

Das Common Warehouse Metamodel als Referenzmodell
für Metadaten im Data Warehouse und dessen
Erweiterung im SAP® Business Information Warehouse

Dr. Michael Hahne

3. März 2005

Agenda

- Metadaten im Data Warehouse
- Das Common Warehouse Metamodel (CWM)
- SAP Business Information Warehouse (BW)
- Implementierung des CWM im BW
- Zusammenfassung

Agenda

- **Metadaten im Data Warehouse**
- Das Common Warehouse Metamodel (CWM)
- SAP Business Information Warehouse (BW)
- Implementierung des CWM im BW
- Zusammenfassung

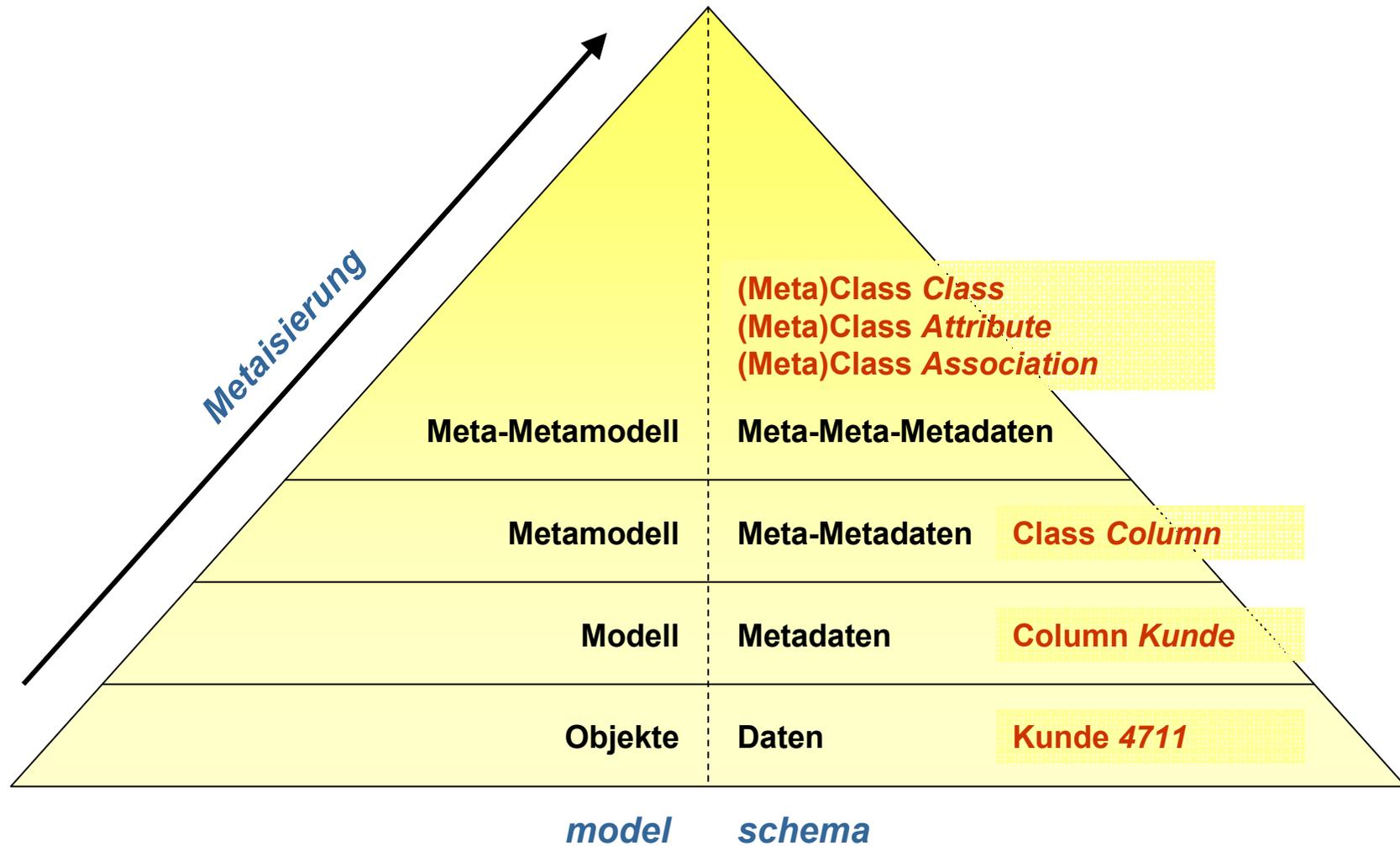
Begriff der Metadaten

- Prägnant:
Daten über Daten
aber nicht hinreichend genau
- Präziser:
Data that describes the meaning and structure of business data, as well as how it is created, accessed and used
[Devlin, 1997]
- Geläufig:
Jede Art von Information, die für den Entwurf, die Konstruktion und die Benutzung eines Informationssystems benötigt werden
[Bauer/Günzel, 2001]

Modellbegriff

- konstruktivistischer Ansatz der Modellierung:
Modelle sind das Ergebnis eines Modellierungsvorganges, bei dem die reale Diskurswelt, die so genannte Miniwelt, durch Abstraktion abgebildet wird
- Klassifikation:
Beschreibungs-, Erklärungs- und Entscheidungsmodelle
- Modell:
einerseits Modellsprache (engl. **model**)
Andererseits Ausprägung (engl. **schema**)

