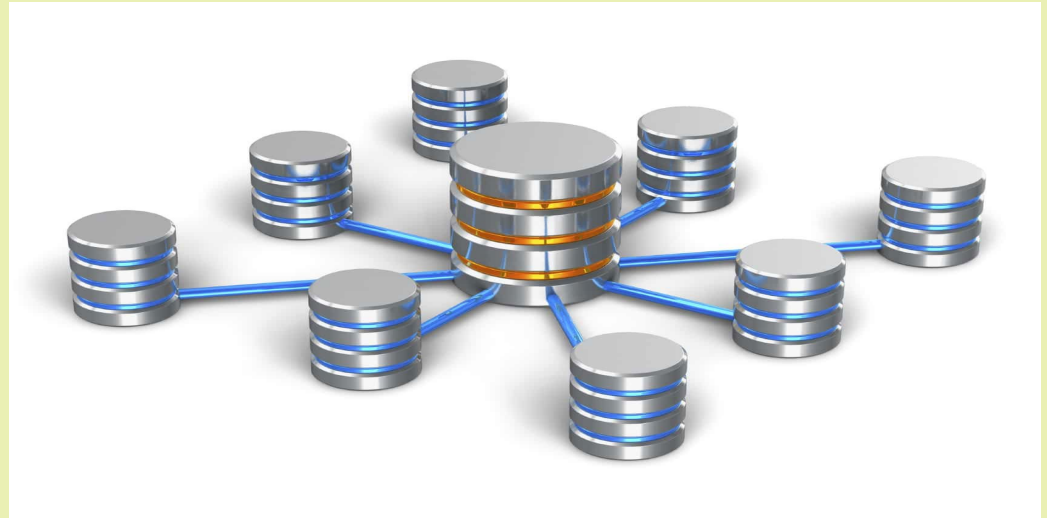


## 2: Datenintegrationsarchitektur

**Prof. Dr. Markus Grüne, FB03**  
**Wirtschaftsinformatik**



# Lernziele

Datenintegrationsarchitekturen kennen lernen

Komponenten von Datenintegrationsarchitekturen kennen lernen

Grundlegende Definitionen kennen lernen

- MDBMS
- SOA
- MDM

# Datenintegrationsarchitekturen

Viele unterschiedliche Architekturen

Kontinuum: Warehousing ... virtuelle Integration

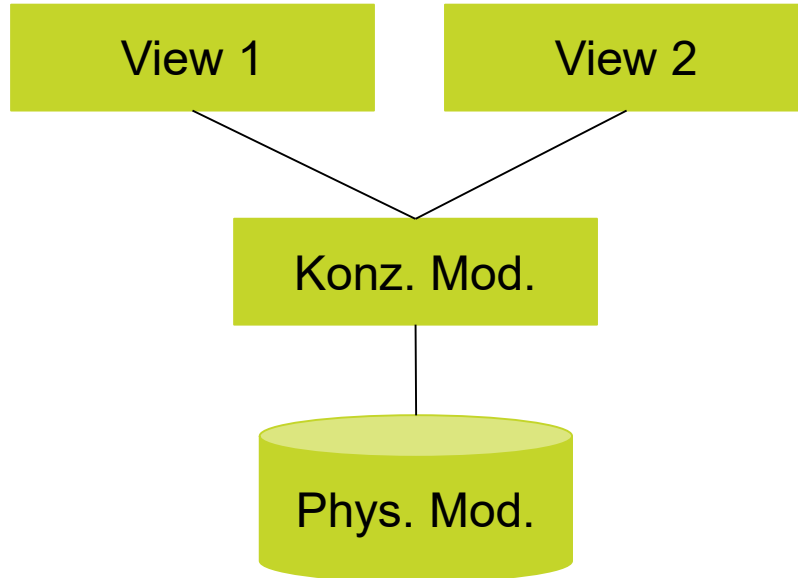
- Warehousing: Daten von Quellen werden geladen und in einer Datenbank materialisiert (warehouse)
- Virtuelle Integration: Daten verbleiben in Quellen und der Zugriff erfolgt erst während der Abfragezeit

