

Data Warehouse

für den Microsoft
SQL SERVER 2000/2005

Begriffe 1

DWH (Data Warehouse) ist eine fachübergreifende Zusammenfassung von Datentabellen.

DataMart ist die Gesamtheit aller Datentabellen für einen fachlich abgegrenzten Bereich. Zum Beispiel: Vertrieb, Finanzen, Produktion.

Dimension, ist eine Menge miteinander verknüpfter Entitäten, nach der Maßzahlen aggregiert werden können und die zu ihrer Interpretation herangezogen wird. Beispiel: zeitliche Dimension (Tag, Monat, Quartal, Jahr), Produktkategorie

Measure, ist eine Information, die zur Bewertung der Leistung oder des Verhaltens eines Geschäftsprozesses dienen kann.

Faktentabelle, ist eine Menge miteinander verknüpfter Maßzahlen und ihrer zugehörigen Dimensionen.

Begriffe 2

Stagingarea ist eine Datenbank in der alle extrahierten Daten eines Quellsystems gespeichert werden.

Deltaarea besteht oft aus mehreren Datenbanken. Das Ziel ist, zu einem möglichst frühen Zeitpunkt nur die tatsächlichen Veränderungen des Quellsystems weiter zu verarbeiten. Performanceoptimierung!

Transformationsarea, eine Datenbank, in der die Deltadaten einem clearing (d.h. inhaltlich und faktisch bereinigt) unterworfen werden und vorhandene Daten je nach Bedarf transformiert werden. Z.B. entsteht aus einem Begriff ein Schlüssel, es werden Werte umgerechnet usw.

DataWarehouse, aus der Transformationsarea wird das DWH aufgebaut.

DataMart, aus dem DataWarehouse werden die DataMarts erstellt.

Cubebereich, aus dem DataWarehouse/DataMarts werden die Dimensionen und Faktentabellen für die Cubes erstellt.

Modellierung

Modellierung



Aufgabenblöcke

- Modellieren Stagingarea mit den einzelnen Ebenen
- Modellierung der Ebenen der Stagingarea je Quellsystem
- Modellierung der Transformationsarea je Quellsystem
- Modellierung der Dimensionen und Faktentabellen
- Modellierung der Dimensionen und Faktentabellen je Quellsystem
- Integration Dimensionen/Faktentabellen je Quellsystem ins DWH
- Modellierung des DWH mit nach gelagerten Systemen
- Modellierung von DataMarts und Cubes
- Entwicklung von Design Guides für einzelne Aufgabenblöcke
- Konkretisierung der Vorgehensweise (Teil 1 – Teil 5)

Modellieren Stagingarea mit den einzelnen Ebenen



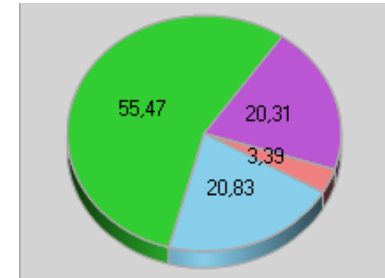
Stagingarea

Je Quellsystem:

- Datenbank Extraktion
- Datenbank Delta
- Datenbank Transformation
- Fehlerdatenbank

Übergeordnet:

- Qualitätsdatenbank
- Datenbank Hilfstabellen



Ebenen der Stagingarea

Modellierung der Ebenen der Stagingarea je Quellsystem

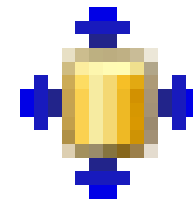
Unterstützung bei der Modellierung der einzelnen Ebenen innerhalb der Stagingarea für das jeweilige Quellsystem.



Stagingarea

Je Quellsystem:

- Datenbank Extraktion
- Datenbank Delta
- Datenbank Transformation
- Fehlerdatenbank



Modellierte Ebenen
der Stagingarea

Übergeordnet:

- Qualitätsdatenbank
- Datenbank Hilfstabellen

Modellierung der Transformationsarea je Quellsystem

- **Schlüsselbehandlung**
 - z.B. ein Präfix = KD oder CRM
- **Anpassung Datentyp**
 - z.B. von Char “010” in int 010
- **Konvertierung von Codierungen**
 - über Ersetzungstabelle den Wert “2” in “PRODUKT” ändern
- **Ersetzen Zeichenfolge**
 - Ändern Feldinhalte über Ersetzungstabelle oder neues Feld erzeugen
- **Vereinheitlichen Zeichenketten**
 - z.B. eliminieren von Leerstellen
“ Adam” in “Adam”
- **Datum in einem Feld**
 - z.B. “1.August 2003” in “01.08.2003”
- **Datum in mehreren Feldern**
 - aus Tabelle Monat und Tabelle Jahr in neues Feld = 12.2006
- **Berechnungen**
 - z.B. Menge * Preis = Wert
- **Kombination Attributwerte**
 - mehrere Feldinhalte in einem Feld zusammen fassen
- **Separierung Attributwerte**
 - Zerlegen eines Feldinhaltes in mehrer Felder

Modellierung der Dimensionen und Faktentabellen

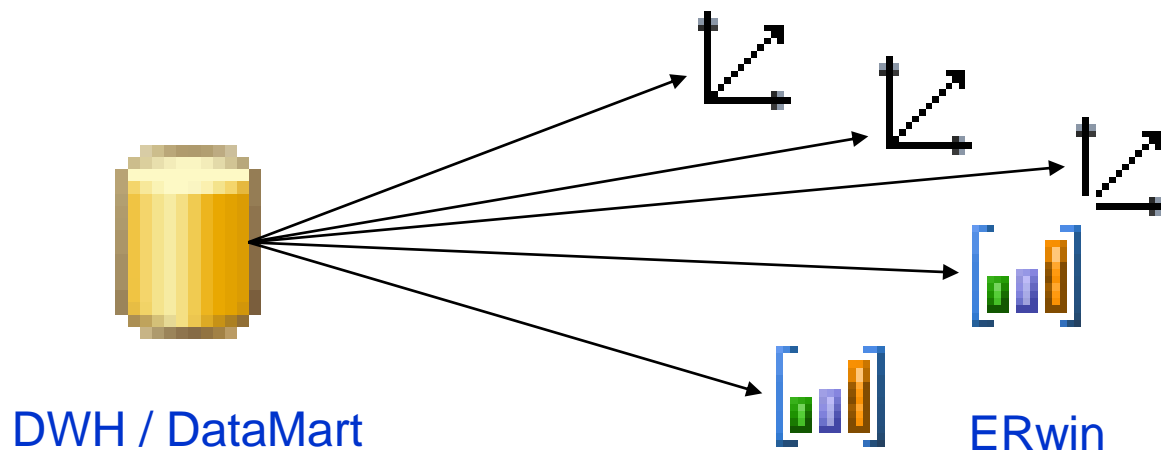
Gliederung der transformierten Felder in Feldgruppen

Feldgruppen...

