS täglich mit aktuellen Daten aus den Quellsystemen überschrieben. Dadurch kann die Zeitspanne zwischen zwei Aktualisierungszeitpunkten des DWH überbrückt werden [Muck06, 136]. Im Rahmen des **Realtime-** bzw. **Near-Realtime-Data-Warehousing** [Sche06] wurde in den letzten Jahren mit dem Einsatz leistungsstarker ETL- bzw. EAI-Werkzeuge die zeitnahe Abbildung aktueller Daten direkt im DWH angestrebt. Dadurch wird das Konzept des ODS obsolet.

- Für die Datenhaltung in der Historisierungs- und Datenbereitstellungsschicht ergeben sich, abhängig von der Verteilung der Systemkomponenten (siehe Abschnitt 5) und der Gestaltung der Datenbereitstellungsschicht (Abb. 4), unterschiedliche Kombinationsmöglichkeiten. In der **Historisierungsschicht** erfolgt der Aufbau der zeitbezogenen Datenbestände innerhalb des DWH (Abschnitt 3). Die hierfür notwendige dauerhafte Datenhaltung kann entweder in der Historisierungs- oder der Datenbereitstellungsschicht erfolgen. Die Datenhaltung in der Historisierungsschicht wird häufig auch als „Basisdatenbank“ bezeichnet [BaGü08, 53 ff.]. Sie erfolgt in der Regel auf Grundlage relationaler Datenstrukturen (3NF-Schema) und ist im Gegensatz zur Datenhaltung in der Datenbereitstellungsschicht nicht für spezifische Analysewerkzeuge optimiert. Die Notwendigkeit einer eigenständigen Datenhaltung innerhalb der Historisierungsschicht ist abhängig vom Bedarf, das temporale Basisdatenschema als persistente Grundlage für eine oder mehrere Datenbereitstellungsschichten zur Verfügung zu stellen (siehe Abschnitt 5.3).

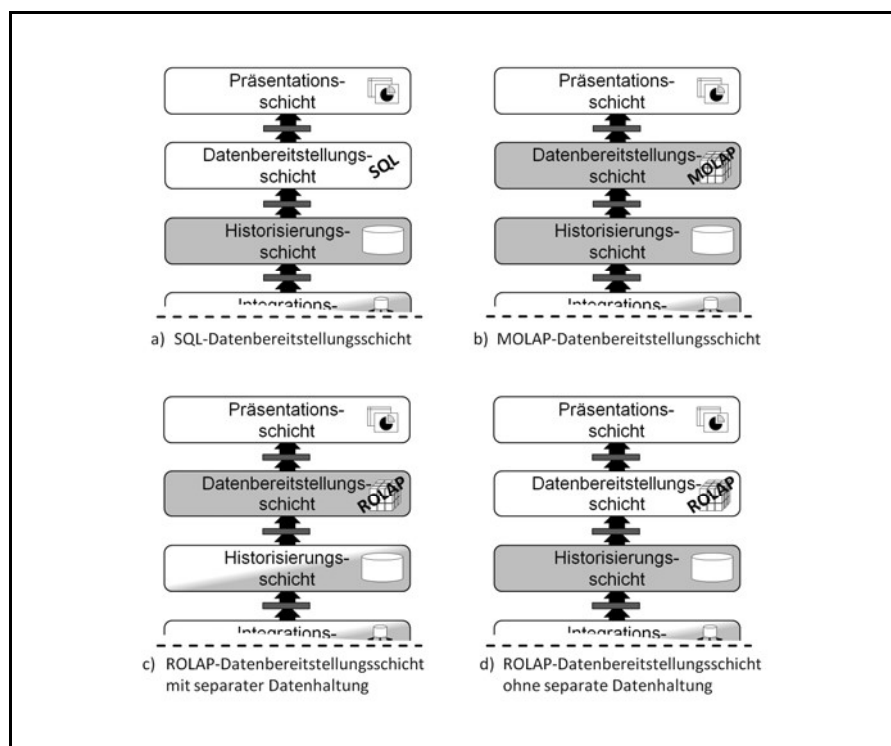


Abb. 4: Alternative Persistenzformen der Datenbereitstellungs- und Historisierungsschicht

- In Abb. 4 sind die Alternativen für die Bereitstellung der Daten im DWH und die daraus resultierenden Anforderungen an die Datenhaltung für die **Datenbereitstellungsschicht** dargestellt:

- a) Erfolgt der Zugriff auf das DWH ausschließlich über eine relationale Anfrageschnittstelle (SQL), dann ist die Datenhaltung in der Historisierungsschicht häufig ausreichend.
- b) Beim Einsatz multidimensionaler OLAP-Werkzeuge (MOLAP) werden die Analysedaten in spezifischen multidimensionalen Datenstrukturen eines multidimensionalen Datenbanksystems innerhalb der Datenbereitstellungsschicht gespeichert [GICH06, 155]. Da eine inkrementelle Aktualisierung häufig nicht unterstützt wird, müssen die Datenstrukturen bei jeder Aktualisierung des DWH mit Hilfe des Datenbestands in der Historisierungsschicht neu aufgebaut werden.
- c) Bei relationalem OLAP wird für die Nutzer eine multidimensionale Sicht auf Grundlage der Daten in einer relationalen Datenbank erzeugt. Häufig kommt hierfür eine eigenständige Datenhaltung innerhalb der Datenbereitstellungsschicht zum Einsatz. Diese basiert in der Regel auf einem Star- oder Snowflake-Schema [Hahn06, 191 ff.] und ist für das eingesetzte Produkt optimiert. Zusätzlich können spezielle Indizes [GICH06, 167] und vorberechnete Werte (Preaggregationstabellen) zur weiteren Verbesserung der Anfrageperformance eingesetzt werden. Eine redundante Datenhaltung in der Historisierungsschicht ist nicht erforderlich, kann jedoch z. B. zur Entkopplung der Ladeprozesse von Vorteil sein.
- d) Einige ROLAP-Werkzeuge unterstützen auch normalisierte Datenschemata. In diesem Fall ist eine separate Datenhaltung in der Datenbereitstellungsschicht nicht zwingend erforderlich. Die multidimensionale Sicht wird dann auf Basis des Datenbestands in der Historisierungsschicht erzeugt. Aufgrund der fehlenden Optimierung der Basisdatenbank für das eingesetzte Analysewerkzeug kann dies insbesondere bei größeren Datenvolumen zu Performanceproblemen führen.

Abhängig von der Verteilung der einzelnen Systemkomponenten ist eine redundante Datenhaltung in der Historisierungsschicht unumgänglich (Abschnitt 5).

Unter dem Begriff **virtuelles DWH-System** (Abb. 5) versteht man in der Literatur ein DWH-System ohne eigenständige Datenhaltung (u. a. [MuBe00, 55 f.]). Die Datenbereitstellung erfolgt hierbei analog zu einem föderierten Datenbanksystem [Conr97] auf Basis der Datenhaltung in den Quellsystemen. Neben einer schlechten Anfrageperformance ergibt sich dadurch der Nachteil, dass die Analyse von Zeitreihen aufgrund der fehlenden historisierten Datenbasis nicht möglich ist. Prinzipiell widerspricht die Architektur eines virtuellen DWH-Systems der Definition eines DWH mit einer eigenständigen physischen Datenhaltung aus Abschnitt 1 und wird daher nicht weiter behandelt.

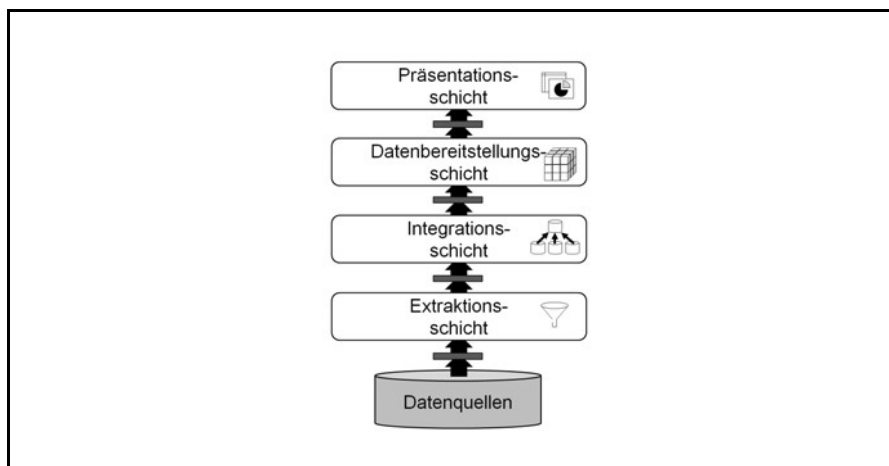


Abb. 5: Virtuelles Data-Warehouse-System

5 Verteilungsaspekte von Data-Warehouse-Systemen

Die Abstimmung mit der Organisations- und Führungsstruktur des Unternehmens erfolgt aus Architektursicht durch die Wahl einer geeigneten Verteilung der Softwareschichten des DWH-Systems. In den folgenden Abschnitten werden hierzu ausgewählte Architekturformen vorgestellt.

5.1 Zentrales Data-Warehouse-System

Die idealtypische Vorstellung der Informationsversorgung aller Führungskräfte in einem Unternehmen spiegelt sich in der Architektur eines **zentralen DWH-Systems** (Abb. 6) wider. Diese Architekturform wird in der Literatur häufig auch als „Enterprise Data Warehouse“ (u. a. [KiRo02, 8]) bezeichnet. Alle entscheidungsrelevanten Daten werden in einer zentralen Datenbasis, dem so genannten „single point of truth“ (SPOT; u. a. [Zeh03, 36]), gespeichert. Die Nutzung erfolgt über die zentrale Datenbereitstellungsschicht. Durch das einheitliche Begriffssystem mit identischen Definitionen für Kennzahlen und Dimensionen wird die Abstimmung und Koordination zwischen den Organisationseinheiten unterstützt. Allerdings können individuelle fachliche und technische Anforderungen einzelner Unternehmensbereiche häufig nicht berücksichtigt werden. Aufgrund der konsolidierten Sicht auf das gesamte Unternehmen ist diese Architekturform für Unternehmen mit einer (teil-) autonomen Organisations- und Führungsstruktur eher ungeeignet.

