

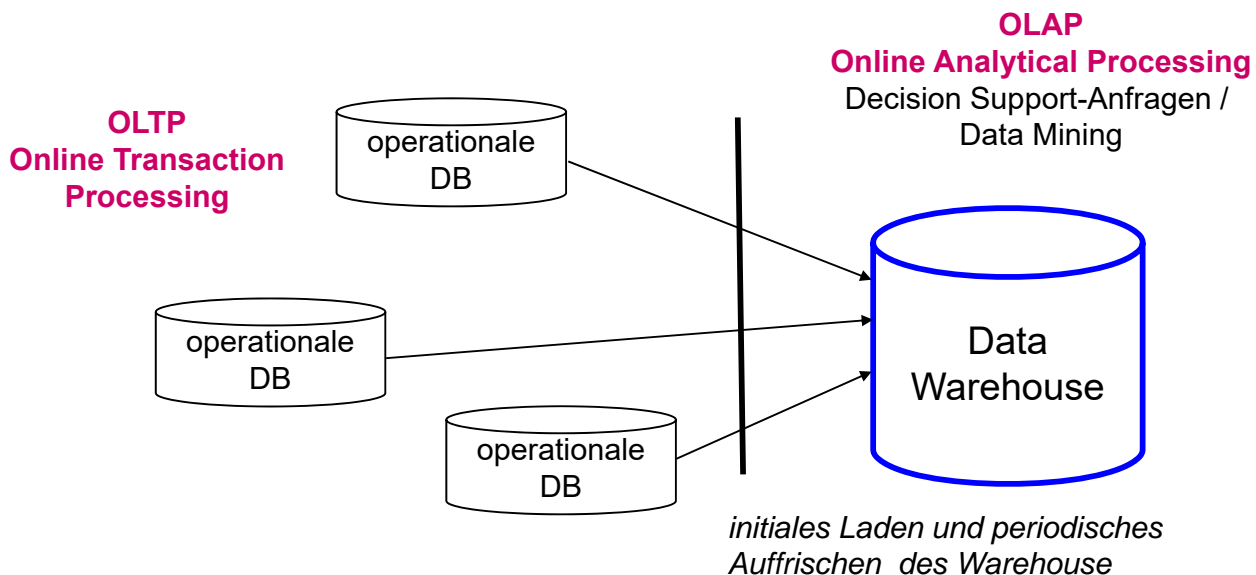
# 1. Data Warehouses - Einführung

- Definitionen und Merkmale
  - Grobdefinition
  - Einsatzbeispiele
  - DW-Merkmale nach Imnon
  - OLTP vs. OLAP
  - Grobarchitektur
  - virtuelle vs. physische Datenintegration
- Mehrdimensionale Datensicht
  - Stern-Schema und -Anfragen
- Analysearten (OLAP, Data Mining)
- Big Data