- **Schlüsselfeld**
 - z.B. Produkt-Nr., Kunden-Nr. (Key's)
- **Wertefeld**
 - z.B. Menge, Beträge
- **Informationsfeld**
 - z.B. Produktbeschreibung, Informationen
- **sonstige Felder**
 - alles anderen Informationen, die nicht in Schlüsselfeld, Informationsfeld und Wertefeld stehen

Modellierung der Dimensionen und Faktentabellen

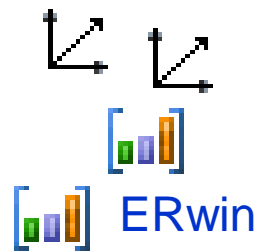
Modellierung der Dimensionen und Faktentabellen anhand des DWH/DataMart.



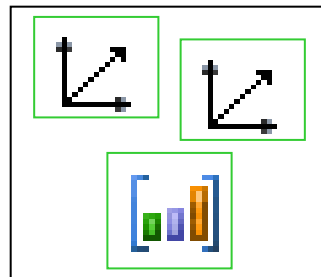
Integration Dimensionen/Faktentabellen in den Cubebereich

Die Dimensionen und Faktentabellen werden mit anderen Quellsystemen abgeglichen und integriert.

Von Quellsystemen
unabhängig modellierte
Dimensionen und Fakten

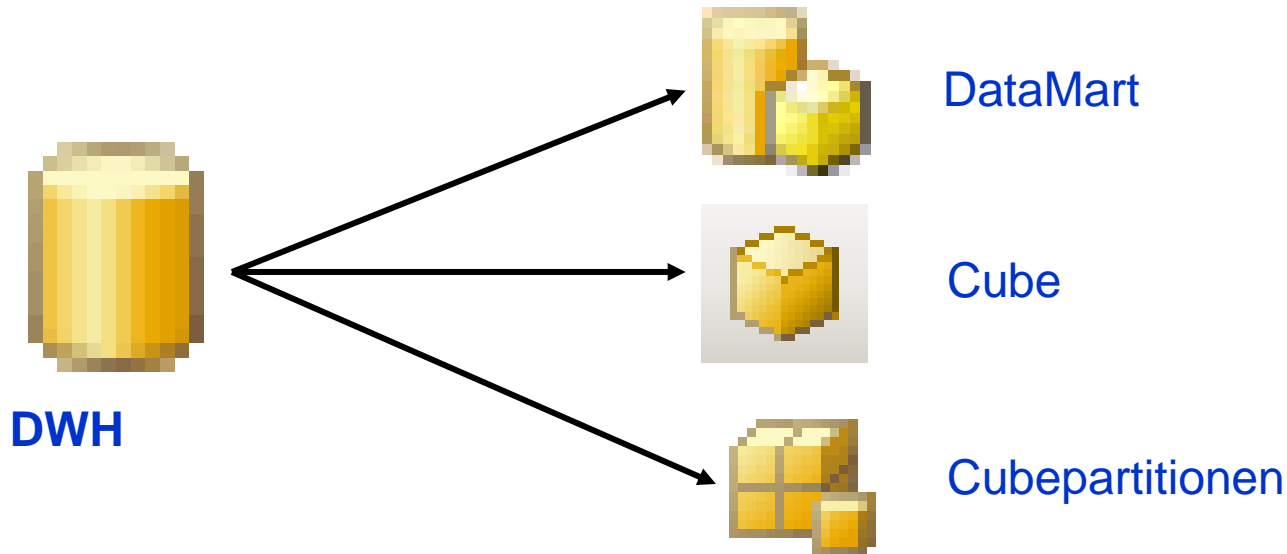


Im Cubebereich
vorhandene
Dimensionen und Fakten



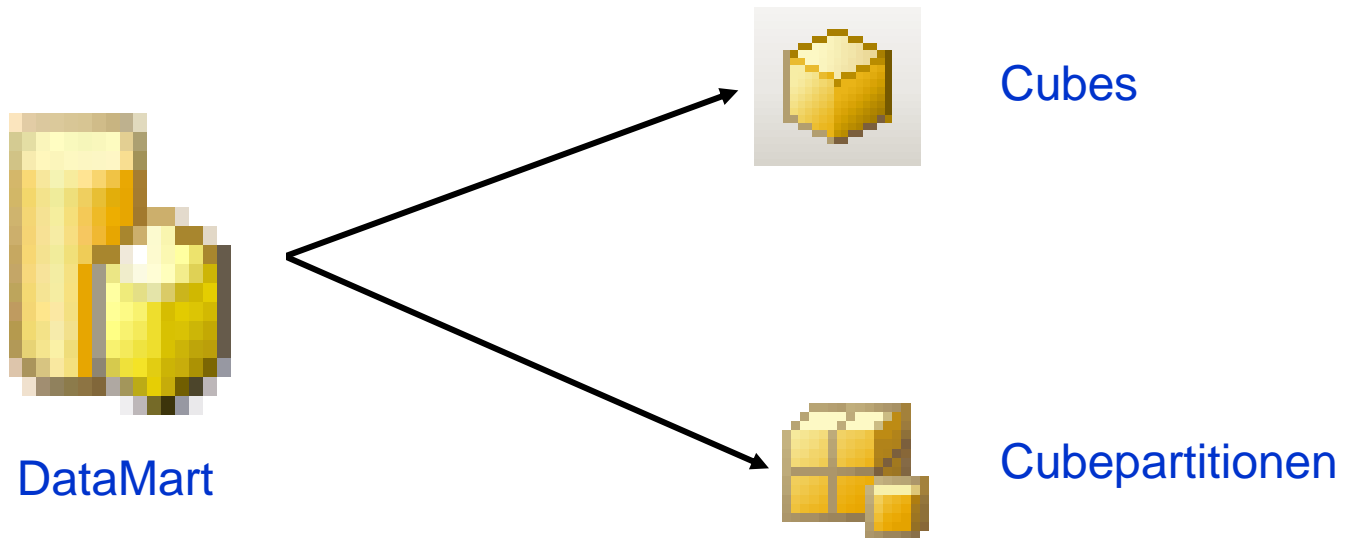
Modellierung des DWH mit nach gelagerten Systemen

Übergeordnet zu allen Quellsystemen wird ein DWH mit abhängigen DataMarts, Cubes und Partitionen für Cubes modelliert.