Abstraktionsebenen der Modellierung



Bedeutung von Metadaten im Data Warehousing

- Für Benutzer: Informationen im Data Warehouse weitgehend autonom nutzen
 - Definitionen der verwendeten Geschäftsbegriffe
 - Verbindung des Fachvokabulars mit den Datenobjekten im Data Warehouse
 - Berichtsbeschreibungen, Ansprechpartner, Zugangsberechtigungen
- Für Entwicklung und Administration: technisch orientierte Metadaten
 - Angaben über Datenquellen, Regeln zur Verbesserung der Datenqualität
 - Regeln für Transformations- und Konsolidierungsschritte
 - Zuordnung zwischen Datenquellen und den Data Warehouse-Modellen
- „best-of-breed“ vs. „end-to-end“
 - heterogene komplexe Systemlandschaft → einheitliche Data Warehouse Lösung
 - Verbesserte Interoperabilität durch Metadatenaustausch
- Metadatenmanagement ist eine zentrale Funktion zur Steigerung von Effektivität und Effizienz einer Data Warehouse-Lösung

Klassifikation von Metadaten

- DSS vs. operativ:
 - **operationalen Metadaten:** Quellsysteme, Originaldatenquellen, Datenstrukturen, Feldbezeichnungen sowie Transformationsprozess und Zieldatenquellen
 - **DSS-Metadaten:** Speicherorte und –strukturen im Data Warehouse, Objekte des Business Modells, Navigationsmöglichkeiten
- **Hauptfunktionen** eines Data Warehouse:
 - **Meta-Data:** Metadatenverwaltung, Fundament für die anderen Funktionen
 - **Source:** Identifikation von Datenquellen
 - **Load:** Datentransfer- und Transformationsprozess
 - **Storage:** Speicherform, Datenverteilung und Archivierung
 - **Query:** Abfrage- und Auswertefunktion, OLAP, Simulationsmethoden und Data Mining-Techniken
- **Dreidimensionale Klassifikation:**
 - **Prozess:** Design, Populate, Administer, Analyze
 - **Benutzer:** Technical User oder Business User
 - **Datensicht:** Operational, Data Warehouse, Data Mart