Trotz Unterschiedlichkeit der Verfahren, viele gleichartige Herausforderungen

(Doan, Halevy, Ives 2012)

# Überblick – DB-Architekturen

## Monolithische Systeme ...

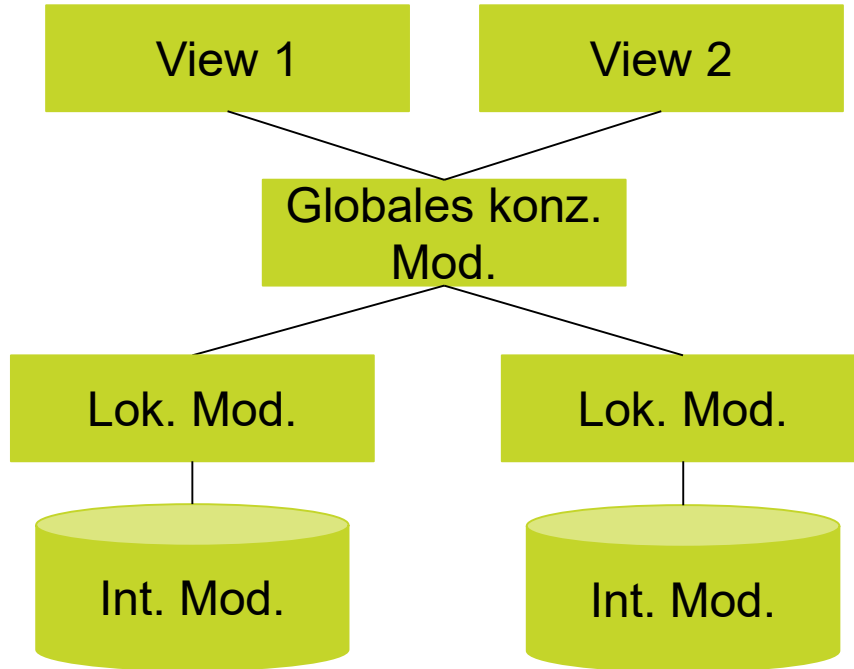


### ANSI-SPARC-3-Schicht

- Externe Schicht
  - für Externe
- Konzeptuelle Schicht
  - Data Model
- Interne Schicht
  - phys. Speicherung

# Überblick – DB-Architekturen

## Verteilte DBMS-Systeme ...



## 4-Schicht-Architektur

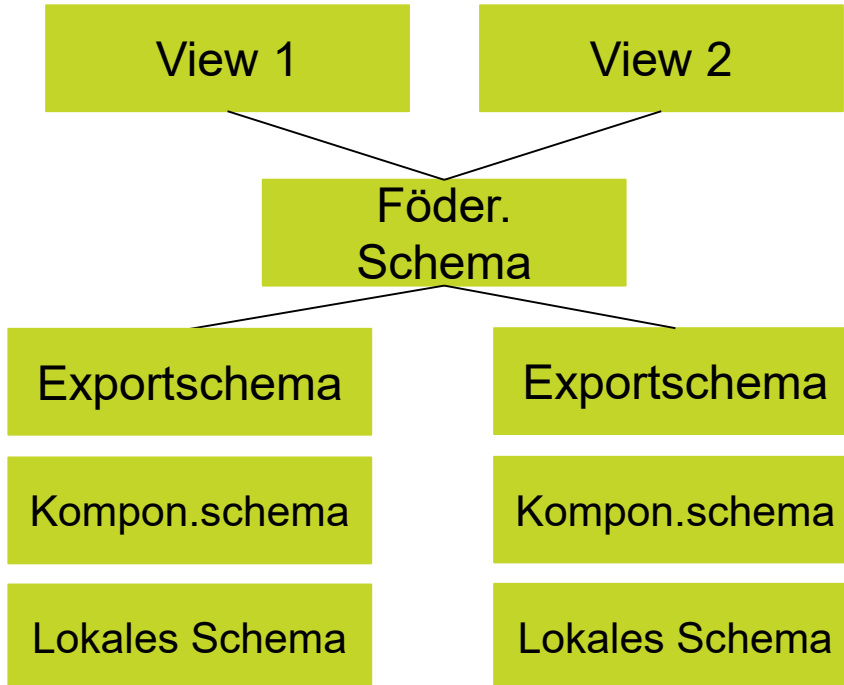
- Externe Schicht
  - für Externe
- globales konzep. Schema
  - Integriert die ..
- lokalen konzep. Schemas
- Interne Schemas
  - phys. Speicherung