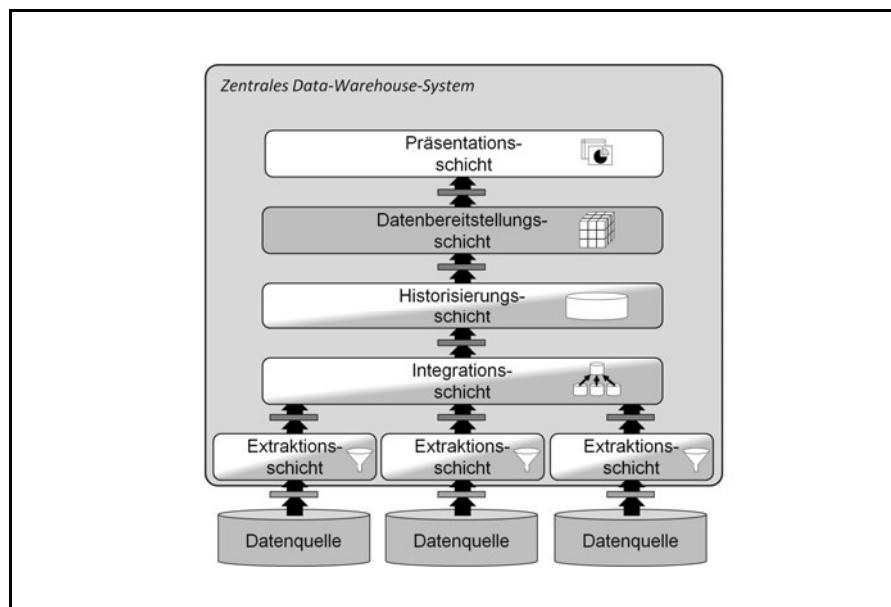


Abb. 6: Zentrale Data-Warehouse-Architektur

Die Implementierung eines zentralen DWH-Systems ist insbesondere für große Unternehmen sehr aufwändig (vgl. u. a. [ChDa97, 518], was einem schnellen Aufbau der Informationsversorgung entgegen steht. Ein Großteil des Aufwands besteht in der unternehmensweiten Abstimmung des einheitlichen Begriffssystems und der Realisierung der konsolidierten Datenbasis. Auf die Implementierung einer Basisdatenbank in der Historisierungsschicht wird aufgrund des Datenumfangs und -volumens meist verzichtet.

Die zentrale Architektur erleichtert grundsätzlich den Betrieb und die Pflege des Systems. Beispielsweise stehen Anpassungen am DWH, z. B. aufgrund von Änderungen an den Datenstrukturen der Quellsysteme, allen Nutzern zeitgleich zur Verfügung. Der zentrale Aufbau des Systems stellt jedoch hohe Anforderungen an die zugrunde liegenden Systemkomponenten [LaSt08, 149]. Einerseits soll die Datenbereitstellungsschicht und die Datenhaltung das geforderte Antwortzeitverhalten für den gleichzeitigen Datenzugriff der Nutzer verwirklichen. Andererseits muss die Aktualisierung des DWH in einem festgelegten Zeitfenster erfolgen, um sowohl die Nutzer als auch die Quellsysteme nicht unnötig zu beeinträchtigen.

5.2 Unabhängige Data-Marts

Im Gegensatz zur zentralen Architektur dienen **unabhängige Data-Marts** zur schnellen Realisierung der Informationsversorgung einzelner Organisationseinheiten eines Unternehmens. Unter dem Begriff Data-Mart versteht man im Allgemeinen ein DWH mit beschränkter Reichweite, das auf den individuellen Informati-

onsbedarf eines Teilbereichs des Unternehmens oder auf einen eingeschränkten Themenbereich ausgerichtet ist (vgl. [ChDa97, 518], [Golf09]). Eine Architektur mit mehreren unabhängigen Data-Marts stellt somit ein vollständig verteiltes System dar (Abb. 7). Selbst die Zugriffe auf einzelne Datenquellen können vollständig unabhängig voneinander erfolgen. Aufgrund der spezifischen Ausrichtung und individuellen Entwicklung der einzelnen Data-Marts besteht die Gefahr inkonsistenter Datenbestände mit unterschiedlichen Definitionen für gleiche Kennzahlen bzw. Dimensionen ([ChDa97, 518], [KiRo02, 81], [LaSt08, 147]). Übergreifende Auswertungen über mehrere Data-Marts sind nicht ohne Weiteres (Abschnitt 5.5) möglich [WaAr05, 11]. Die Abstimmung und Koordination zwischen einzelnen Organisationseinheiten wird somit von der Architektur mit unabhängigen Data-Marts nicht unterstützt. Sie eignet sich jedoch für Unternehmen mit weitgehend autonomen Organisationseinheiten und zugehörigen Führungssystemen, da deren individuelle Anforderungen besser berücksichtigt werden können.