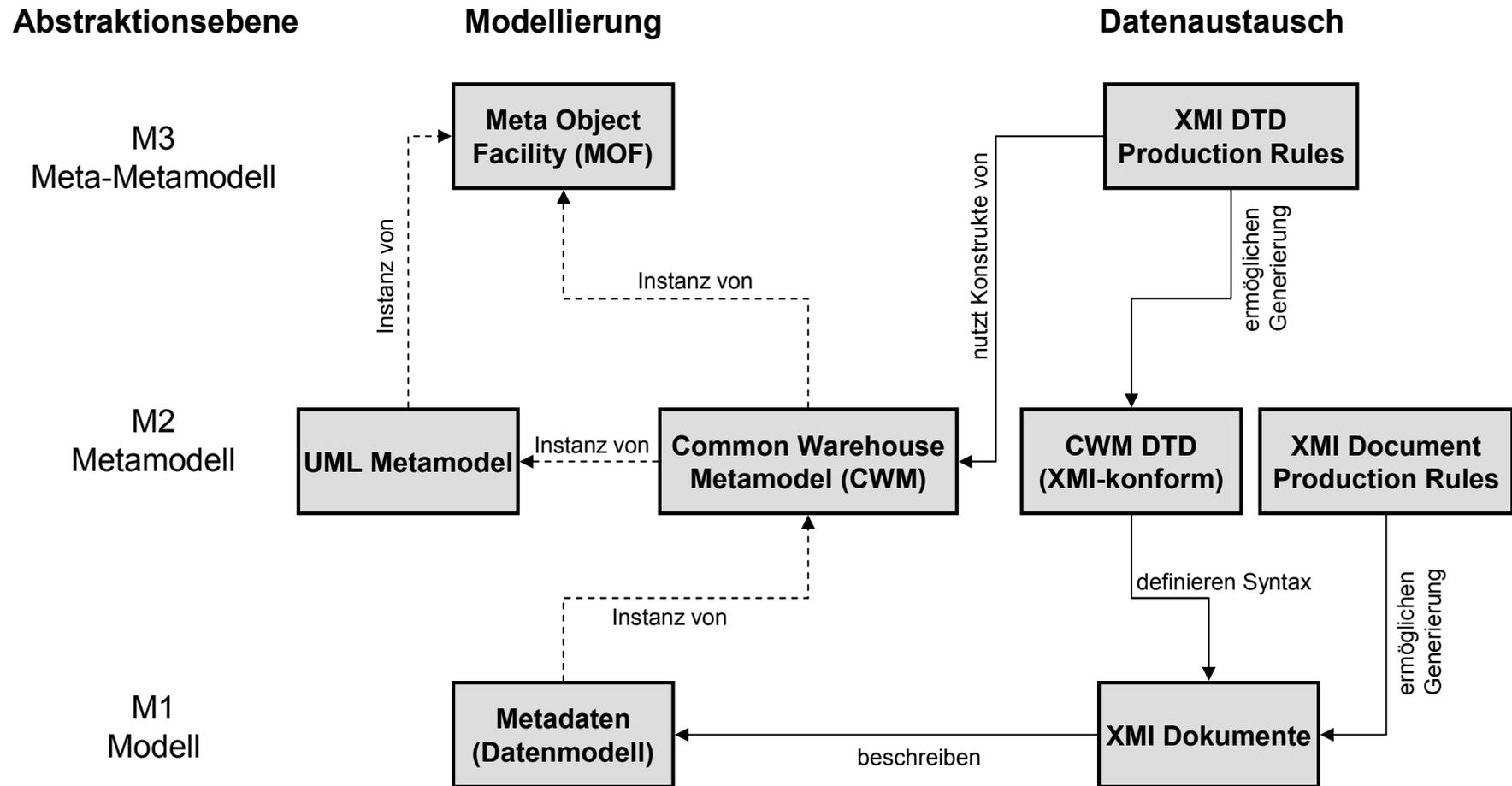
Agenda

- Metadaten im Data Warehouse
- **Das Common Warehouse Metamodel (CWM)**
- SAP Business Information Warehouse (BW)
- Implementierung des CWM im BW
- Zusammenfassung

Common Warehouse Metamodel (CWM)

- Modell für Metadaten im Data Warehousing
- seit 1998 unter aktiver Beteiligung u.a. von IBM, NCR, Hyperion Solutions und Oracle entwickelt
- von der Object Management Group (OMG) erstmals im Jahr 2000 als Standard verabschiedet
- Eingebettet in Model Driven Architecture (MDA) der OMG
- basiert auf über 200 Klassen und über 150 Assoziationen in einem UML-Klassenmodell
- Metadatenaustausch datenstrombasiert über XMI (XML Metadata Interchange) und funktionsbasiert über CORBA-Schnittstellen
- Leitlinien: Vollständigkeit, Verständlichkeit, Unabhängigkeit, Kompatibilität, Interoperabilität, Erweiterbarkeit und Verfügbarkeit