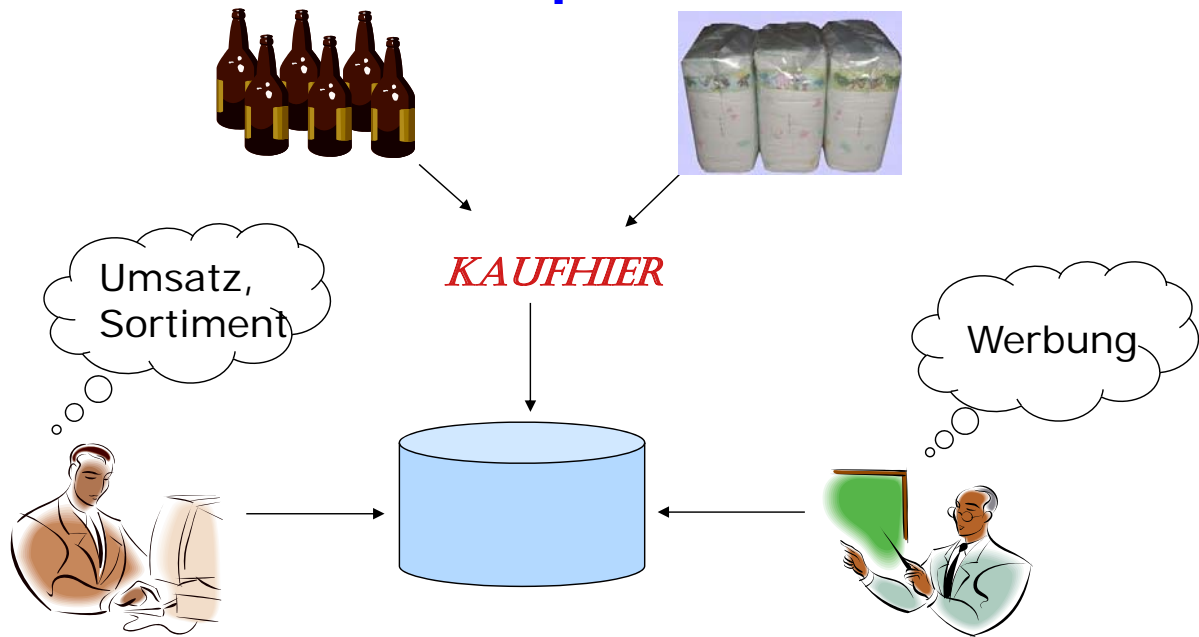


## Data Warehouses

- Ausgangsproblem
  - viele Unternehmen haben Unmengen an Daten, ohne daraus ausreichend Informationen und Wissen für kritische Entscheidungsaufgaben ableiten zu können
- **Data Warehouse (Def.):** für Analysezwecke optimierte zentrale Datenbank, die Daten aus mehreren, i.a. heterogenen Quellen zusammenführt und verdichtet (Integration und Transformation)



# Szenario: Supermarktkette



## ■ Anfragen:

- Wie viele Pakete Windeln wurden letzten Monat verkauft?
- Wie hat sich der Verkauf von Bier und Wasser im letzten Jahr entwickelt?
- Wo sind unsere Top-Filialen?
- Von welchem Lieferanten beziehen wir das meiste Bier?
- Wie wirkten sich die Werbepreise für Produkt X aus? ...



# Einsatzbeispiele

## ■ Warenhauskette

- Verkaufszahlen und Lagerbestände aller Warenhäuser
- mehrdimensionale Analysen: Verkaufszahlen nach Produkten, Regionen, Warenhäusern
- Ermittlung von Kassenschlagern und Ladenhütern
- Analyse des Kaufverhaltens von Kunden (Warenkorbanalyse)
- Erfolgskontrolle von Marketing-Aktivitäten
- Minimierung von Beständen
- Optimierung der Produktpalette, Preisgestaltung •••

## ■ Versicherungsunternehmen

- Bewertung von Filialen, Vertriebsbereichen, Schadensverlauf, ...
- automatische Risikoanalyse
- schnellere Entscheidung über Kreditkarten, Lebensversicherung, Krankenversicherung ...

## ■ Banken, Versandhäuser, Restaurant-Ketten

## ■ wissenschaftliche Einsatzfälle •••

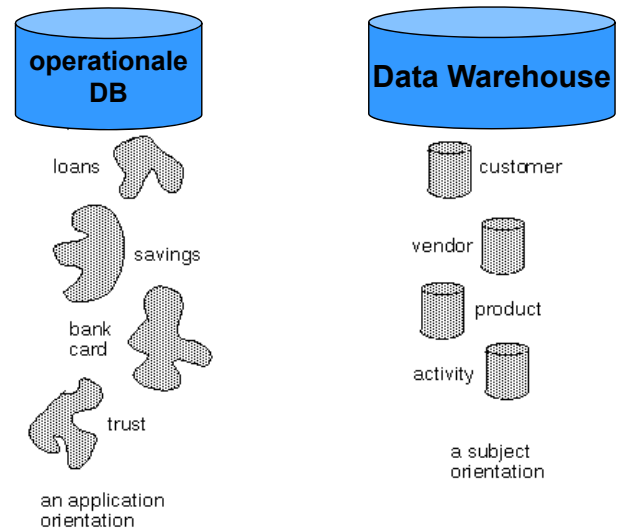


# DW-Eigenschaften nach Inmon

A Data Warehouse is a *subject-oriented, integrated, non-volatile, and time variant* collection of data in support of management decisions (W. H. Inmon, *Building the Data Warehouse*, 1996)