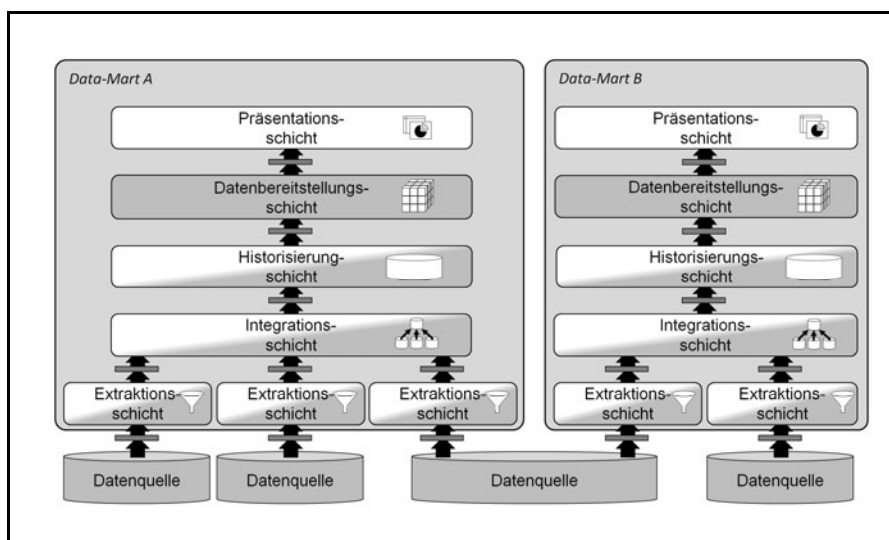


Abb. 7: Unabhängige Data-Marts

Der wesentliche Vorteil unabhängiger Data-Marts gegenüber einer zentralen Architektur besteht in der geringeren Entwicklungszeit bis zur Realisierung der Informationsversorgung der potentiellen Nutzer [ChDa97, 518]. Aufgrund der inhaltlichen Fokussierung entfällt der Aufwand für die Abstimmung und Konsolidierung des zu realisierenden Informationsangebots mit anderen Organisationseinheiten. Das Problem inkonsistenter Datenbestände zwischen einzelnen Data-Marts kann durch einheitliche Vorgaben für die Definition von Kennzahlen und Dimensionen beispielsweise mithilfe eines zentralen Metadaten-Repository reduziert werden (zur Überwindung der Heterogenität bei den Metadaten multipler DWH-Systeme siehe [Hart08]).

Die dezentrale Architektur führt grundsätzlich zu einem höheren Aufwand für den Betrieb und die Pflege des Systems. Beispielsweise müssen Schemaänderun-

gen in einer Datenquelle unter Umständen in mehreren Data-Marts berücksichtigt werden. Dabei kann nicht sichergestellt werden, dass die Anpassungen in allen Data-Marts zeitgleich den Nutzern zur Verfügung stehen. Gegenüber der zentralen Architektur sind zusätzlich die geringeren Anforderungen an die zugrunde liegende Hardware hervorzuheben.

5.3 Hub-and-Spoke-Architektur

Die **Hub-and-Spoke-Architektur**, auch „Corporate Information Factory“ genannt [Inmo05], besteht aus einem zentralen DWH mit abhängigen Data-Marts (Abb. 8). Die Informationsversorgung der einzelnen Organisationseinheiten erfolgt unter Nutzung des analyseoptimierten Datenbestands (Analyseschema) im entsprechenden Data-Mart. Dieser basiert auf dem relevanten Ausschnitt des konsolidierten Datenbestands im zentralen DWH (Basisdatenschema). Inkonsistenzen zwischen Berichten aus unterschiedlichen Data-Marts werden vermieden, da allen Data-Marts das einheitliche Begriffssystem des konsolidierten Datenbestands im zentralen DWH zugrunde liegt. Übergreifende Auswertungen über Daten aus mehreren Data-Marts können direkt auf Basis des zentralen DWH berechnet werden. Zusätzlich können individuelle fachliche und technische Anforderungen der jeweiligen Organisationseinheiten innerhalb des jeweiligen Data-Marts berücksichtigt werden [LaSt08, 149]. Beispielsweise können zusätzliche Datenquellen eingebunden werden, die nicht im zentralen DWH benötigt werden. Die Architekturform eignet sich speziell für Unternehmen, deren Organisationseinheiten und zugehörige Führungssysteme spezifische Anforderungen an die Managementunterstützung richten, ohne dabei eine zentrale Datenversorgung grundsätzlich in Frage zu stellen.

Entsprechend der zentralen Architektur besteht auch bei der Hub-and-Spoke-Architektur ein großer Aufwand für die unternehmensweite Abstimmung des einheitlichen Begriffssystems und für die Realisierung der konsolidierten Datenbasis.

Durch die zentrale Datenintegration und -aufbereitung können die Anzahl der Schnittstellen zu den Quellsystemen minimiert und ein redundanter Zugriff auf die Quellsysteme vermieden werden. Dies erleichtert den Betrieb und die Pflege des Gesamtsystems. Aufgrund der getrennten Datenhaltung in der Historisierungs- und der Datenbereitstellungsschicht kann die aufwändige Aktualisierung des zentralen DWH unabhängig von der Nutzung des Systems erfolgen.