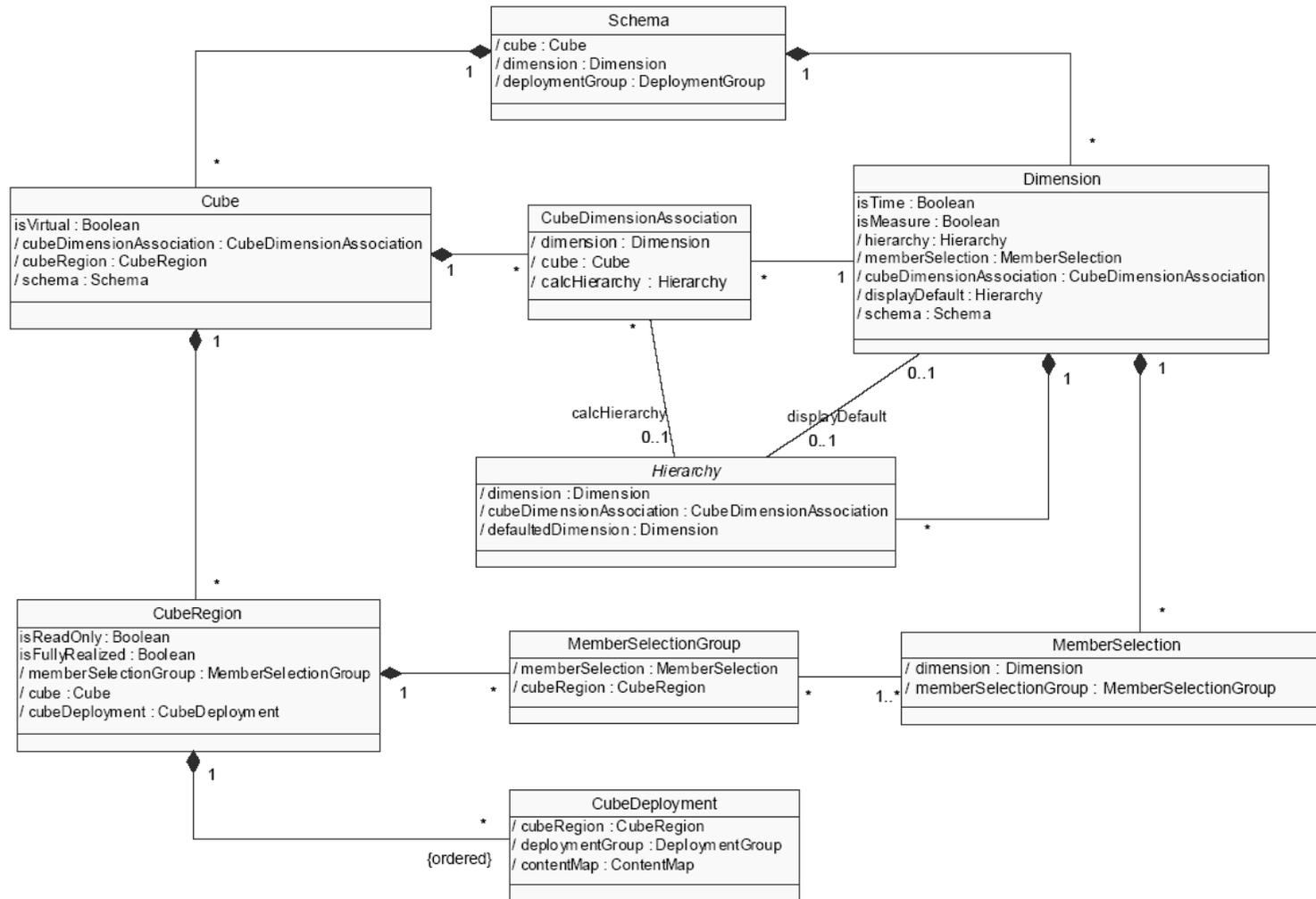
Metamodellarchitektur der OMG



CWM Paketstruktur

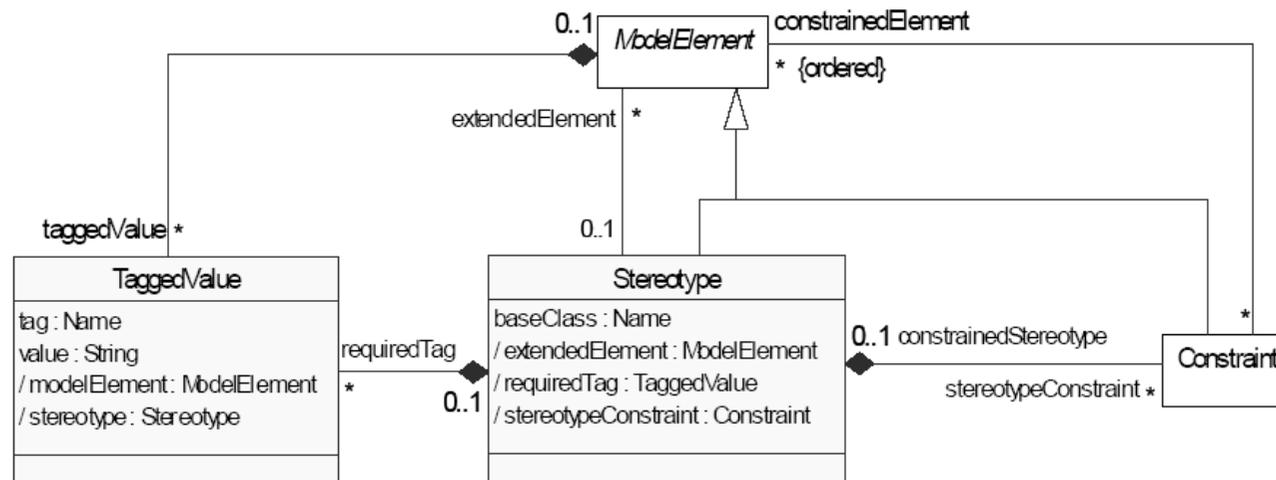
Management	Warehouse Process			Warehouse Operation		
Analysis	Transformation		OLAP	Data Mining	Information Visualization	Business Nomenclature
Resource	Object Model	Relational	Record	Multidimensional		XML
Foundation	Business Information	Data Types	Expression	Keys and Indexes	Type Mapping	Software Deployment
Object Model	Core		Behavioral	Relationships		Instance

OLAP Metamodel



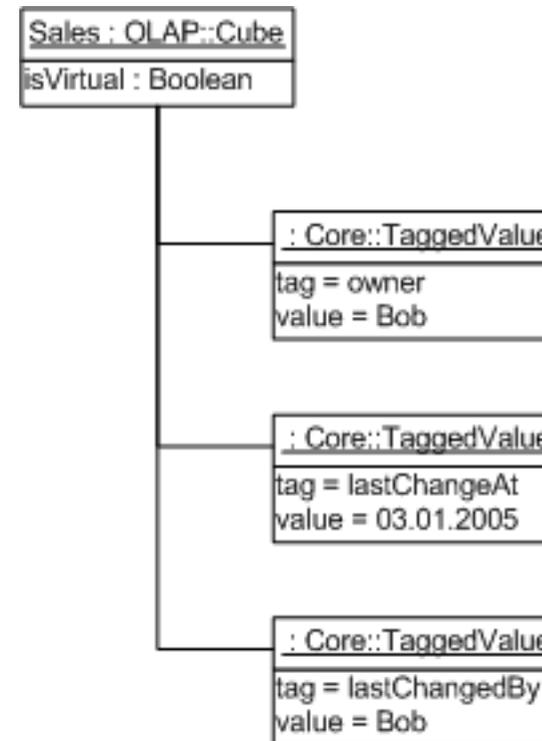
Erweiterungsmöglichkeiten des CWM

- Tagged Values auf Basis des Paketes *Core*
- Stereotypes ebenfalls auf Basis des Paketes *Core*
- Erweiterungsklassen als flexibelste Erweiterungsmöglichkeit