Modellierung von DataMarts und Cubes

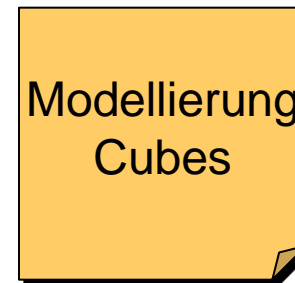
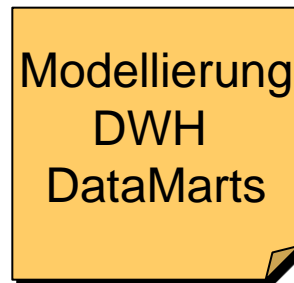
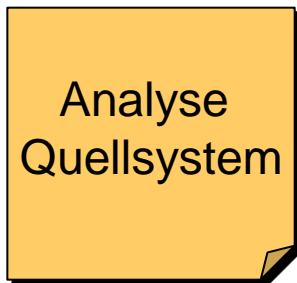
Für fachlich abgegrenzte Quellsysteme können unabhängige DataMarts, Cubes und Partitionen für Cubes modelliert werden.



Entwicklung von Design Guides für einzelne Aufgabenblöcke

Für die einzelnen Aufgabenbereiche der Modellierung und Integration werden spezifische Design Guides entwickelt.

Zum Beispiel:



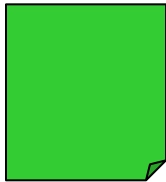
Konkretisierung der Vorgehensweise

Ziele der Vorgehensweise

- Revisionsfähigkeit
- Wartungsfreundlichkeit
- Einfache Erweiterbarkeit
- Einfache Migration auf die nächste SQL-Server Version oder auf ein anderes DWH-System

Konkretisierung der Vorgehensweise Teil 1

Extraktion



- Definition der Extraktionsregeln
- Erstellen des Datenmodells für die Extraktions - DB



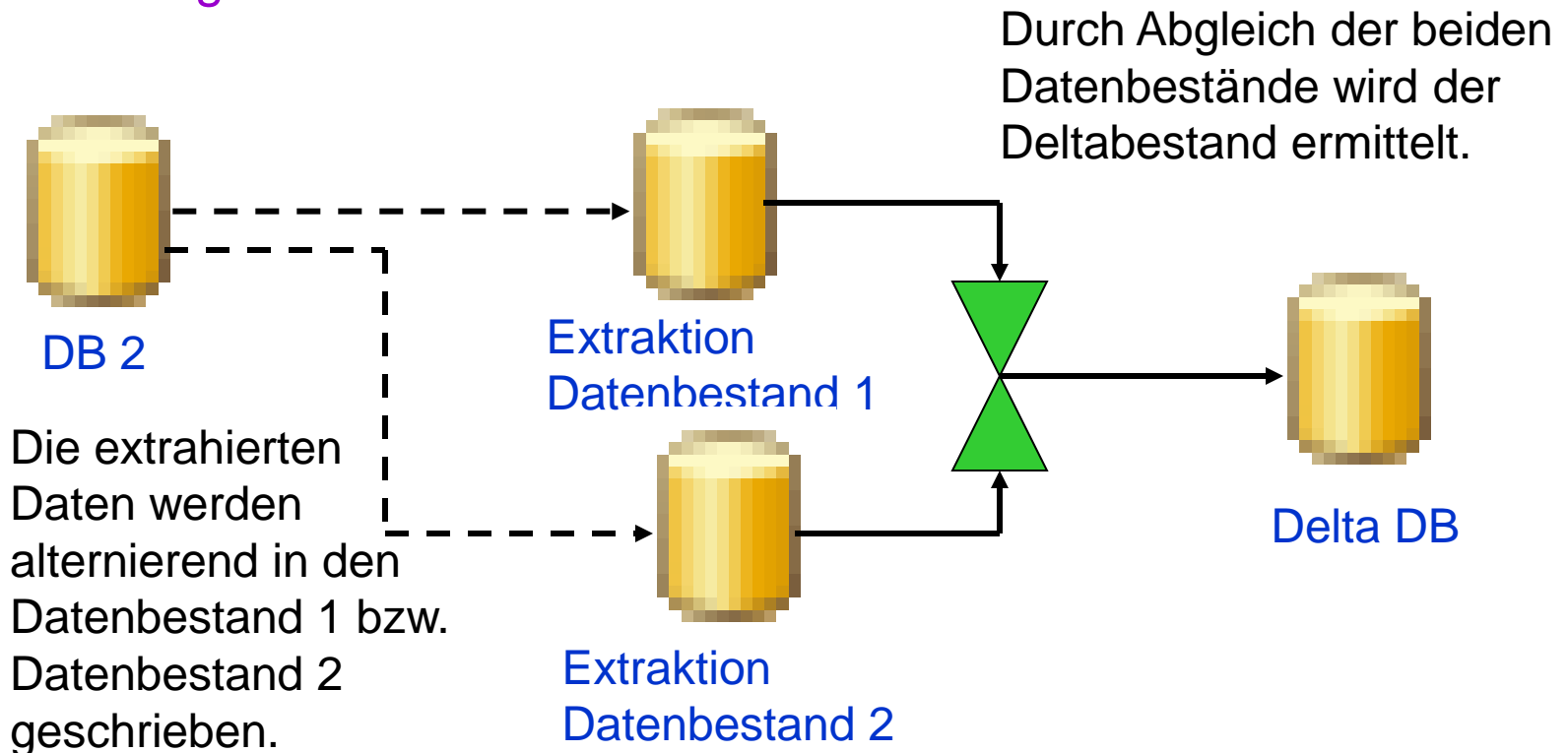
DB2



Extraktions - DB

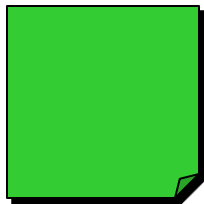
Konkretisierung der Vorgehensweise Teil 2

