

# Data Warehousing

Unternehmen und Daten

# Begriffe

## **Datenverfügbarkeit / Data Availability**

Dienste oder Programme, welche die

→ Verfügbarkeit von Daten zu einem vorgegebenen Leistungsniveau unter allen Umständen sicherstellen.

→ Frage nach Speicherort und Zugang

### **In Unternehmen:**

Fragestellung für Business Intelligence (BI)  
wer hat Zugriff und worauf?

Wie schnell kann man an die Daten gelangen  
und durch welche Instanzen/Tools?

## **Data-Mining**

*„We are drowning in data but starving for knowledge“*

→ Methoden und Algorithmen zur Aufdeckung diverser Zusammenhänge der Unternehmensdaten

→ Unterstützung taktischer und strategischer Unternehmensentscheidungen.  
NICHT eine homogene Methode, sondern ein Sammelbegriff

## **Data-Warehouse / Data Warehousing**

Zusammenführung von Daten aus diversen Quellen in einer physischen Datenbank für spätere Analysezwecke

# Gründe für die Notwendigkeit des Data Warehousing

- in operativen Systemen des Unternehmens nur kurzfristige Speicherung
- Mitbetrachtung historischer Daten für aussagekräftige Informationen
- Datenlast würde operatives System ausbremsen oder überfordern  
→ Auslagerung in separates Datawarehouse

## **Merkmale:**

- zentrale Verwaltung
- Archivierung
- schneller Zugriff
- Unabhängigkeit von Datenquellen und Analyseinstanz
- Benutzerunabhängigkeit

→ langfristige Speicherung

→ Bereitstellung für Auswertung und Analyse

## Betreiben eines Datawarehouse

- Aufbau einer separaten Entscheidungsdatenbasis zur Unterstützung dispositiver Aufgaben
- Im Idealfall unternehmensweit ausgerichtet
- Abdeckung des Informationsbedürfnisses verschiedenster Anwendergruppen
- Entkoppelung eines zentralen Data Warehouse von den datenliefernden Vorsystemen
- führt einerseits zu einer Entlastung der operativen Systeme und eröffnet andererseits die Option, das analyseorientierte System auf die Belange von Auswertungen und Berichten hin zu optimieren
- Versorgung und Unterhaltung der untergeordneten "Data Marts" (separate Datenbestände, kleine Datenpools)