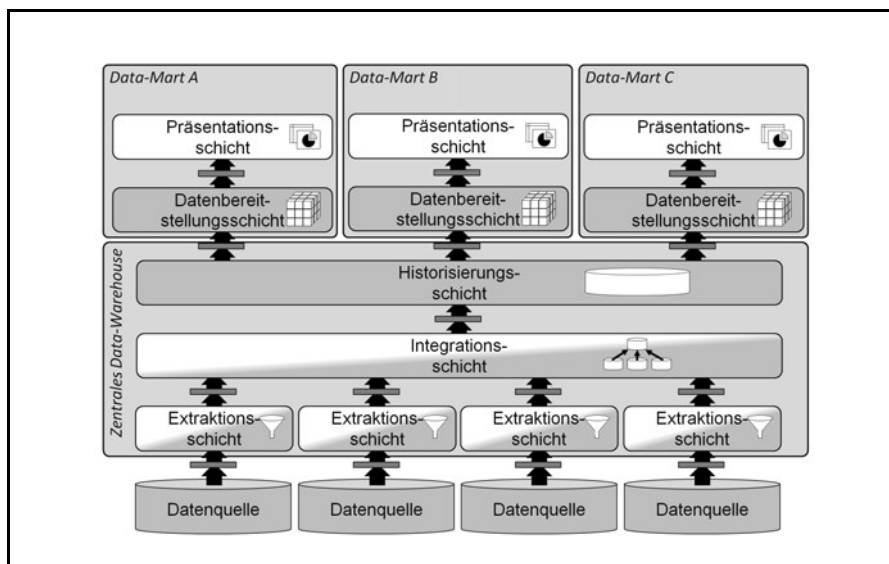


Abb. 8: Hub-and-Spoke-Architektur

5.4 Data-Mart-Bus-Architektur

Im Gegensatz zur Hub-and-Spoke-Architektur wird bei der **Data-Mart-Bus-** bzw. **DWH-Bus-Architektur** nach KIMBALL [KiRo02, 78 ff.] auf eine zentrale Datenhaltung verzichtet (Abb. 9). Die Konsistenz zwischen den Data-Marts wird stattdessen durch ein zentrales Metadaten-Repository und den Austausch gemeinsamer Dimensionsdaten und Kennzahlen („conformed dimensions“ bzw. „conformed facts“ [KiRo02, 82 ff.]) über einen logischen Bus sichergestellt. Dadurch wird die Abstimmung und Koordination zwischen den Unternehmensbereichen unterstützt und gleichzeitig können individuelle fachliche und technische Anforderungen in den einzelnen Data-Marts realisiert werden. Das Laden und die Pflege der gemeinsamen Dimensionen und Kennzahlen erfolgt im Verantwortungsbereich ausgewählter Data-Marts.

Der wesentliche Vorteil dieser Architekturform besteht in dem einheitlichen Begriffssystem, das ohne den Aufwand für die Entwicklung eines konsolidierten Gesamtdatenbestands „virtuell“ realisiert wird. Allerdings ist bei der Entwicklung der notwendige Aufwand zur Abstimmung der Definitionen für die gemeinsamen Dimensionen und Kennzahlen nicht zu unterschätzen. Die hier betrachtete Architekturform eignet sich für weitgehend autonome Organisationseinheiten, die zur besseren Koordination der Lenkungssysteme ihre DWH-Systeme partiell abstimmen.

Die dezentrale Architektur führt grundsätzlich zu einem höheren Aufwand für den Betrieb und die Pflege des Systems. Durch die Verteilung der gemeinsamen Dimensionsdaten und Kennzahlen über den logischen Bus kann jedoch die redun-

dante Pflege dieser Objekte und der redundante Zugriff auf die Quellsysteme vermieden werden. Das Fehlen der zentralen Datenhaltung führt zudem zu geringeren Anforderungen an die zugrunde liegende Hardware.