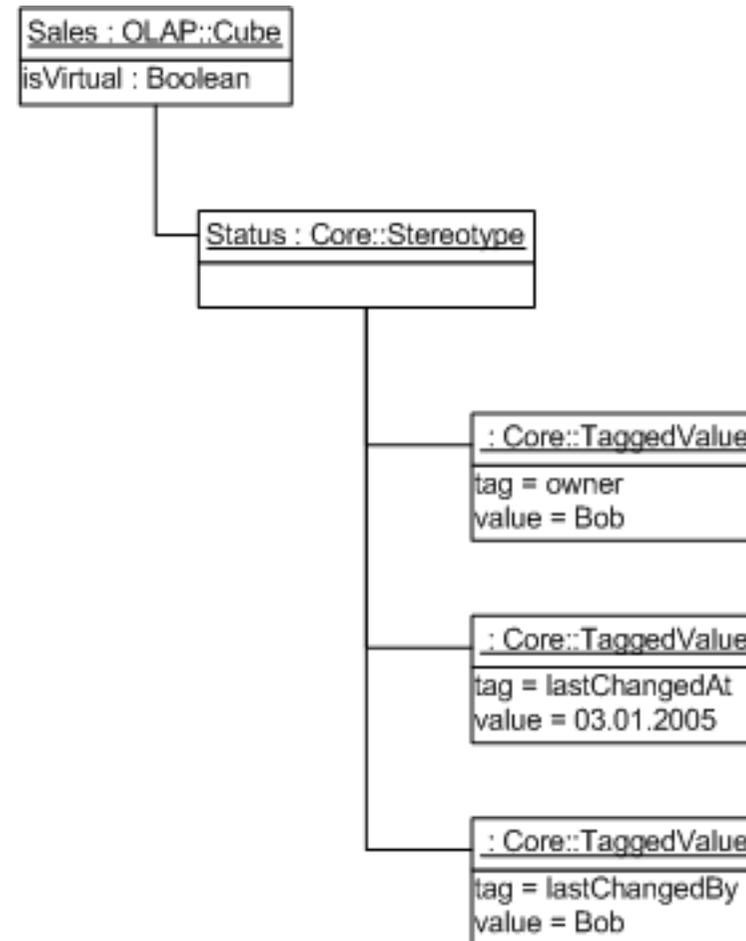
Erweiterungsmöglichkeiten des CWM: Tagged Values

- Tagged Values: Aus der UML-Spezifikation ins CWM übernommen (Paket *Core*)
- Erweiterung um weitere Attributwerte der Form Attributname (tag) = Attributwert (value)
- Im Allgemeinen unabhängig von Versionswechseln der CWM-Spezifikation
- Auf Werte von Typ String beschränkt
- Kein direkter Zugriff (Der Benutzer muss von den definierten Tagged Values Kenntnis haben)
- Keine zusätzlichen Assoziationen im Metamodell



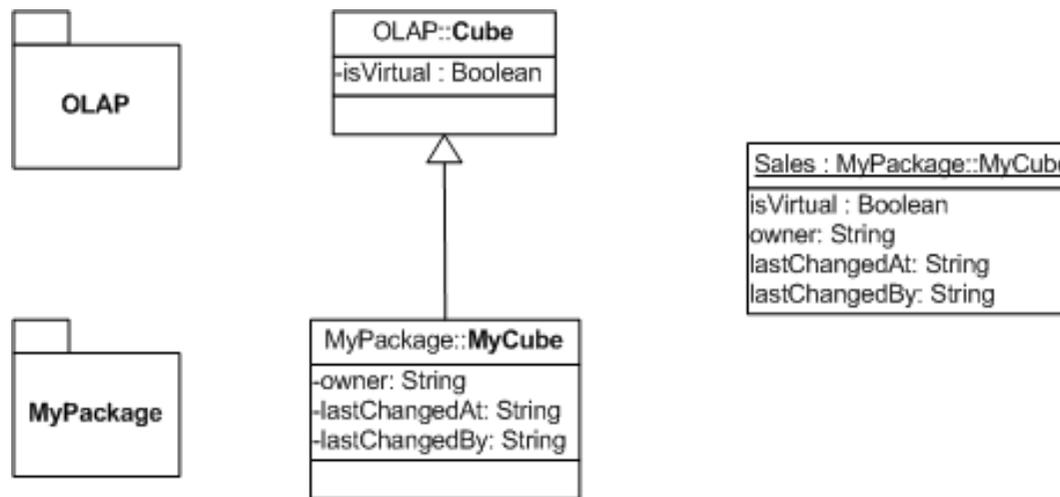
Erweiterungsmöglichkeiten des CWM: Stereotypen

- Stereotype: Aus der UML-Spezifikation ins CWM übernommen (Paket *Core*)
- Im Allgemeinen unabhängig von Versionswechseln der CWM-Spezifikation
- Keine zusätzlichen Assoziationen im Metamodell
- Über Stereotypen können u. a. mehrere Tagged Values inhaltlich gruppiert werden



Erweiterungsmöglichkeiten des CWM: Erweiterungsklassen

- flexibelste Möglichkeit der Erweiterung: Eigene Klassen, Assoziationen und Pakete
- Hoher Freiheitsgrad in der Gestaltung
- Gefahr der Entfernung vom Standard
- Vererbungskonzept sollte genutzt werden
- Abhängigkeiten bei Release-Wechsel des CWM
- Extension Packages der OMG z. B. zu Essbase oder Express basieren auf diesem Konzept