Deltabildung

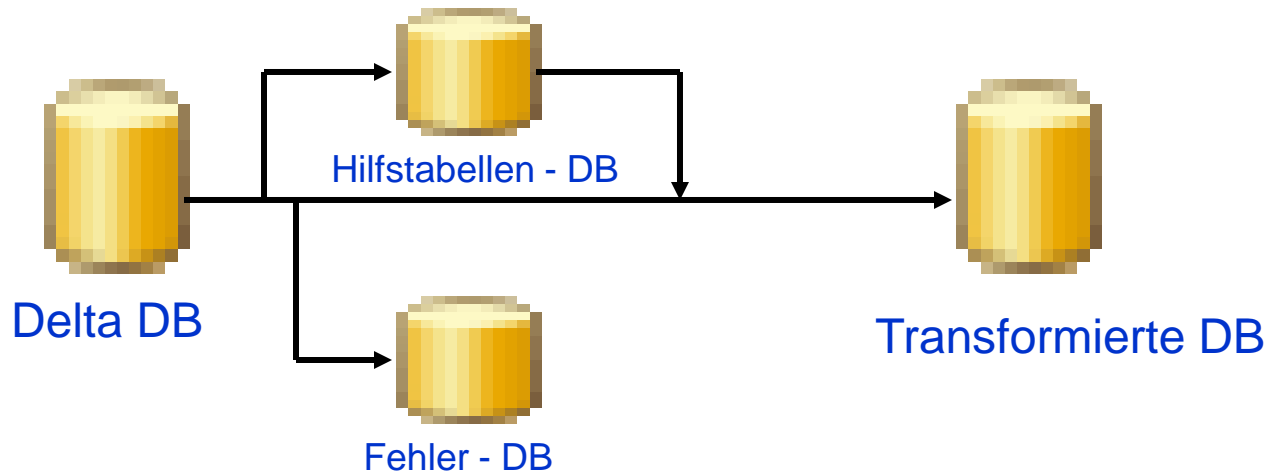


Konkretisierung der Vorgehensweise Teil 3

Transformation

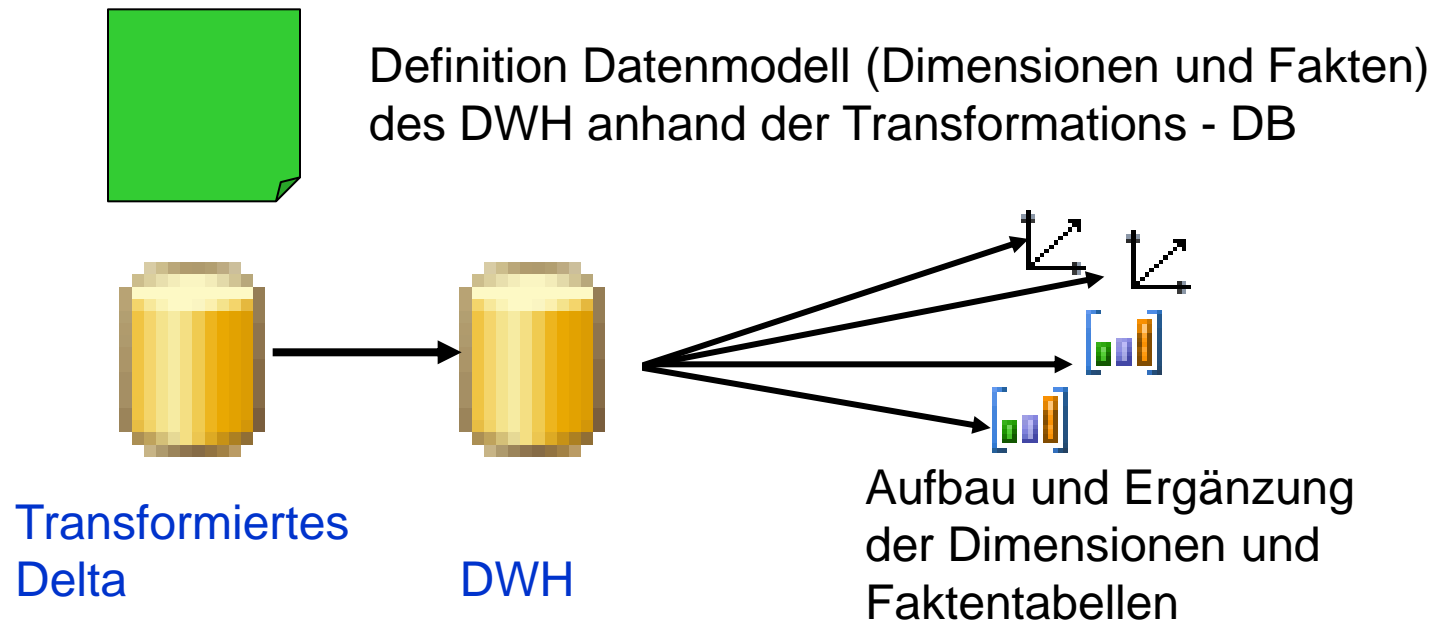


Definition der Transformationsregeln
Definition des Datenmodells der transformierten Daten



Konkretisierung der Vorgehensweise Teil 4

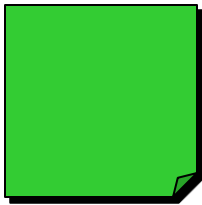
Laden der Deltadaten in das DWH



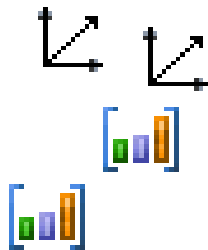
Konkretisierung der Vorgehensweise

Teil 5

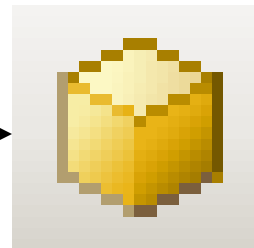
Erstellen der Cubes



- Definition Datenmodell des Cube
- Definition berechneter Felder (Berechnungsregeln)
- Definition berechneter Zellen (Teilbereiche des Cubes)



Cubebereich



Cube

Business Intelligence

Business Intelligence heißt, eine globale Sicht auf die Transaktionen zu gewinnen, um daraus bessere und schnellere Entscheidungen für das Geschäft abzuleiten.

Business Intelligence

Was ist Business Intelligence?