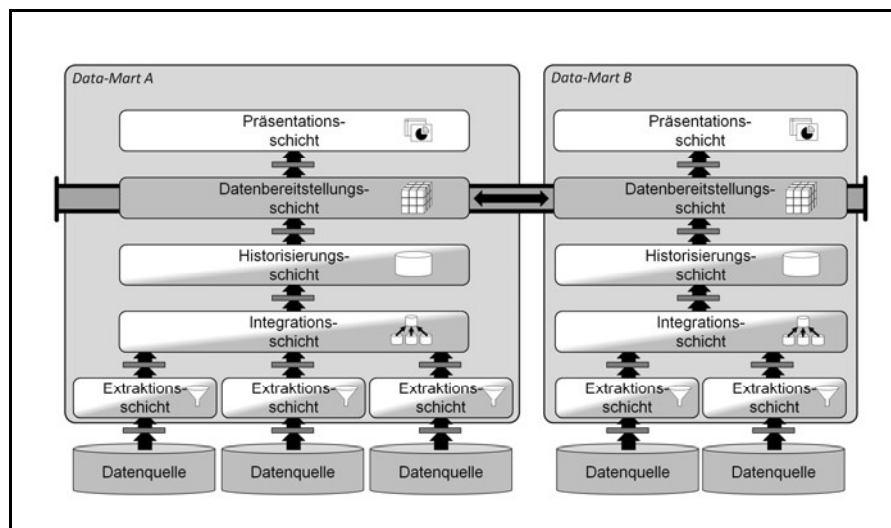


Abb. 9: Data-Mart-Bus-Architektur

5.5 Föderierte und hierarchische Data-Warehouse-Systeme

Föderierte und hierarchische DWH-Systeme resultieren aus der Weiterentwicklung einer bestehenden DWH-Architektur mit unabhängigen Data-Marts. Sie stellen im Gegensatz zur idealtypischen, zentralisierten Architektur für viele Unternehmen eine praktikable, organisatorisch durchsetzbare Lösung für eine unternehmensweite Informationsversorgung dar [LaSt08, 150 ff.]. Sie ermöglichen die übergreifende Auswertung von heterogenen Daten aus unabhängigen Data-Marts, ohne die bestehende Infrastruktur aufzulösen [WaAr05, 13]. Dadurch kann eine konsistente und einheitliche Sichtweise auf die wichtigsten Daten für das Unternehmen sichergestellt werden [LaSt08, 150]. Diese Architekturform unterstützt beispielsweise das Management von Unternehmen mit (teil-)autonomen Organisationseinheiten, indem die Autonomie der einzelnen DWH-Teilsysteme gewahrt bleibt (z. B. im Hochschulwesen [SBPU01]).

Bei **föderierten DWH-Systemen** (Abb. 10) wird die konsolidierte Sichtweise lediglich virtuell erzeugt. Die Berechnung der Anfragen erfolgt entsprechend einem föderierten Datenbanksystem [Conr97] auf Basis der Datenhaltung in den zugrunde liegenden Data-Marts. Im Gegensatz dazu werden beim **hierarchischen DWH-System** (Abb. 11) die Datenbestände aus den unterschiedlichen Data-Marts in einer separaten Datenbasis zusammengeführt. Dadurch können zusätzlich weitere Datenquellen eingebunden werden.

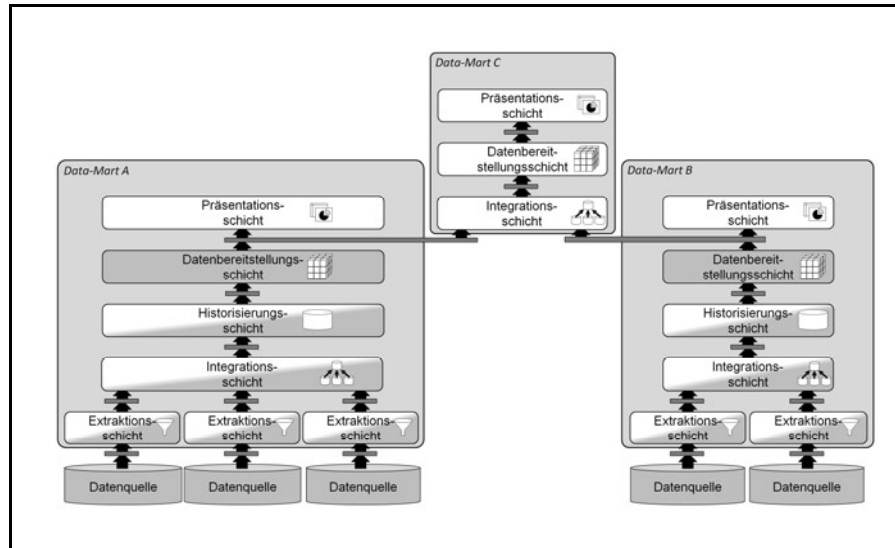


Abb. 10: Föderierte Data-Warehouse-Systemarchitektur

Das größte Problem bei der Entwicklung von föderierten oder hierarchischen DWH-Systemen besteht in der virtuellen oder physischen Konsolidierung der heterogenen Datenbestände aus den unterschiedlichen Data-Marts. Zusätzlich können bei föderierten Systemen Probleme mit der Konnektivität und dem Datendurchsatz auftreten [LaSt08, 150 ff.].