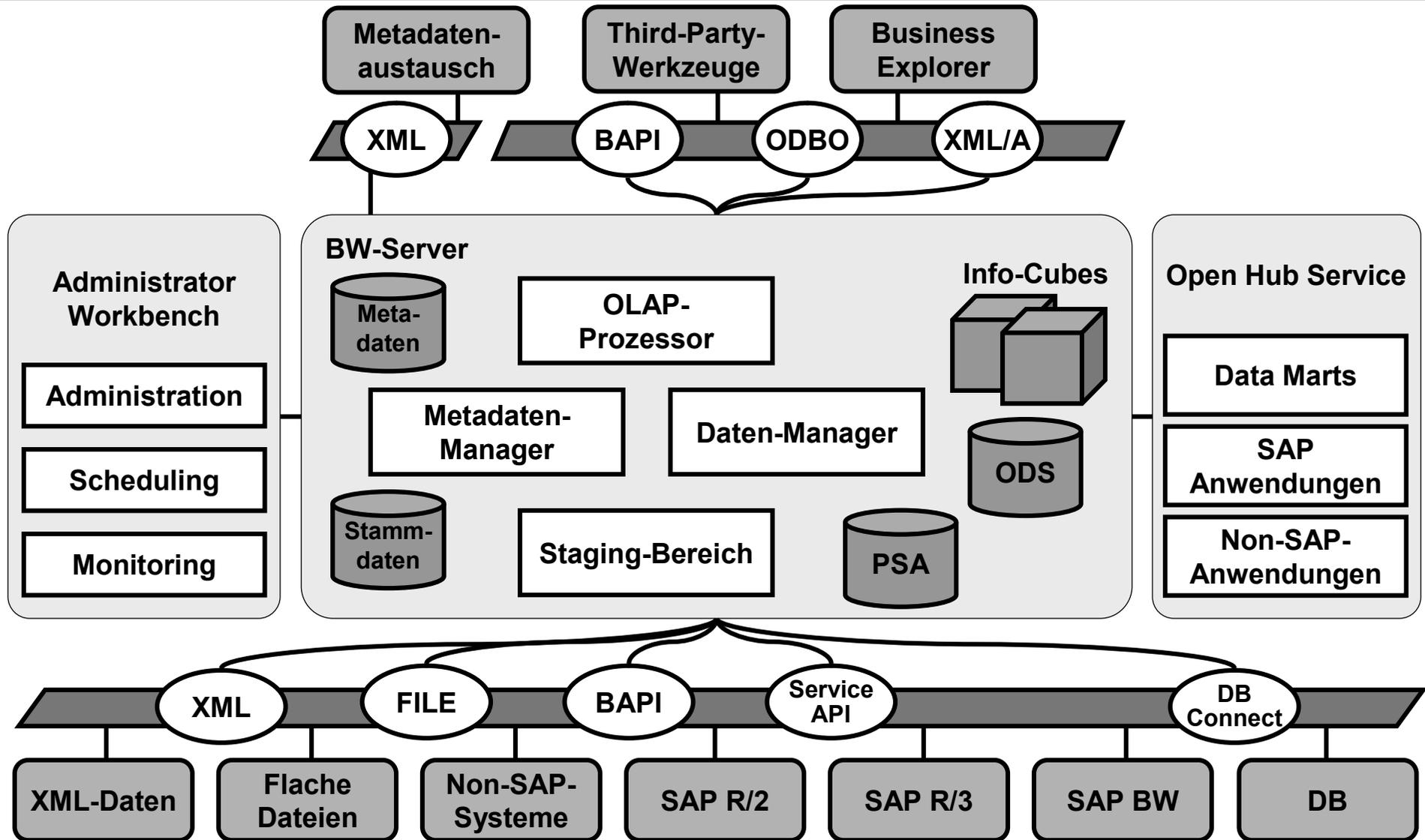
Erweiterungsbedarf des CWM

- CWM sehr stark auf technische Metadaten fokussiert
- Keine Möglichkeit der Abbildung von Berechtigungen
 - Heterogene DW-Landschaften weisen oftmals viele verschiedene Konzepte zur Verwaltung von Benutzern, Rollen und Berechtigungen auf
 - Möglich durch Standardisierung von berechtigungsbezogenen Metadaten:
 - Abgleich von Berechtigungen
 - Erhöhung der Transparenz zur Vermeidung von Sicherheitslücken
- Fehlende Konstrukte für Qualitätsmanagement
 - konzeptionell, funktional, technisch
 - Akzeptanzbereiche und Alerting-Mechanismen
- Unzureichende Berücksichtigung vielfältiger Mechanismen der Berichtsgruppierung und Verteilung über unterschiedlichste Kanäle sowie Konstrukte der Berichtsdefinitionen
- Für die Berücksichtigung der organisatorischen Data Warehouse-Prozesse bietet sich eine Erweiterung des CWM an