## Weitere Architekturen

- Multi-DBMS
  - Lokales Schema wird von den Quellen als Export-Schema umformatiert und bereitgestellt.
  - Multidatenbanksprache ermöglicht den Externen Schemas die Abfrage unterschiedlicher Quellen

# Überblick – DB-Architekturen

## Föderiertes Schema ...



## 5-Schicht-Architektur

- Föderiertes Schema = globales konz. Schema
- Exportschemas = Teilmenge Komponentenschema
- Komponentenschema = Normierung, Mappings, Überwindung Heterogenität
- Lokales Schema = konzept. Schema
- Interne Schemas nicht betrachtet
- Föderiertes Schema wird aus den Exportschemas integriert

# Wichtige Begriffe

## Mediator

- Software, die das Wissen über Daten nutzt, um Informationen für höherwertige Anwendungen zu generieren

## Aufgaben Mediator

- Informationsintegration
- Datentransformation
- Metadaten-Bereitstellung
- Sortieren
- Schlüsselgenerierung
- ...
- Für Experten geschaffen



# Wichtige Begriffe

## Wrapper

- Software, die Kommunikation / Datenfluss zwischen Mediatoren und Quellen herstellen

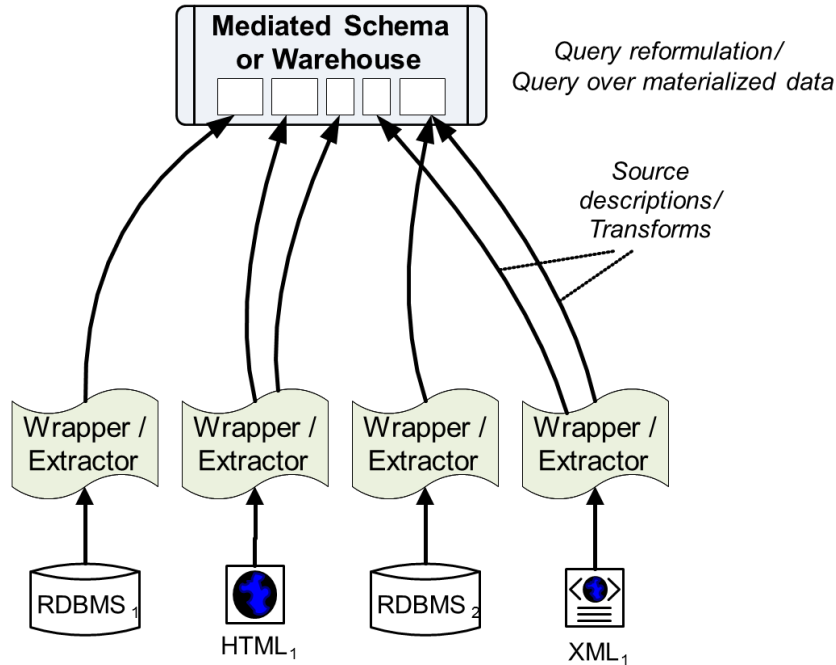
## Aufgaben Wrapper

Heterogenität durch geeignete Schnittstellen auflösen: z.B.

- Datenmodellheterogenität
- Heterogenität der Abfragesprachen

...

# Logische Komponenten im Zusammenspiel Mediated Schema ODER Warehouse



Daten Quellen: RDBMS, XML, Web Forms, ...

Programme, die mit den Quellen kommunizieren

Wrappers: virtuelle Integration

Extractors: materialisiert (Warehouse)

Mediated Schema: Die User interagieren mit dem Data Integration Schema über ein einzelnes Schema

(Doan, Halevy, Ives 2012)

# Logische Komponenten

## I) **Mediated Schema** / Virtual Integration

Die Beschreibungen der Quellen sind essenziell, um ein integriertes Schema zu bauen

- Verbindung zwischen "Mediated Schema" und Schema der Quellen
- Semantische Mappings beschreiben, wie die Attribute in der Quelle mit den Attributen im Mediated Schema zusammen hängen.