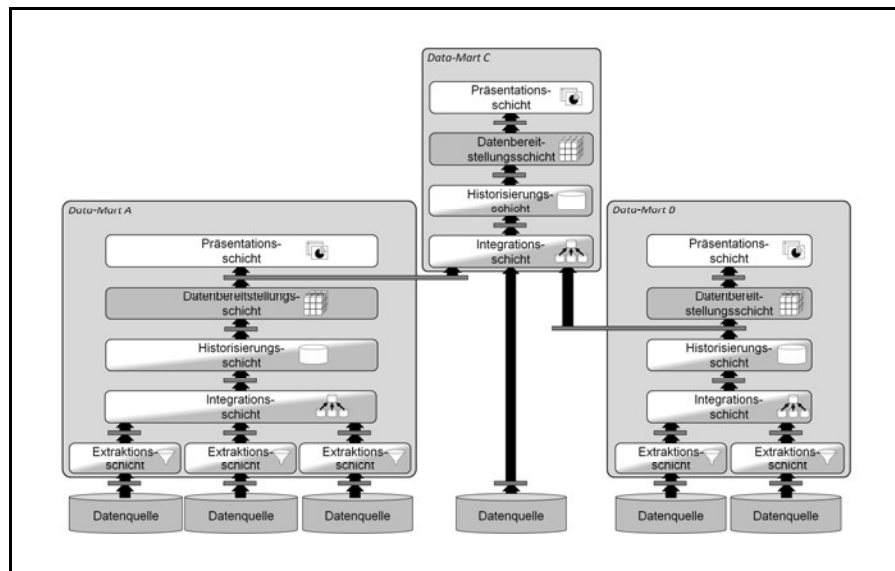


Abb. 11: Hierarchisches Data-Warehouse-System

6 Zusammenfassung

Auf der Basis eines generischen Rahmens in Form einer Hierarchie von Software-schichten wurden im vorliegenden Beitrag typische Architekturkonzepte von DWH-Systemen beschrieben. Die einzelnen Architekturformen unterstützen jeweils spezifische Anforderungen der Organisations- und Führungsstruktur des zugrunde liegenden Unternehmens sowie spezielle Anforderungen an die Datenbereitstellung.

Alle genannten Architekturformen haben in der Praxis Verbreitung gefunden. Gemäß einer empirischen Studie aus dem Jahr 2006 [ArWe06] werden die einzelnen Architekturformen in den befragten Unternehmen wie folgt eingesetzt: Hub-and-Spoke-Architektur (39%), Data-Mart-Bus-Architektur (26%), zentrales DWH-System (17%), unabhängige Data-Marts (12%), föderierte bzw. hierarchische DWH-Architektur (4%).

Literatur

- [ArWe06] Ariyachandra, T.; Watson, H. J.: Which Data Warehouse Architecture Is Most Successful? In: Business Intelligence Journal 11 (1), 2006, S. 4-6.
- [BaGü08] Bauer, A.; Günzel, H. (2008): Data-Warehouse-Systeme. Architektur, Entwicklung, Anwendung. 3. Aufl., dpunkt, Heidelberg, 2008.
- [ChDa97] Chaudhuri, S.; Dayal, U.: An Overview of Data Warehousing and OLAP Technology. In: ACM SIGMOD Record 26 (1), 1997, S. 65-74.
- [ChGH05] Chamoni, P.; Gluchowski, P.; Hahne, M.: Business Information Warehouse. Perspektiven betrieblicher Informationsversorgung und Entscheidungsunterstützung auf der Basis von SAP-Systemen. Springer, Berlin, 2005.
- [ChGI00] Chamoni, P.; Gluchowski, P.: On-Line Analytical Processing (OLAP). In: Mucksch, H.; Behme, W. (Hrsg.): Das Data Warehouse-Konzept - Architektur, Datenmodelle, Anwendungen. 4. Aufl., Gabler, Wiesbaden, 2000, S. 333-376.
- [Conr97] Conrad, S.: Föderierte Datenbanksysteme. Konzepte der Datenintegration. Springer, Berlin, 1997.
- [Fisc01] Fischer, C.: Externe Daten als Achillesferse von Data-Warehouse-Projekten - Probleme und Lösungsansätze. In: Schütte, R.; Rott Howe, T.; Holten, R. (Hrsg.): Data Warehouse Managementhandbuch - Konzepte, Software, Erfahrungen. Springer, Berlin, 2001, S. 107-118.

- [FeSi08] Ferstl, O. K.; Sinz, E. J.: Grundlagen der Wirtschaftsinformatik. 6. Aufl., Oldenbourg, München, 2008.
- [GlCh06] Gluchowski, P.; Chamoni, P.: Entwicklungslinien und Architekturkonzepte des On-Line Analytical Processing. In: Chamoni, P.; Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Business Intelligence-Technologien und -Anwendungen. 3. Aufl., Springer, Berlin, 2006, S. 143-176. (VERWEIS AUF AKTUELLE AUFLAGE?)
- [Golf09] Golfarelli, M.: Data warehouse life-cycle and design. In: Özsu, M. T.; Liu, L. (Hrsg.): Encyclopedia of Database Systems. Springer, Berlin, 2009.
- [Hahn06] Hahne, M.: Mehrdimensionale Datenmodellierung für analyseorientierte Informationssysteme. In: Chamoni, P.; Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Business Intelligence-Technologien und -Anwendungen. 3. Aufl., Springer, Berlin, 2006, S. 177-206. (VERWEIS AUF AKTUELLE AUFLAGE?)
- [Hart08] Hartmann, S.: Überwindung semantischer Heterogenität bei multiplen Data-Warehouse-Systemen. University of Bamberg Press, Bamberg, 2008.
- [Inmo96] Inmon, W. H.: Building the data warehouse. 2. Aufl., Wiley, New York, 1996.
- [Inmo99] Inmon, W. H.: Building the operational data store. 2. Aufl., Wiley, New York, 1999.
- [Inmo00] Inmon, W. H.: ODS Types. In: Online-Magazine DMReview, Januar 2000, <http://www.information-management.com/issues/20000101/1749-1.html> (Abruf am 2009-06-03).
- [Inmo05] Inmon, W. H.: Building the data warehouse. 4. Aufl., Wiley, New York, 2005.
- [KeFi06] Kemper, H.-G.; Finger, R.: Transformation operativer Daten - Konzeptionelle Überlegungen zur Filterung, Harmonisierung, Aggregation und Anreicherung im Data Warehouse. In: Chamoni, P.; Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Business Intelligence-Technologien und -Anwendungen. 3. Aufl., Springer, Berlin, 2006, S. 113-128. (VERWEIS AUF AKTUELLE AUFLAGE?)
- [KiRo02] Kimball, R.; Ross, M.: The Data Warehouse Toolkit. The complete Guide to Dimensional Modeling. 2. Aufl., Wiley, New York, 2002.
- [Küpp97] Küpper, H.-U.: Controlling, Konzeption, Aufgaben und Instrumente. 2. Aufl., Schäffer-Poeschel, Stuttgart, 1997.
- [LaSt08] Lahrmann, G.; Stroh, F.: Systemarchitekturen für die Informationslogistik. In: Dinter, B.; Winter, R. (Hrsg.): Integrierte Informationslogistik. Springer, Berlin, 2008, S. 137-164.