Physische Basis für BI

Möglichkeit 1: keine aggregierten Daten

Möglichkeit 2: keine atomaren Daten

Möglichkeit 3: bereinigte, aggregierte und historische Daten

Business Intelligence

Verlässliche, flexible und übergreifende Reports

- Nur **eine** Interpretation der Kennzahlen
- Schnelle Lieferung Ihres Wunschreports
- "Alle" Zahlen sind verfügbar
- Benchmarking mit Konkurrenz (externe Daten)

Komplexe Analysen und Darstellungen

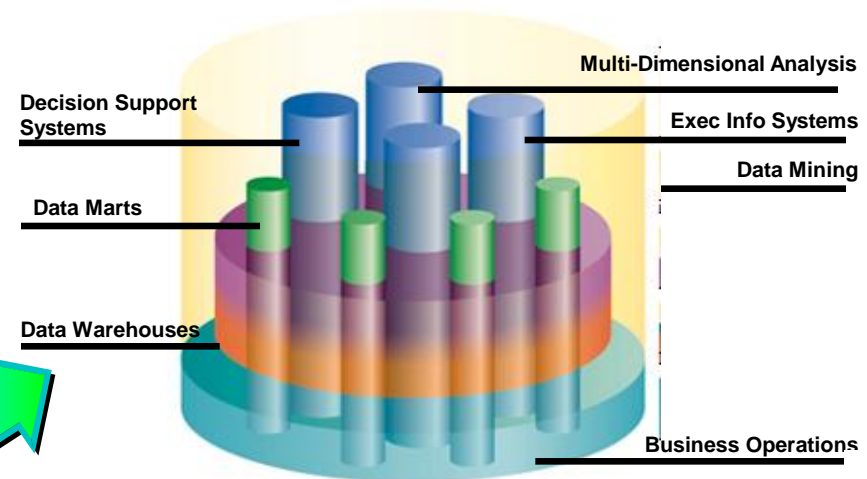
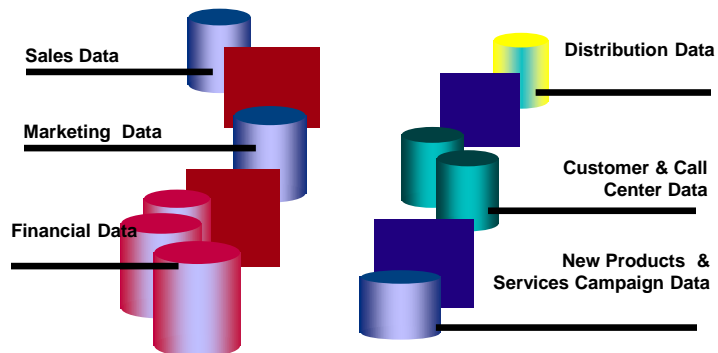
- Multidimensionale Analysen, z.B. was, wann, wo
- Planungen, was-wäre-wenn Analysen
- Unterschiedliche Abstraktionsstufen mit Möglichkeiten zum Nachbohren im Detail
- Geographische Darstellungen, Managementcockpits

Aufdeckung vorher unentdeckter Zusammenhänge

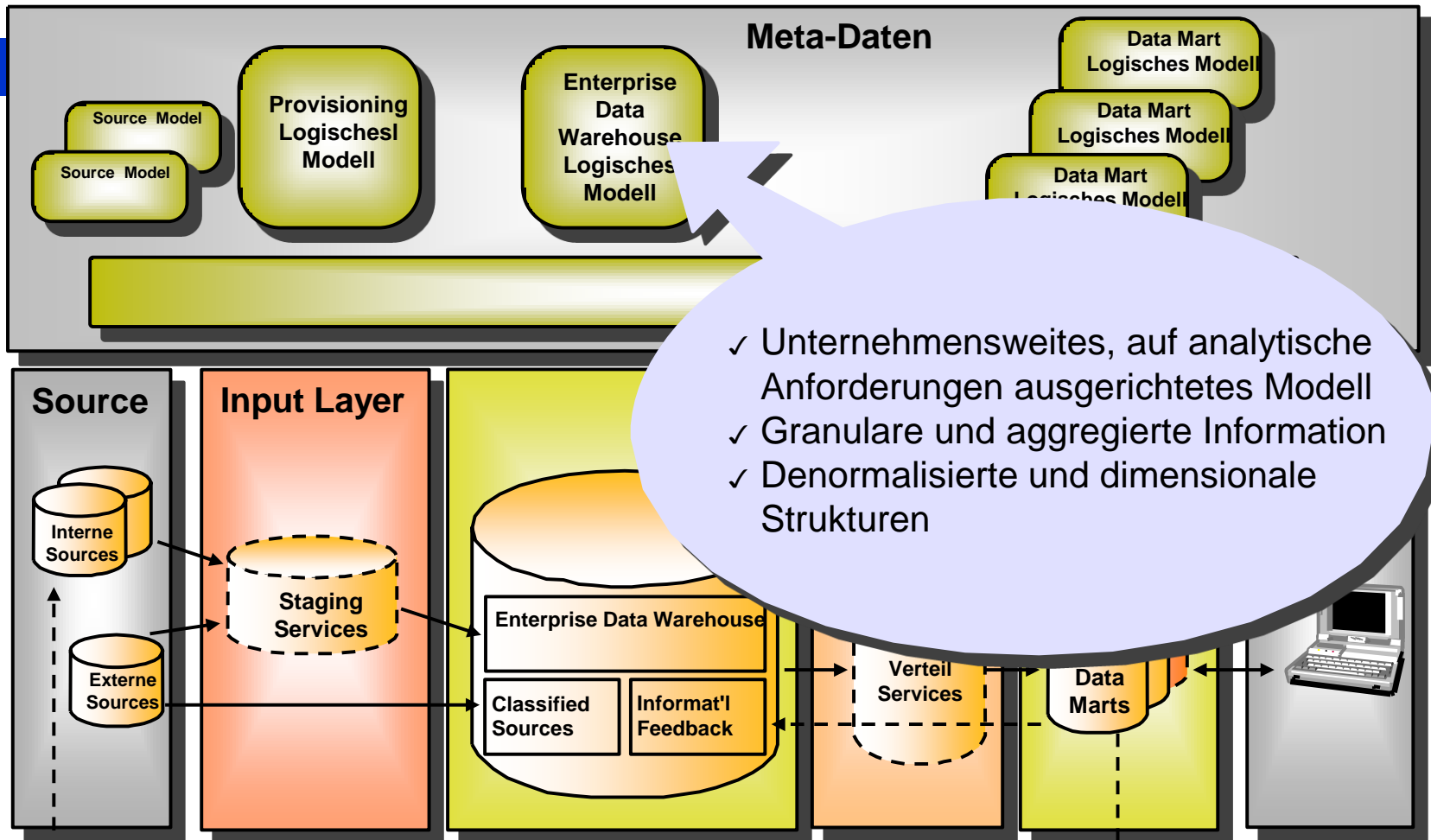
- Segmentierung
- Verhaltensmusteranalyse
- Optimierung von Wegen, Plazierungen, etc.