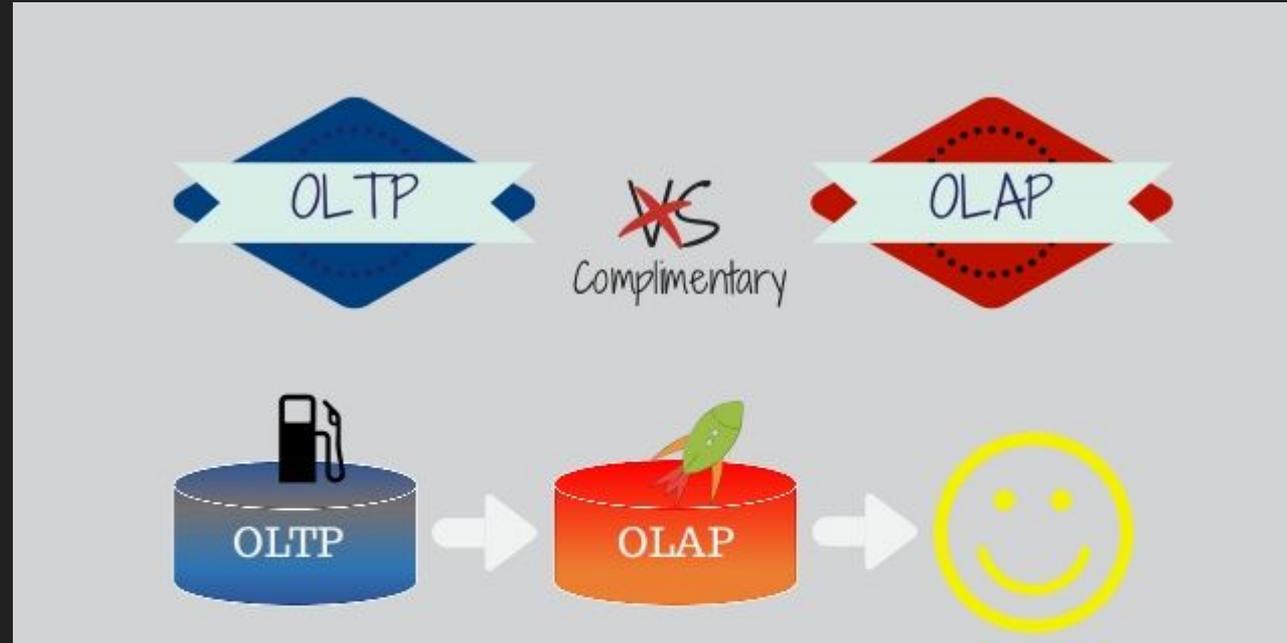
# OLTP vs. OLAP

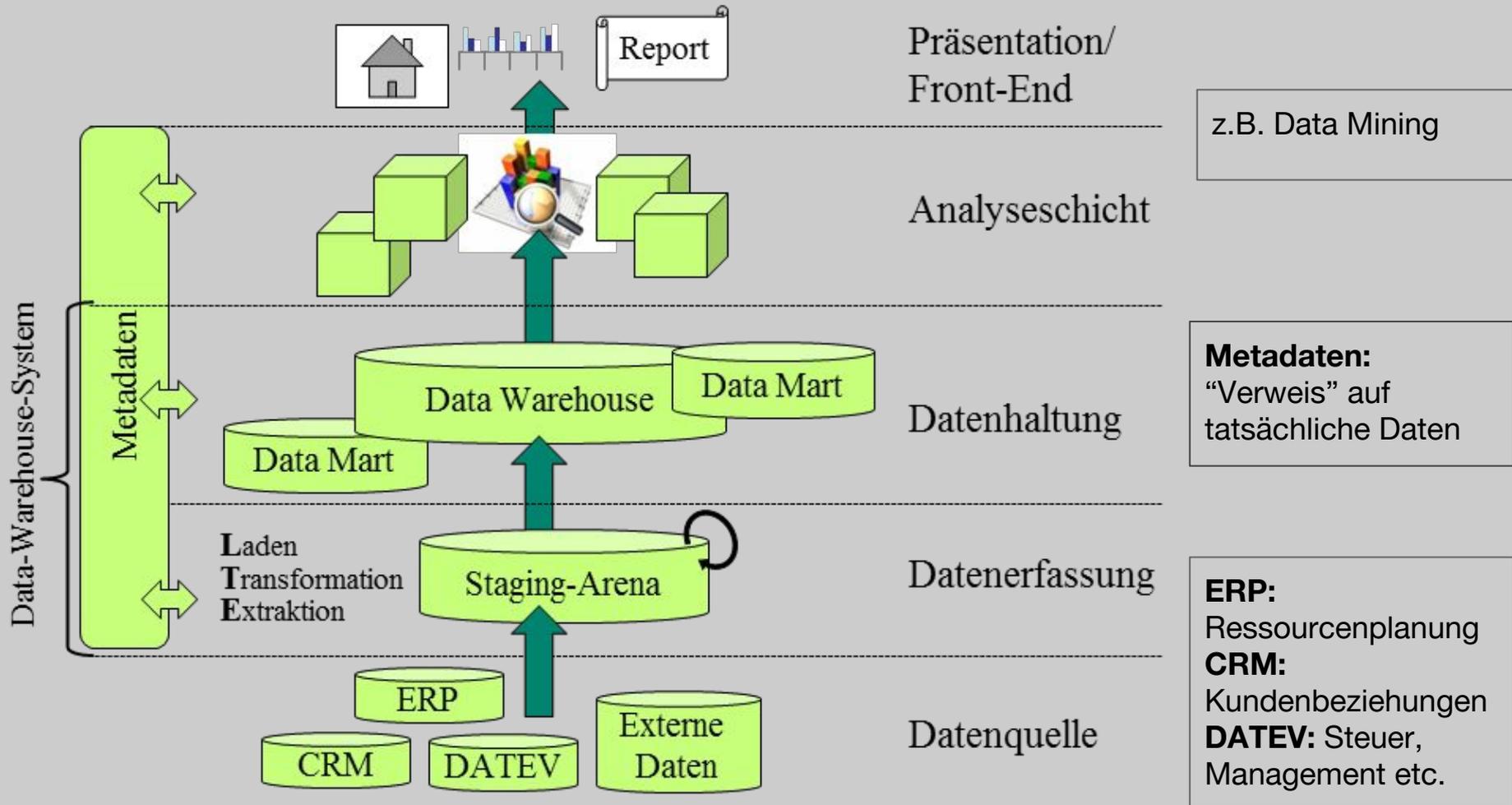
OLTP (Online Transactional Processing)