## II) **Data Warehouse** / [i.d.R.]

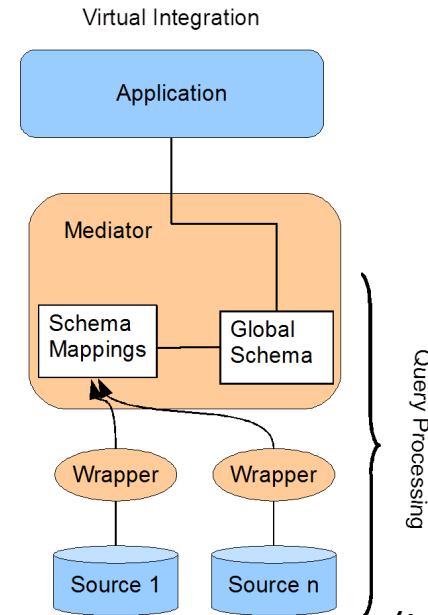
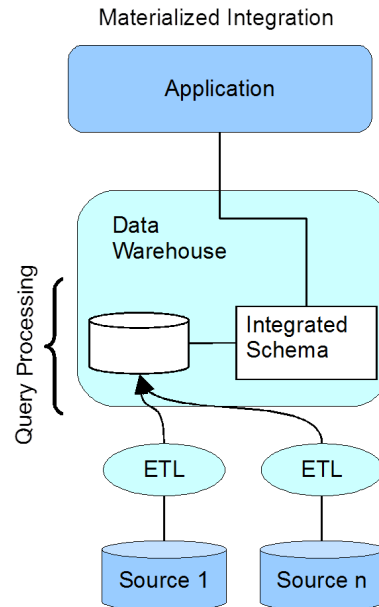
Materialized Integration

User stellen Anfragen an das DWH-Schema

ETL-Pipelines extrahieren Daten von den Quellsystemen

(Doan, Halevy, Ives 2012)

# Logische Komponenten



Ref: [LeNa07 p. 87]

(Leser, Naumann 2007)

# Materialized / Virtual Integration

## Materialized Data Integration

- Integrierte System, die Daten speichern und bereit stellen an andere Applikationen → **Data Warehouse**
- Abfragen auf die Daten werden auf dem integrierten System selbst ausgeführt.
- Originaldaten bleiben in den Quellen
- Daten werden kopiert und transformiert.

## Virtual Data Integration

- Daten werden in kleinen Portionen – on demand – an die Anwendung geschickt
- Kein Kopieren, mögliche Transformationen nur während der Laufzeit, ohne Änderung der Originaldaten

(Leser, Naumann 2007)

# Vergleich

	Materialized integration / DWH	Virtual integration
Datenaktualität	Geringe, abhängig von Aktualisierungsfrequenz	Hoch, direkte Abfragen
Anwortzeit	Kurz	Lang, netzwerkabhängig
Komplexität	Niedrig, ähnlich zu monolithischen Datenbanken	Eher hoch, abhängig von der Varietät und Anzahl der Datenquellen
Query facilities	SQL und Derivate	Abhängig von der Art der Quellen
Speicheranforderungen	Hoch: Komplette Duplizierung der Daten im schlimmsten Fall	Niedrig: nur die Metadaten müssen gespeichert sein
Datenbereinigung	Möglich	Nur schwer möglich

in Anl. an (Leser, Naumann 2007)

# Service-orientierte Architektur

IT-Komponenten wie Datenbanken, Server und Websites werden in Dienste gekapselt  
Die Aufgabe besteht darin, einzelne Dienste so miteinander zu orchestrieren, dass aus ihnen höherwertige Dienste entstehen, z.B. Anwendungen.