Aus technischer Sicht sorgt Business Intelligence für eine konsolidierte Datensicht auf Automationsinseln

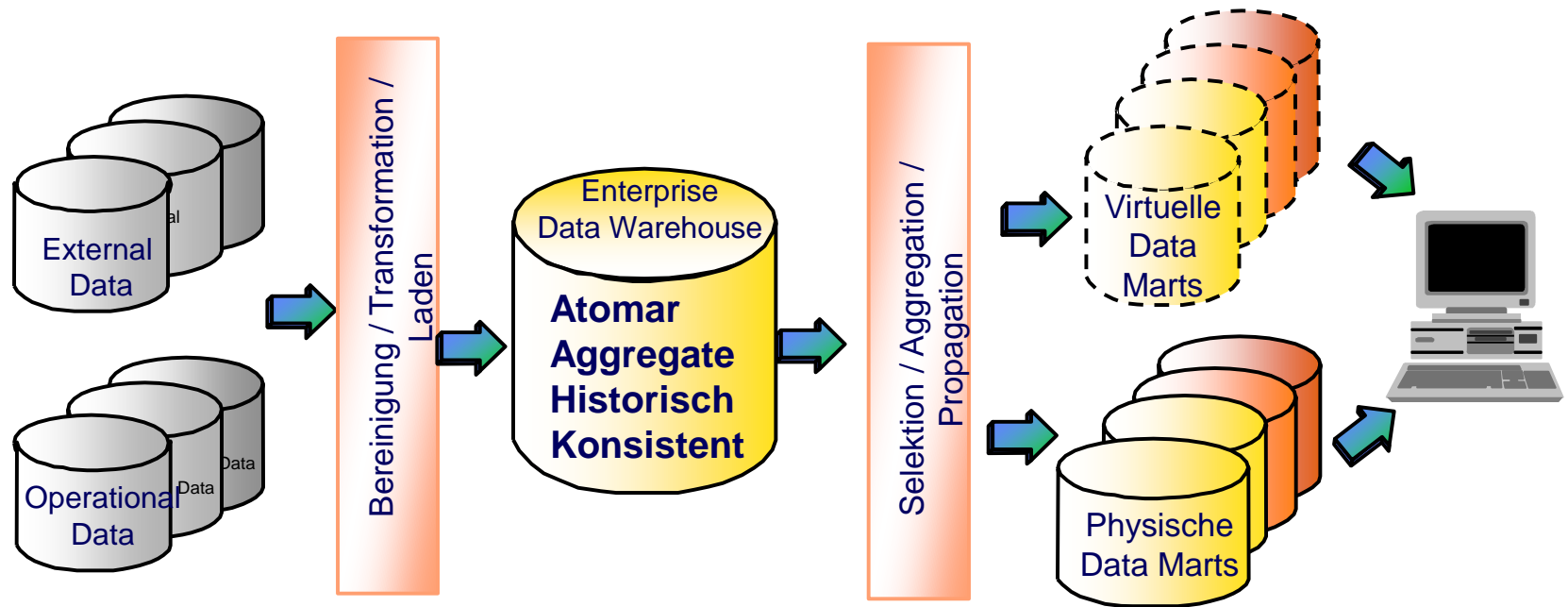
- bessere Datenkonsistenz
- bessere Datenqualität
- Historie
- übergreifender Zugriff
- Auswertung mit modernen Werkzeugen
- vereinfachte Systemverwaltung
- weniger Last auf OLTP Systemen