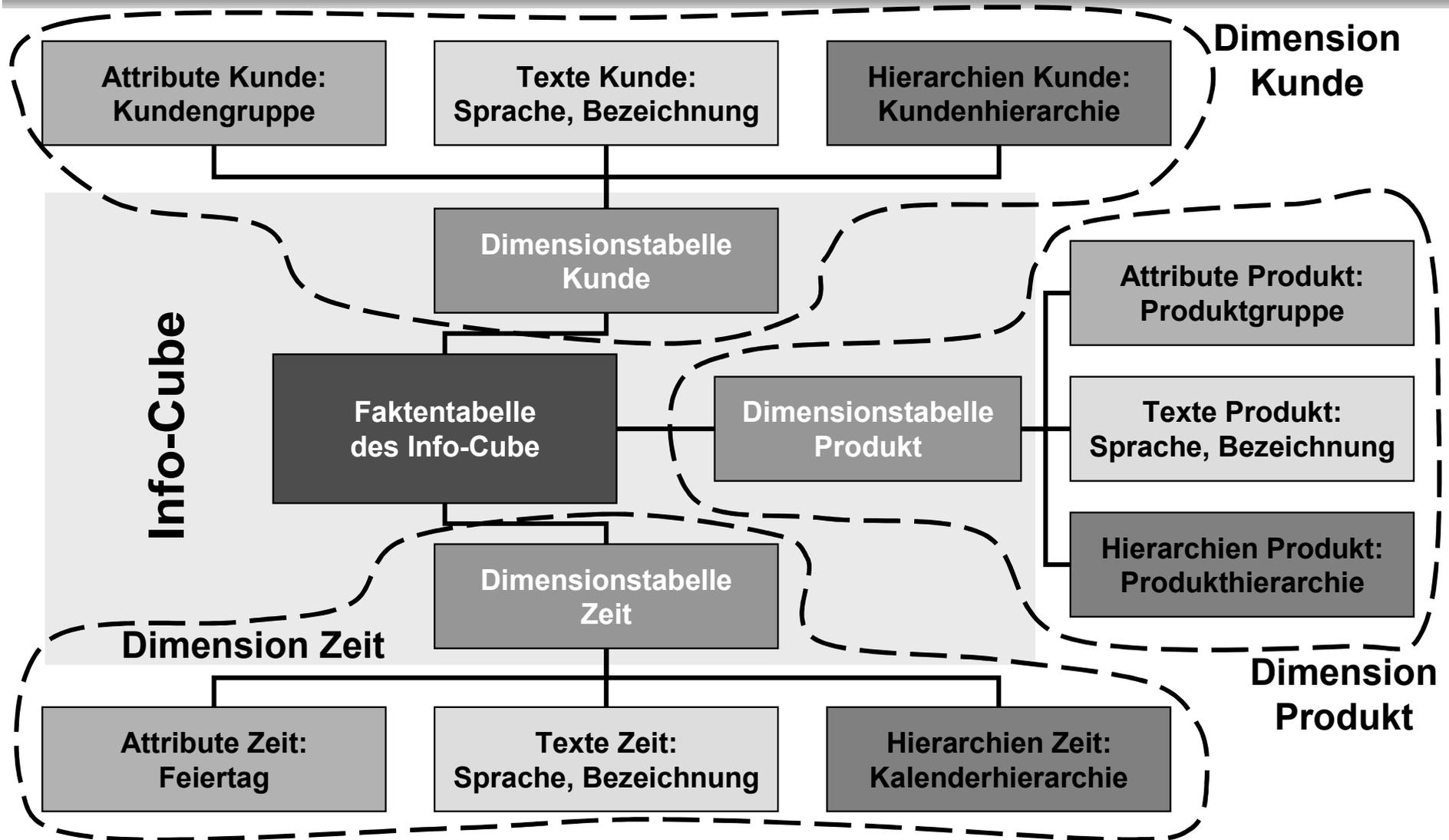
Agenda

- Metadaten im Data Warehouse
- Das Common Warehouse Metamodel (CWM)
- **SAP Business Information Warehouse (BW)**
- Implementierung des CWM im BW
- Zusammenfassung

Architektur des SAP BW



Datenmodell des SAP BW



Agenda

- Metadaten im Data Warehouse
- Das Common Warehouse Metamodel (CWM)
- SAP Business Information Warehouse (BW)
- **Implementierung des CWM im BW**
- Zusammenfassung

BW CWM Paketstruktur: *types*

Management	whm		
Analysis	transformation	bex	authorization
Resource	core	relational	olap
Foundation	types		foundation

- BW-spezifische Datentypen:
 - Boolean
 - Integer
 - Name
 - String
 - Time

BW CWM Paketstruktur: *foundation*

Management	whm		
Analysis	transformation	bex	authorization
Resource	core	relational	olap
Foundation	types		foundation

- Assoziationen z.B. für die Verbindung der Objekte entlang des Datenflusses im BW
- Attribute für Objekt-Eigenschaften wie etwa der Eigner, der Status und beschreibende Texte
- Business Content Informationen

BW CWM Paketstruktur: *core*

Management	whm		
Analysis	transformation	bex	authorization
Resource	core	relational	olap
Foundation	types		foundation

- Klassendefinitionen für
 - Info-Objekte
 - Währungsumrechnungsarten,
 - Personen
 - ABAP-Routinen
 - diverse (Quell-)Systeme

BW CWM Paketstruktur: *relational*

Management	whm		
Analysis	transformation	bex	authorization
Resource	core	relational	olap
Foundation	types		foundation

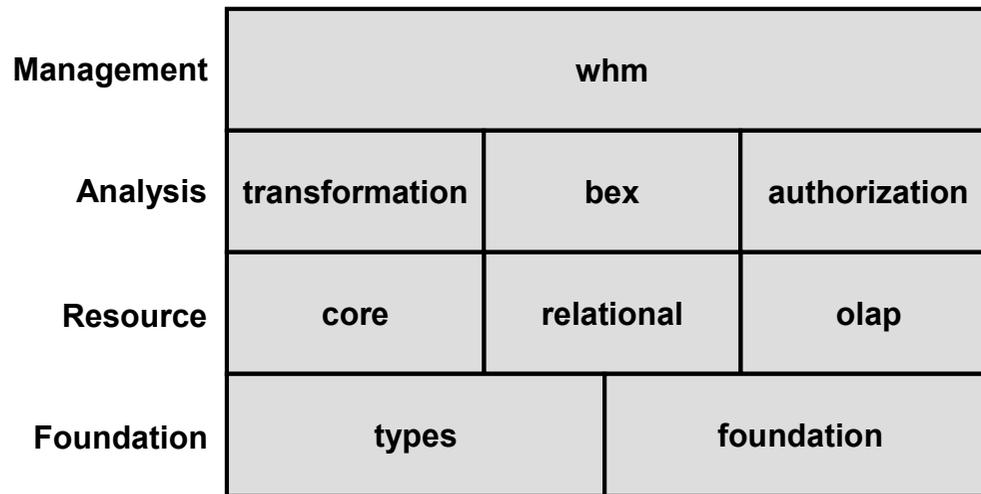
- ODS-Objekte
- Info-Sets
- Indizes

BW CWM Paketstruktur: *olap*

Management	whm		
Analysis	transformation	bex	authorization
Resource	core	relational	olap
Foundation	types		foundation