Beispiel:

Dienst1:Kundendatenverwaltung; Dienst2:EinkäufeKunde → D3:Rechnungserstellung

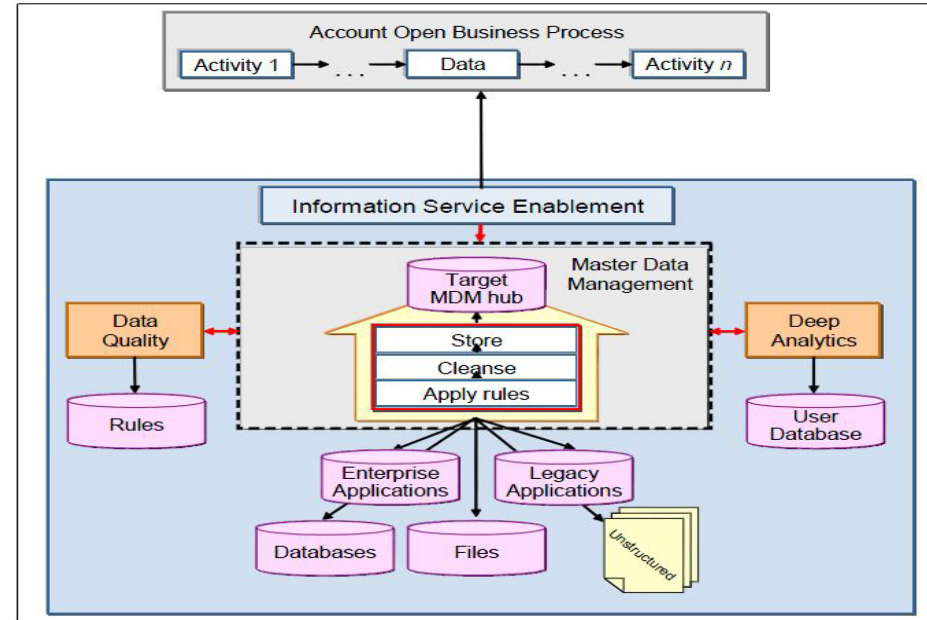
Bei SOA handelt es sich somit um eine Struktur, welche die Unternehmensanwendungsintegration ermöglicht, indem die Komplexität der einzelnen Anwendungen („Applications“) hinter den standardisierten Schnittstellen verborgen wird.

Wichtig: standardisierte Schnittstellen, Verzeichnis zum Auffinden der Dienste,  
Integration der Dienste durch Programmierung

In Anl. (Wikipedia)

# Master Data Management

Menge von Diensten für den Zugriff und das Management von Key Business Data. Agiert als verbindliche, autoritäre Quelle. Verwaltet komplexe Hierarchien von Geschäftsdaten.



(Keen et al. 2008)



# Master Data

- Master Data: Daten über Kunden, Supplier, Partner, Produkte, Material, Konten und andere "critical entities", die i.d.R. in IT-Systemen gespeichert und verwaltet werden.
- Master Data sind die wertvollen Kerninformationen, die verwendet werden, um kritische Geschäftsprozesse im gesamten Unternehmen durchzuführen. Sie sind der Kern jeder Geschäftstransaktion, Anwendung, jedes Reports und jeder Entscheidung.

<http://www-01.ibm.com/software/data/master-data-management/overview.html>

# Key Takeaways

Datenintegration kann auf unterschiedliche Art erfolgen.

Je nach Anwendungsfall kann es notwendig sein, Daten bei der Integration zu laden, zu speichern und weiterzuverarbeiten oder nur während der Laufzeit Daten abzufragen und diese direkt zu transformieren.

Die Datenintegration hängt engt mit dem Bereich Master Data Management zusammen. Bei der Integration von Masterdaten / Stammdaten ist besondere Sorgfalt erforderliche, da es sich um die essenziellen Daten des Unternehmens handelt (Basis der Geschäftstätigkeit).

# Literatur

Doan A, Halevy A, Ives ZG (2012) Principles of data integration. Elsevier/Morgan Kaufmann, Amsterdam, London

Keen, M. et.al. (2008): Case Study: Information as a Service SOA Scenario. IBM Red Paper

Leser, U.; Naumann, F. (2007): Informationsintegration, 1. Auflage dpunkt

Linthicum, D.S. (2006): Understanding the Critical Need for Data Services when Building a SOA Linthicum Group 2006. [www.informatica.com](http://www.informatica.com) Whitepaper