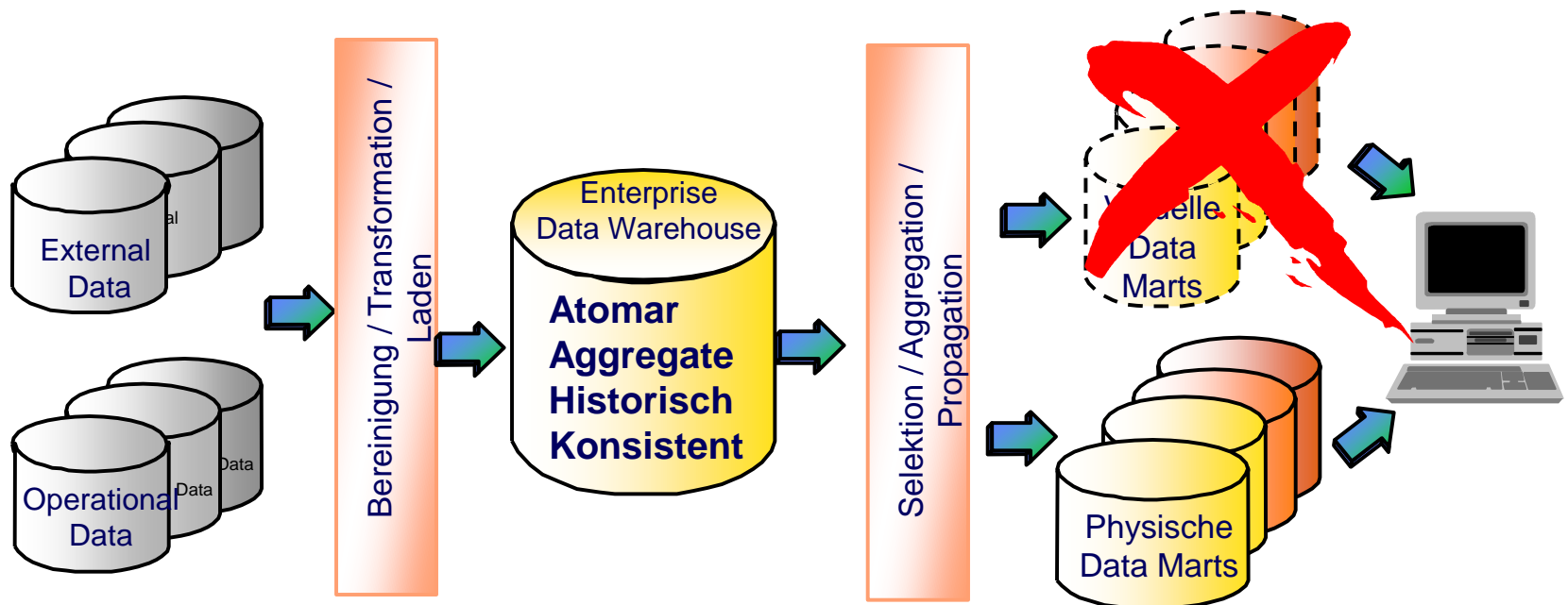
Enterprise Data Warehouse Modell



Generell gibt es zwei Möglichkeiten zur Implementierung von Data Marts



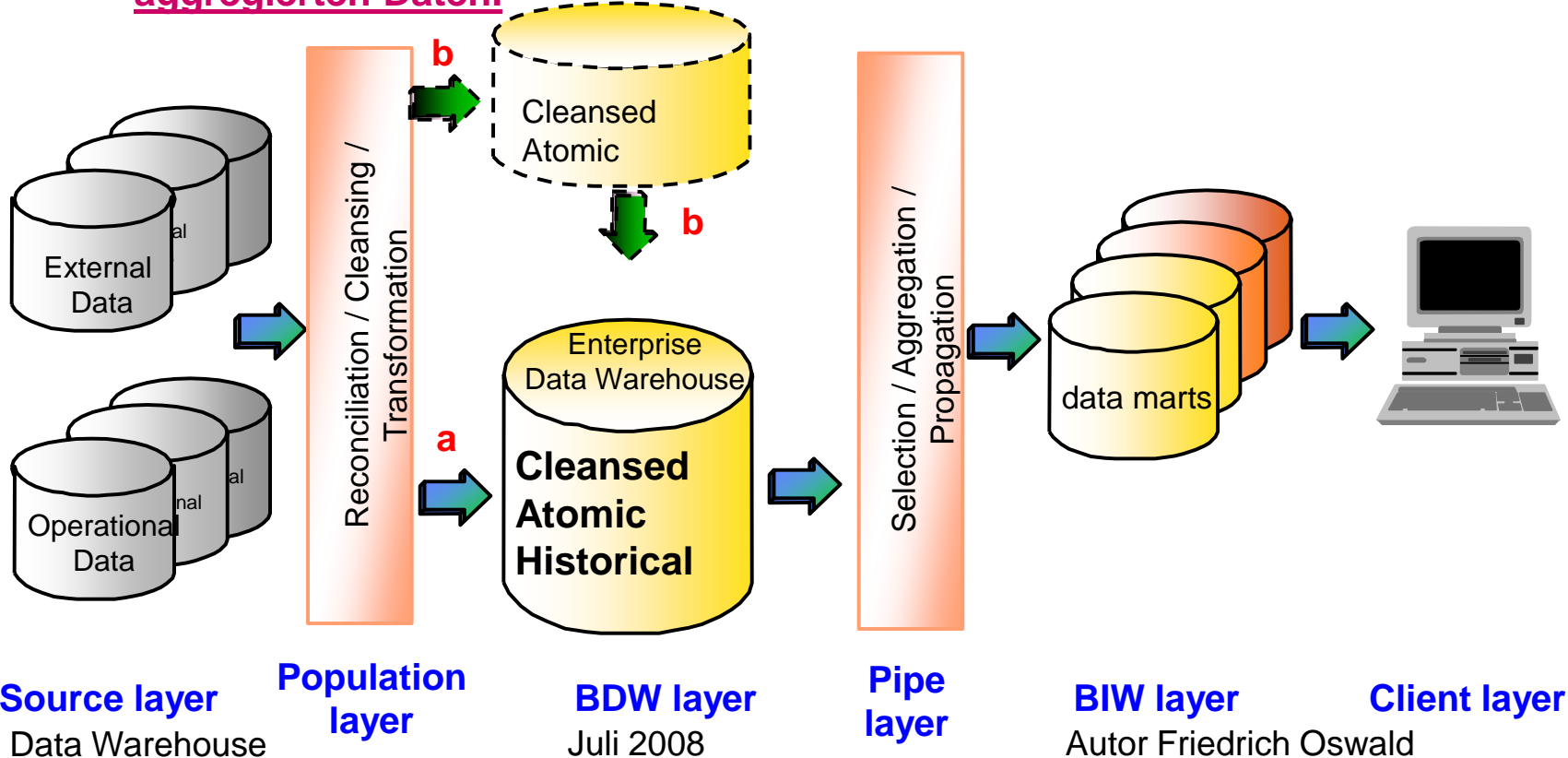
... doch nur eine hat sich in der Praxis bewährt



Physische Implementierung: 1. Möglichkeit



Die BDW Schicht enthält bereinigte, atomare und historische Daten, keine aggregierten Daten.



Physische Implementierung:

1. Möglichkeit

Vorteile

- Spart Plattenplatz, da keine aggregierten Daten im Enterprise Data Warehouse gehalten werden.
- Empfehlenswert, solange nur wenige Data Marts.
- Einfache Erweiterungen.