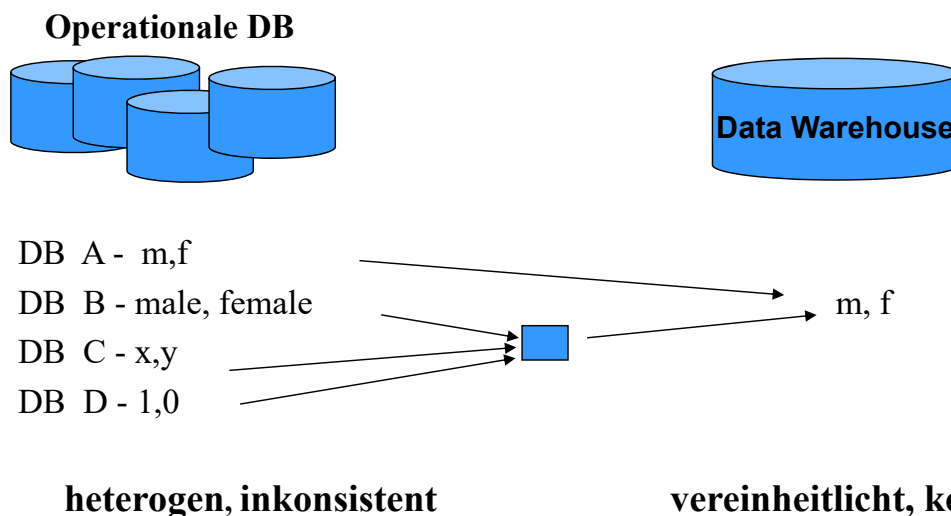
## ■ Subjekt-orientiert

- Zweck des Systems ist nicht Erfüllung einer dedizierten operationalen Aufgabe (z.B. Personaldatenverwaltung),
- alle Daten - unternehmensweit - über ein Subjekt (Kunden, Produkte, Regionen ...) und nicht „versteckt“ in verschiedenen Anwendungen
- Unterstützung übergreifender Auswertungsmöglichkeiten aus verschiedenen Perspektiven



## DW-Eigenschaften nach Inmon (2)

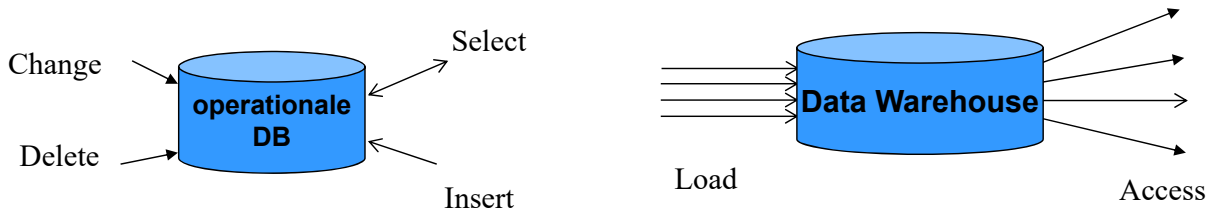
- integrierte Datenbasis (integrated): konsolidierte Daten aus mehreren verschiedenen Datenquellen



## DW-Eigenschaften nach Inmon (3)

### ■ dauerhafte Datenbasis (non-volatile):

- Daten im DW werden i.a. nicht mehr geändert
- stabile, persistente Datenbasis



regelmäßige Änderungen von Sätzen



## DW-Eigenschaften nach Inmon (4)

### ■ historische Daten (time-variant):

- Vergleich der Daten über Zeit möglich (Zeitreihenanalyse)
- Speicherung über längeren Zeitraum



aktuelle Datenwerte:

- Zeitbezug optional
- Zeithorizont: 60-90 Tage
- Daten änderbar



Schnappschuß-Daten

- Zeitbezug aller Objekte
- Zeithorizont: 2-10 Jahre
- keine Änderung nach Schnappschuß-Erstellung



# Operationale Datenbanken vs. Data Warehouses (OLTP vs. OLAP)

	Operationale Datenbanken /OLTP	Data Warehouses/OLAP
<i>Entstehung</i>	für je eine Applikation / eine Perspektive	mehrere Perspektiven / anwendungs- übergreifend
<i>Bedeutung</i>	Tagesgeschäft	Entscheidungs-/Planungsaufgaben
<i>Nutzer</i>	Sachbearbeiter, Online-Nutzer	Analysten / Manager
<i>Datenzugriff</i>	sehr häufiger Zugriff, kleine Datenmengen pro Operation, Lesen, Schreiben, Modifizieren, Löschen	moderate Zugriffsfrequenz, große Datenmengen, vorwiegend lesender Zugriff
<i>Änderungen</i>	sehr häufig	periodisches Auffrischen
<i>#Datenquellen</i>	meist eine pro Anwendung	mehrere
<i>Datenmerkmale</i>	nicht abgeleitet, autonom, zeitaktuell, dynamisch	abgeleitet, integriert, i.a. leicht veraltet, stabil
<i>Optimierungsziele</i>	hoher Durchsatz, sehr kurze Antwortzeiten (ms .. s), hohe Verfügbarkeit	gute Antwortzeiten für komplexe Analysen