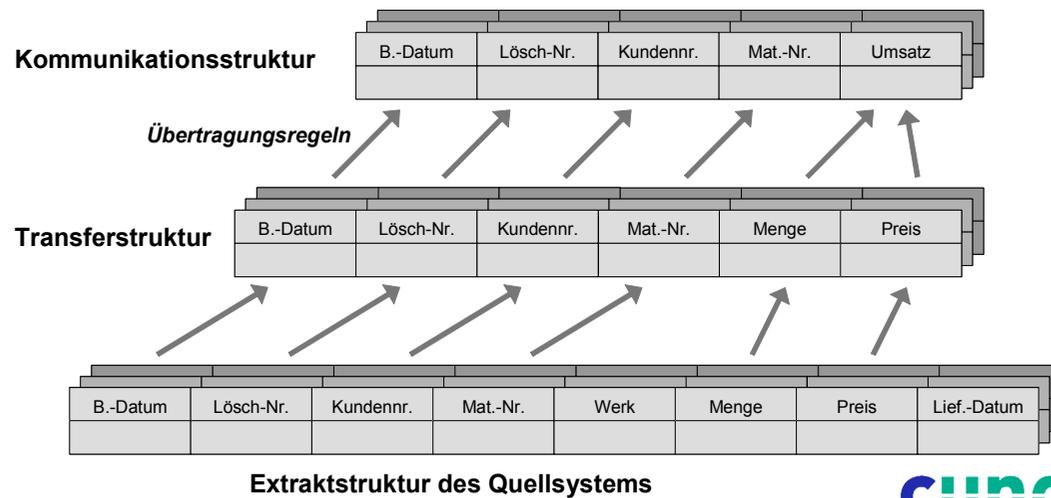
- Klassen für die Info-Provider und Aggregate
 - Klassen *InfoCube*, *InfoProvider*, *MultiProvider*, *VirtualCube*, *Aggregate*
- Reporting Agent
 - Klassen *ReportAgentPackage*, *ReportAgentSetting*

BW CWM Paketstruktur: *transformation*



- Transformationsmöglichkeiten des BW entlang des systeminternen Datenflusses

- Transferstruktur
- Kommunikationsstruktur
- Info-Source
- Update-Rule
- etc.



BW CWM Paketstruktur: *bex*

Management	whm		
Analysis	transformation	bex	authorization
Resource	core	relational	olap
Foundation	types		foundation

- Klassen zur Datenauswertung auf Basis des BEx Analyzer
 - Klassen *ExcelWorkbook*, *InfoSetQuery*, *Query*, *QueryCalculatedKeyFigure*, *QueryElement*, *QueryFormula*, *QuerySelection*, *QueryStructure* und *QueryVariable*

BW CWM Paketstruktur: *authorization*

Management	whm		
Analysis	transformation	bex	authorization
Resource	core	relational	olap
Foundation	types		foundation

- Berechtigungsinformationen entsprechend des SAP Rollenkonzeptes
 - Klassen *Authorization*, *Role* und *UserGroup*

BW CWM Paketstruktur: *whm*

Management	whm		
Analysis	transformation	bex	authorization
Resource	core	relational	olap
Foundation	types		foundation

- Für den Betrieb des Business Information Warehouse notwendige Objekte
 - Klassen *EventChain*, *InfoPackage* und *InfoPackageGroup* zur Darstellung von Prozessketten, Info-Packages und deren Gruppen
- Klassen für die Objektstrukturierung mit Hilfe von Katalogen, Bereichen und Anwendungskomponenten
 - Klassen *InfoObjectCatalog*, *InfoArea* und *Application*

Agenda

- Metadaten im Data Warehouse
- Das Common Warehouse Metamodel (CWM)
- SAP Business Information Warehouse (BW)
- Implementierung des CWM im BW
- **Zusammenfassung**

Zusammenfassung

- Das Common Warehouse Metamodel (CWM) ist ein anerkannter Industriestandard für Modellierung und Austausch von Metadaten
- Das CWM bietet einen viel versprechenden Ansatz zur Integration von Komponenten in heterogenen Data Warehouse-Landschaften
- Das CWM ist architektur- und plattformunabhängig und berücksichtigt daher keine spezifischen Objekte dedizierter Werkzeuge, bietet aber flexible Möglichkeiten der Erweiterung
- Wesentliche Schwachpunkte des CWM sind u.a. unzureichende Metamodelle für:
 - Berechtigungen
 - Semantik von Data Warehouse-Inhalten,
 - Qualitätsmanagement
 - Organisatorischer Data Warehouse Prozess
- SAP BW: Implementierung einer CWM konformen Erweiterung:
 - Sinnvolle Erweiterung geeignet als Standardisierungsgrundlage insbesondere im Bereich Berechtigungen