→ operative Dienste

OLAP (Online Analytical Processing)

→ dispositive Dienste, z.B. Data Warehouse und Data Marts





## Data Warehouse

- Gesamtheit aller Unternehmensdaten
- große Datenmenge → für umfassende historischen Analysen
- unabhängig von Anwender und Anwendung
- komplexer gestaltet

→ multidimensionale Einheit am Ende des Vereinheitlichungsprozesses

## Data Mart

- spezifischer Teildatenbestand (z.B. Verkaufszahlen)
- kleinere Datenmenge → schnellerer Zugriff
- Anwender- und anwendungsspezifisch
- weniger komplex gestaltet

→ Insellösungen für einzelne Bereiche

# Beispiel: Blutdatenbank mit Data Warehouse

## Spenderdaten

- Identifikation
- Alter
- Blutgruppe, Antikörper
- Auffälligkeiten
- Spendemenge / Produkt
- Spendezeitpunkt

→ Eingabe über OLTP des jeweiligen Spendeorts



## Blutdatenbank

z.B. landesweit

- Langzeitspeicherung und garantiert Zugriff für weitere Analyse
- Link zwischen Spender und Empfänger
- Abstrakte ID → Datenschutz!



## Empfängerdaten

- Identifikation
- Alter
- Diagnose
- Blutgruppe
- erhaltene Menge
- Komplikationen/Überleben
- Länge des KH-Aufenthalts

→ Eingabe der Daten über OLTP des jeweiligen Krankenhauses



## Analyse-Instanzen (Forschung, Kliniken, Universitäten)

- Risikofaktoren
- Voraussage des Spendeblutbedarfs
- Prozessoptimierung

**SCANDAT database** → Schweden und Dänemark, eingerichtet 2002, beinhaltet alle Spender und Transfusionsdaten seit 1968 (Schweden) and 1980 (Dänemark). Beinhaltet 47 Jahre Daten bezüglich der Heilung, Krankenhausaufenthalte, Krebs und Tod

**Recipient Epidemiology and Donor Evaluation Study (REDS-III) USA** → ähnliche Database wie SCANDAT

**PROTON database** zur Identifikation (**PROles of TransfusiON recipients**) in den Niederlande, mit Daten bzgl. Alter, Geschlecht, Diagnosen und OPs und die Anzahl erhaltener Spenden pro Aufenthalt .

Besonderheit: nur eine Organisation zur Blutspende → einheitliche Spendedaten, Eingabe der Empfängerdaten über öffentliche Gesundheitsdokumentation (electronic health records (EHRs)), welche von den Krankenhäusern weitergeleitet wird.

Ref_Staff_Categories	
<b>PK</b>	staff_category_code
	staff_category_description eg Doctor, Nurse

Addresses	
<b>PK</b>	address_id
	line_1_number_building line_2_number_street line_3_area_locality city zip_postcode state_province_county country other_address_details

Donors_Medications	
<b>PF</b>	donor_id
<b>PF</b>	medication_code
	comments

Medications	
<b>PK</b>	medication_code
	medication_name medication_description other_details eg Antibiotics

Staff	
<b>PK</b>	staff_id
<b>FK</b>	address_id
<b>FK</b>	blood_bank_id
<b>FK</b>	staff_category_code
	gender_mfu staff_job_title staff_first_name staff_middle_name staff_last_name staff_qualifications staff_birth_date other_staff_details