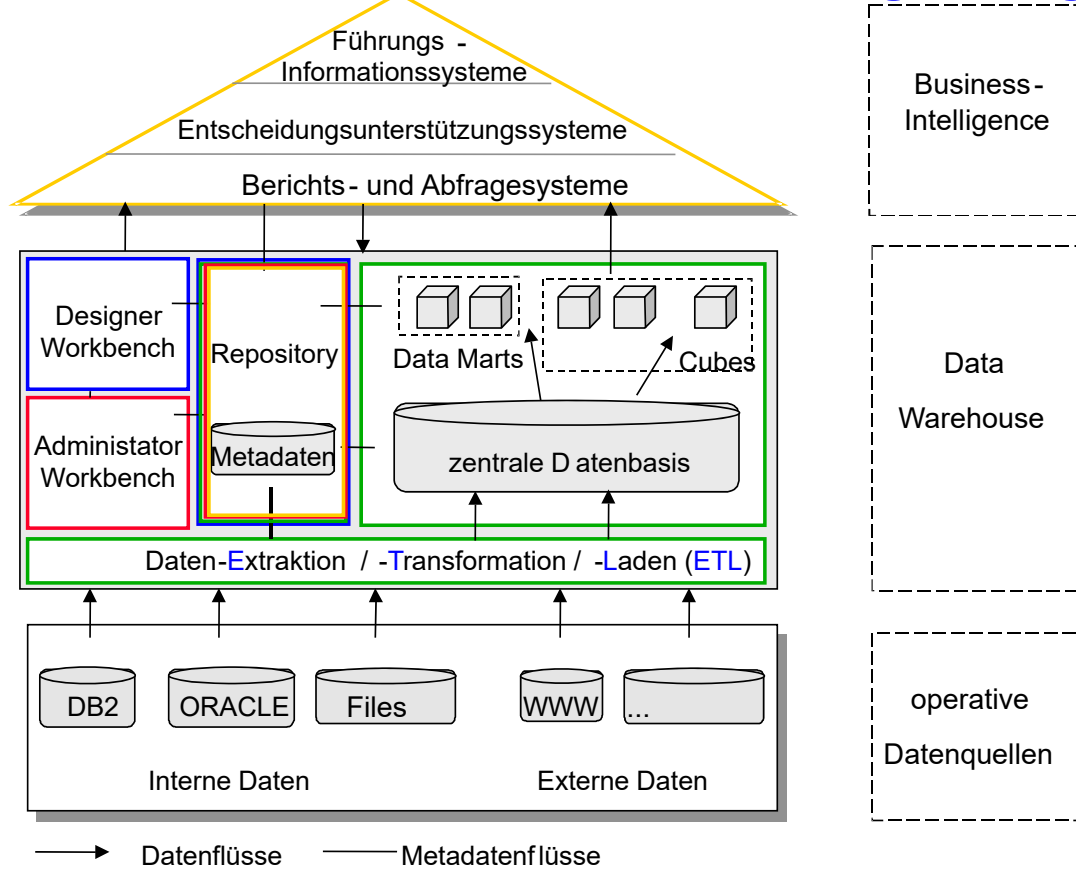


## Warum separates Data Warehouse?

- unterschiedliche Nutzung und Datenstrukturierung
- unterschiedliche Funktionalität
  - historische Daten
  - Konsolidierung (Integration, Bereinigung und Aggregation) von Daten aus heterogenen Datenquellen
- Performance
  - OLTP optimiert für kurze Transaktionen und bekannte Lastprofile
  - komplexe OLAP-Anfragen würden gleichzeitige OLTP-Transaktionen ausbremsen
  - OLAP erfordert speziellen logischen / physischen DB-Entwurf für mehrdimensionale Anfragen
  - Transaktionseigenschaften (ACID) für OLAP weniger wichtig
- Sicherheit
- Nachteile der separaten Lösung
  - Datenredundanz
  - Daten nicht vollständig aktuell
  - hoher Administrationsaufwand
  - hohe Kosten