- [Lehn03] Lehner, W.: Datenbanktechnologie für Data-Warehouse-Systeme. Konzepte und Methoden. dpunkt, Heidelberg, 2003.
- [LeNa07] Leser, U.; Naumann, F.: Informationsintegration. Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen. dpunkt, Heidelberg, 2007.
- [Mehr05] Mehrwald, Christian: Data Warehousing mit SAP BW 3.5. Architektur, Implementierung, Optimierung. 3. Aufl., dpunkt, Heidelberg, 2005.
- [Micr09] o. V.: SQL Server 2008-Online-Dokumentation: Multidimensional Expressions (MDX) – Referenz. Microsoft-Cooperation, 2009, <http://technet.microsoft.com/de-de/library/ms145506.aspx> (Abruf am 2009-06-03).
- [MuBe00] Mucksch, H.; Behme, W.: Das Data Warehouse-Konzept als Basis einer unternehmensweiten Informationslogistik, in: Mucksch, H.; Behme, W. (Hrsg.): Das Data Warehouse-Konzept. 4. Aufl., Gabler, Wiesbaden, 2000, S. 3–80.
- [Muck06] Mucksch, H.: Das Data Warehouse als Datenbasis analytischer Informationssysteme. Architektur und Komponenten. In: Chamoni, P.; Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Business Intelligence-Technologien und -Anwendungen. 3. Aufl., Springer, Berlin, 2006, S. 129–142.
- [SBPU01] Sinz, E. J., Böhnlein, M., Plaha, M., Ulbrich-vom Ende, A.: Architekturkonzept eines verteilten Data-Warehouse-Systems für das Hochschulwesen. In: Buhl, H.-U., Huther, A., Reitwiesner, B. (Hrsg.): Information Age Economy (WI-IF 2001, Augsburg, 19.-21. September), Physica, Heidelberg, 2001.
- [Sche06] Schelp, J.: „Real“-Time Warehousing und EAI. In: Chamoni, P.; Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Business Intelligence-Technologien und -Anwendungen. 3. Aufl., Springer, Berlin, 2006, S. 425–438. (VERWEIS AUF AKTUELLE AUFLAGE?)
- [SeSi05] Sen, A.; Sinha, A. P.: A comparison of data warehousing methodologies. In: Communications of the ACM 48 (3), 2005, S. 79–84.
- [Sinz01] Sinz E.J.: Data Warehouse. In: Küpper, H.-U.; Wagenhofer, A. (Hrsg.): Handwörterbuch Unternehmensrechnung und Controlling. 4. Aufl., Schäffer-Poeschel, Stuttgart, 2001, S. 309–318.
- [WaAr05] Watson, H. J.; Ariyachandra, T.: Data Warehouse Architectures: Factors in the Selection Decision and the Success of the Architectures. Technical Report, 2005, http://www.terry.uga.edu/~hwatson/DW_Architecture_Report.pdf (Abruf am 2009-04-20).

- [Zeh03] Zeh, T.: Data Warehousing als Organisationskonzept des Datenmanagements: Eine kritische Betrachtung der Data-Warehouse-Definition von Inmon. In: Informatik – Forschung und Entwicklung 18 (1), 2003, S. 32–38.

Anmerkungen

- ¹ Die Definition von INMON im Sinne einer Aufzählung der konstituierenden Merkmale eines DWH wird mittlerweile in der Literatur kontrovers diskutiert (siehe u. a. [Zeh03]). Hintergrund sind Erweiterungen der Einsatzbereiche für DWH. Da in diesem Beitrag das DWH in seiner Funktion als Hilfsregelstrecke zur Unterstützung des Managements betrachtet wird, soll diese Diskussion hier nicht weiter verfolgt werden.
- ² Der Begriff Operational-Data-Store wird in der Literatur und der Praxis sehr unterschiedlich verwendet. INMON beschreibt in [Inmo00] ursprünglich fünf Klassen von ODS (siehe hierzu auch [BaGü08, 56]). Die SAP versteht unter einem ODS beispielsweise die relationale Speicherung der integrierten und historisierten Daten ohne multidimensionalen Bezug im Gegensatz zum „InfoCube“ [ChGH05, 49]. Daher wird in diesem Beitrag nur kurz auf die wesentlichen Eigenschaften eines ODS eingegangen.