Blood_Banks	
<b>PK</b>	blood_bank_id
<b>FK</b>	address_id
	blood_bank_details

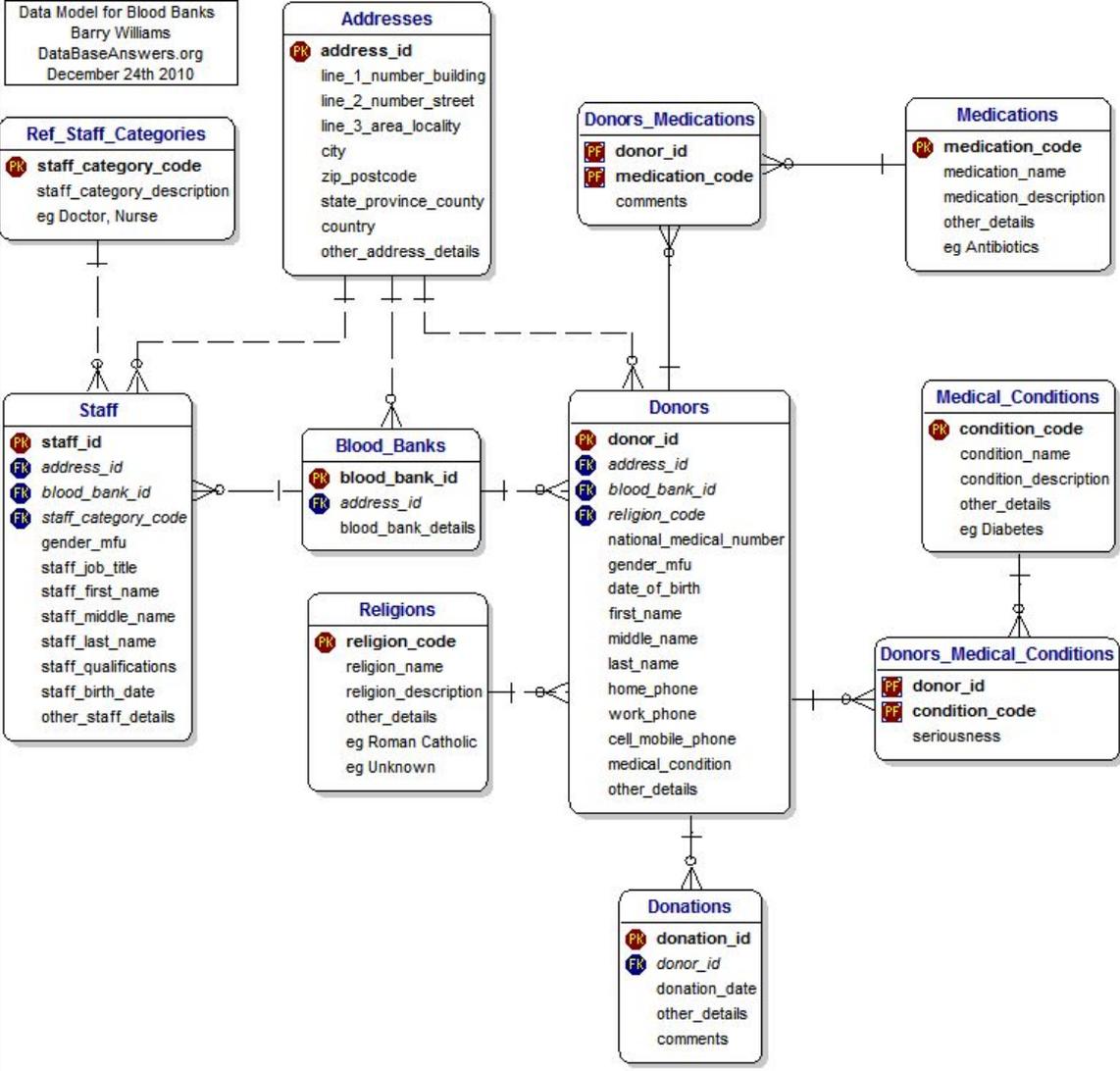
Donors	
<b>PK</b>	donor_id
<b>FK</b>	address_id
<b>FK</b>	blood_bank_id
<b>FK</b>	religion_code
	national_medical_number gender_mfu date_of_birth first_name middle_name last_name home_phone work_phone cell_mobile_phone medical_condition other_details