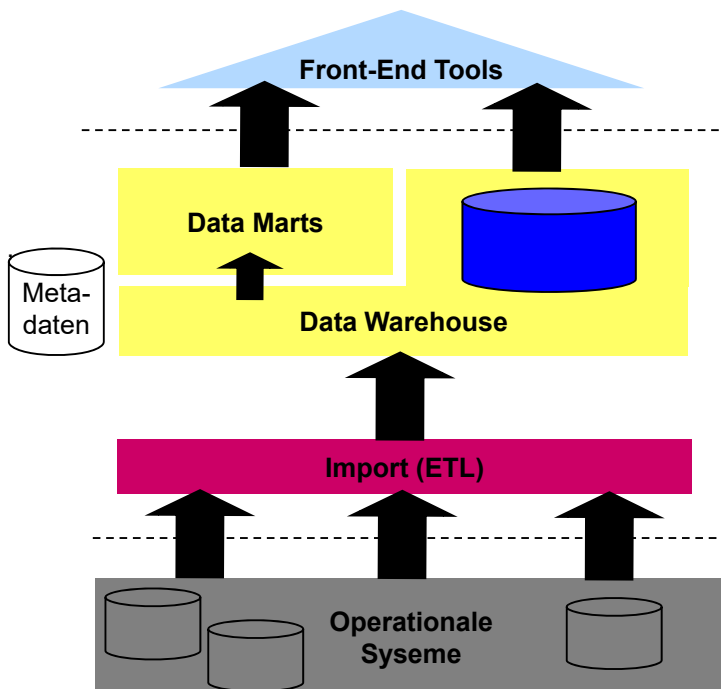


# Grobarchitektur einer DW-Umgebung

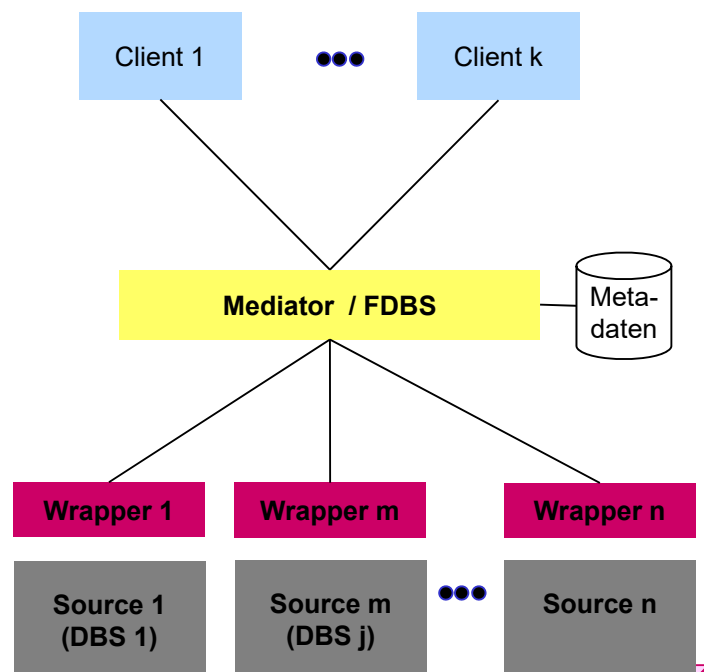


# Datenintegration: physisch vs. virtuell

## Physische (Vor-) Integration (Data Warehousing)



## Virtuelle Integration (Mediator/Wrapper-Architekturen, föderierte DBS)



## Datenintegration: physisch vs. virtuell (2)

	physisch (Data Warehouse)	virtuell
Integrationszeitpunkt: Metadaten	vorab (DW-Schema)	vorab (globale Sicht)
Integrationszeitpunkt: Daten	vorab	dynamisch (zur Anfragezeit)
Aktualität der Daten		
Autonomie der Datenquellen		
Erreichbare Datenqualität		
Analysezeitbedarf für große Datenmengen		
Hardwareaufwand		
Skalierbarkeit auf viele Datenquellen		