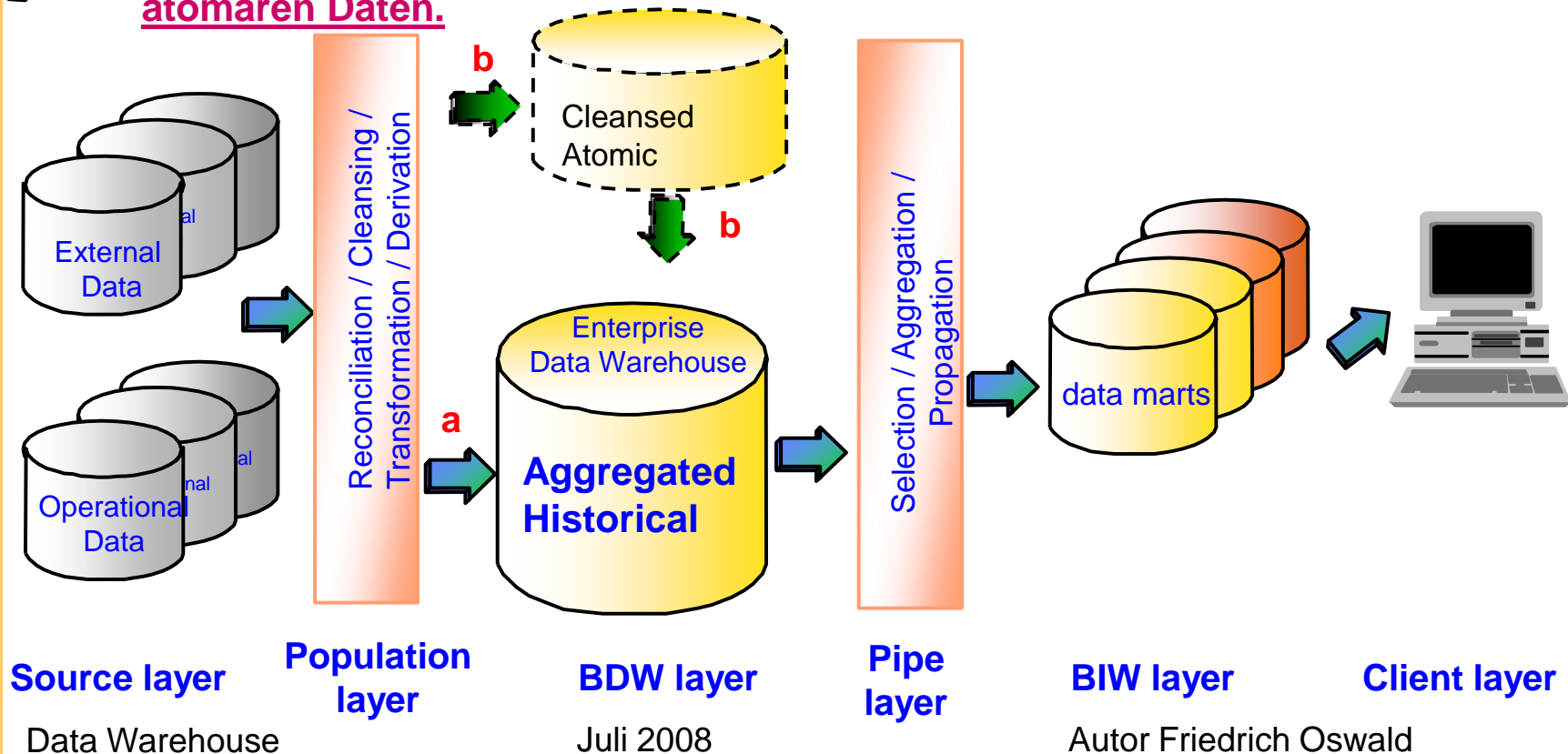
Auswirkungen

- Gefahr von Inkonsistenzen bei der Ableitung und Aggregation, falls es für einzelne Data Marts Gemeinsamkeiten gibt.
- Möglicherweise eine Übergangslösung, bis mehrere Data Marts entwickelt werden.
- Auf Ebene des Enterprise Data Warehouse sind keine Abfragen möglich.

Physische Implementierung: 2. Möglichkeit



Das Data Warehouse enthält aggregierte und historische Daten, keine atomaren Daten.



Physische Implementierung:

2. Möglichkeit

Vorteile

- Spart Plattenplatz, da nur die aggregierten Daten gespeichert werden.
- Empfehlenswert, wenn die operationalen Daten "sauber" sind

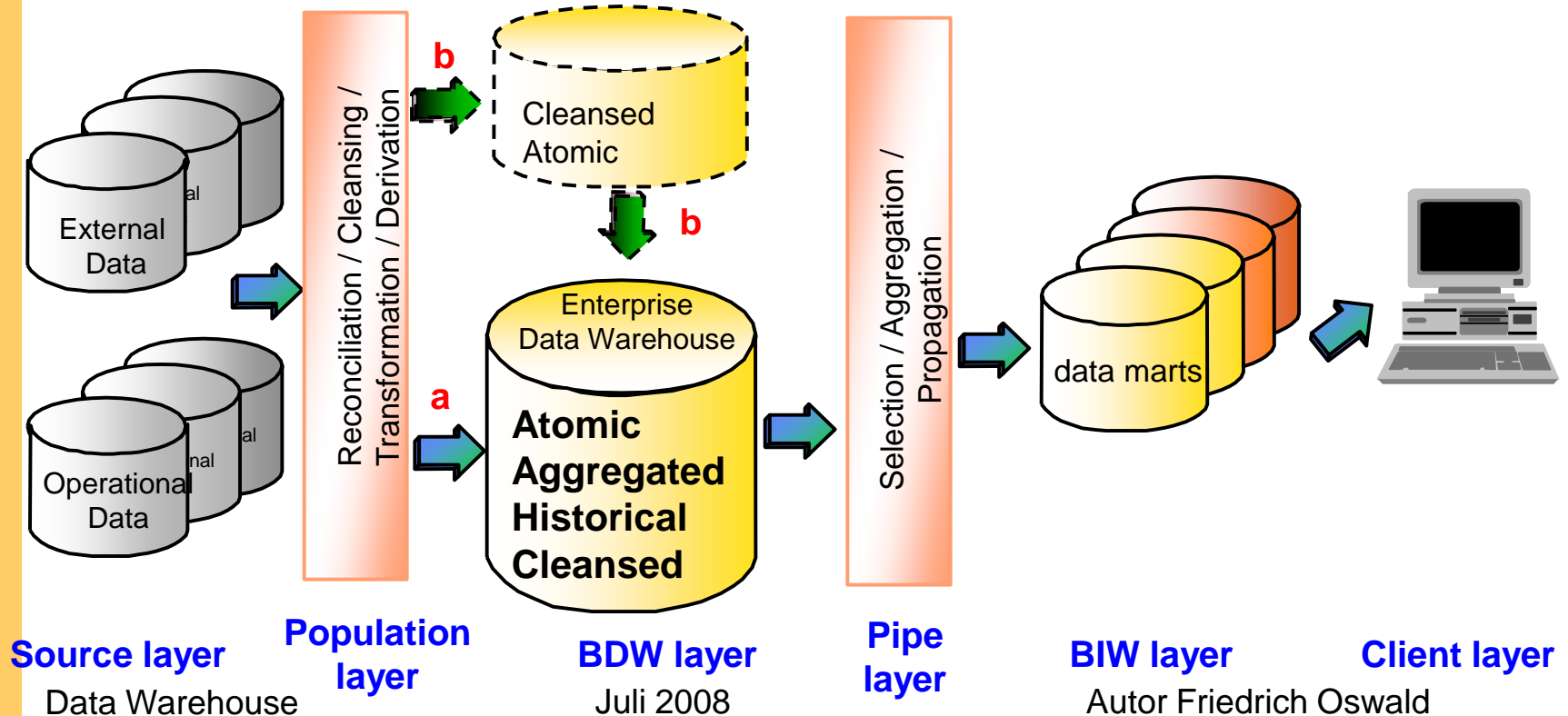
Auswirkungen

- Die Ableitung aus den operationalen Daten ist komplexer.
- Kein Zugriff auf atomare Daten im Enterprise Data Warehouse möglich.

Physische Implementierung: 3. Möglichkeit



Die BDW Schicht enthält bereinigte, atomare, aggregierte und historische Daten.



Physische Implementierung:

3. Möglichkeit

Vorteile

- Die "sauberste" Lösung.
- Klare Trennung zwischen
 - cleansing Schicht
 - transformation Schicht
 - propagation Schicht.

Auswirkungen

- Viel Plattenplatz.
- Aufwendiger zu entwickeln und zu pflegen.
- Für einfache Probleme zu komplex.

Danke...

...für Ihre Aufmerksamkeit