Medical_Conditions	
<b>PK</b>	condition_code
	condition_name condition_description other_details eg Diabetes

Religions	
<b>PK</b>	religion_code
	religion_name religion_description other_details eg Roman Catholic eg Unknown

Donors_Medical_Conditions	
<b>PF</b>	donor_id
<b>PF</b>	condition_code
	seriousness

Donations	
<b>PK</b>	donation_id
<b>FK</b>	donor_id
	donation_date other_details comments



## Data Warehousing vs. Big Data

- Big Data: breit verwendeter Begriff für extrem “große”, heterogene Datenmengen (z.B. aus sozialen Netzwerken), die mit herkömmlichen Methoden nicht mehr zu erfassen und zu verwerten sind
- Big Data Analytics: Methoden zur Analyse von Big Data

### **Data Warehousing:**

- Analyse in Intervallen
- Daten sind heterogen, aber strukturiert
- begrenzter Fokus, historische Perspektive, Langzeitspeicherung

### **Big Data Analytics:**

- Analyse in Echtzeit
- Daten teilweise unstrukturiert (Bilder, Texte, etc.)
- weiter Fokus, eher kurzzeitige Perspektive

# Quellen

## Bilder:

<http://www.intelligence.de/news/technologische-umsetzung-von-business-intelligence-systemen.html> (aufgerufen am 26.4.17, 14:00 Uhr)

[https://www.slideshare.net/BI\\_Solutions/oltp-vs-olap-41776499](https://www.slideshare.net/BI_Solutions/oltp-vs-olap-41776499) (aufgerufen am 7.05.17 um 12:00 Uhr)

[https://scvinodkumar.files.wordpress.com/2012/02/blood\\_banks\\_model.gif](https://scvinodkumar.files.wordpress.com/2012/02/blood_banks_model.gif) (aufgerufen am 7.05.17, 8:00 Uhr)

## Literatur:

Mertens, Peter; Griese, Joachim: Integrierte Informationsverarbeitung, Band 2: Planungs- und Kontrollsysteme in der Industrie. 9. Auflage. Wiesbaden: Gabler, 2002.

Mucksch, Harry; Behme, Wolfgang: Das Data Warehouse-Konzept als Basis einer unternehmensweiten Informationslogistik. In: Mucksch, Harry; Behme, Wolfgang (Hrsg.): Das Data Warehouse-Konzept. 4. Auflage. Wiesbaden: Gabler 2000, S. 3 – 80.

## Web

<http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Business-Intelligence/Data-Warehouse>  
(aufgerufen am 6.05.17 18:00 Uhr)

<https://www.imise.uni-leipzig.de/Archiv/2003/10/10/DiplomarbeitWagner.pdf> (aufgerufen am 6.05.17 18:00 Uhr)

[https://www.fbi.h-da.de/fileadmin/personal/b.humm/Publikationen/Humm\\_Wietek\\_-\\_Architektur\\_DW\\_Informatik-Spektrum\\_2005-01\\_.pdf](https://www.fbi.h-da.de/fileadmin/personal/b.humm/Publikationen/Humm_Wietek_-_Architektur_DW_Informatik-Spektrum_2005-01_.pdf) (aufgerufen am 6.05.17 18:00 Uhr)

<https://www.aerzteblatt.de/archiv/17265/Data-Mining-und-Data-Warehouse-Wissen-aus-medizinischen-Datenbanken-nutzen> (aufgerufen am 5.05.17, 20:00 Uhr)

<http://bmjopen.bmj.com/content/6/8/e010962> (aufgerufen am 6.05.17, 19:00 Uhr)

# Diskussion

Wer sollte die Autorität über die gesammelten Daten besitzen?

Wie effektiv ist Anonymisierung?

Wie lange ist Speicherung historischer Daten möglich und verantwortungsvoll?

Ist Datawarehousing noch zeitgemäß?