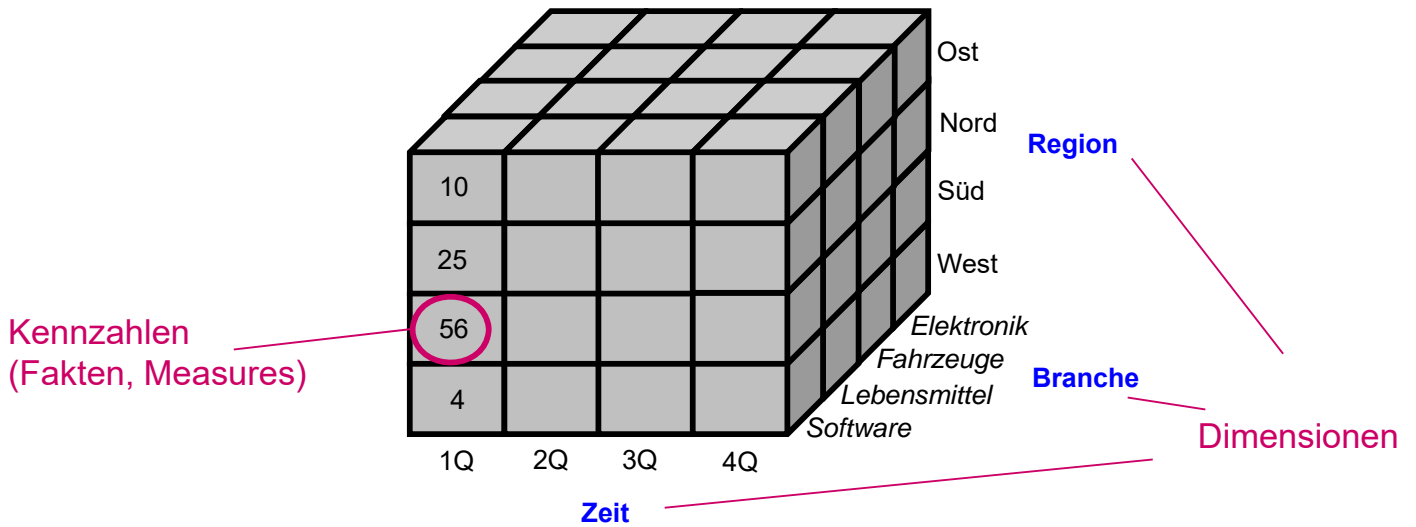


## OLAP (Online Analytical Processing)

- interaktive multidimensionale Analyse auf konsolidierten Unternehmensdaten
- FASMI-Merkmale (Fast Analysis of Shared Multidimensional Information)
  - Skalierbarkeit auf große Datenmengen
    - stabile, volumenunabhängige Antwortzeiten
  - intuitive, interaktive Analyse und Visualisierung
  - Mehrbenutzerunterstützung
    - Client/Server-Architektur
  - mehrdimensionale, konzeptionelle Sicht auf die Daten
    - unbegrenzte Anzahl an Dimensionen und Aggregationsebenen
    - unbeschränkte dimensionsübergreifende Operationen
  - integrierter Zugang zu heterogenen Datenbeständen mit logischer Gesamtsicht



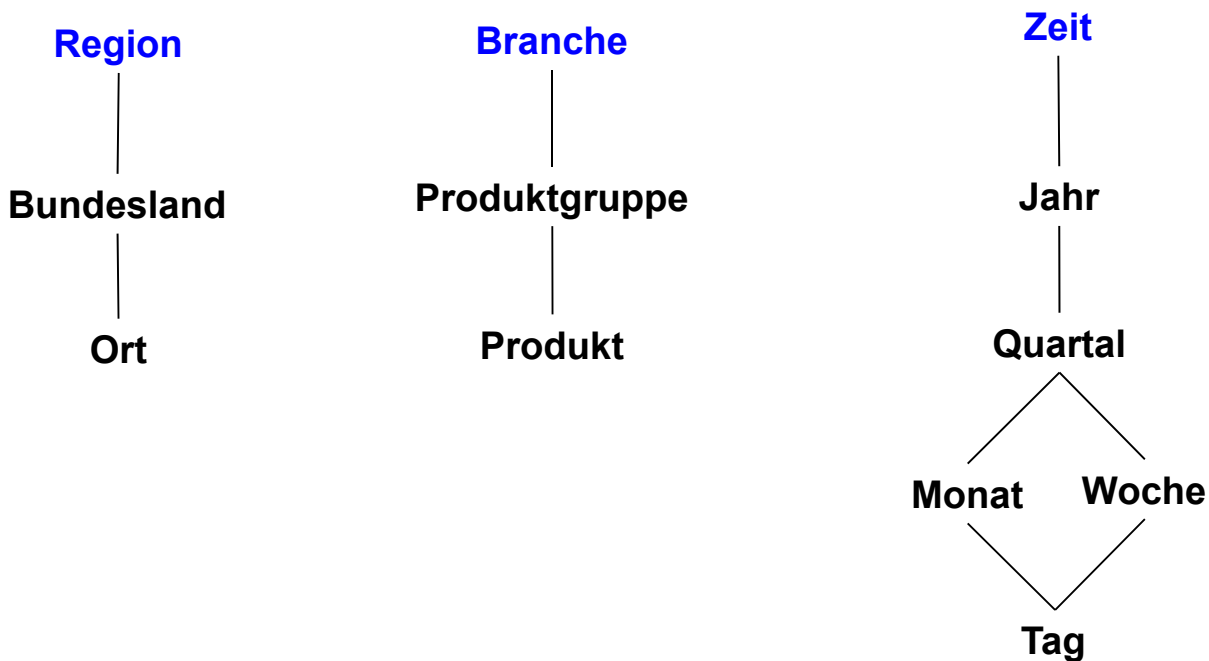
# Mehrdimensionale Datensicht



- Kennzahlen: numerische Werte als Grundlage für Aggregationen/Berechnungen (z.B. Absatzzahlen, Umsatz, etc.)
- Dimensionen: beschreibende Eigenschaften
- Operationen:
  - Aggregation der Kennzahlen über eine oder mehrere Dimension(en)
  - Slicing and Dicing: Bereichseinschränkungen auf Dimensionen



# Hierarchische Dimensionierung



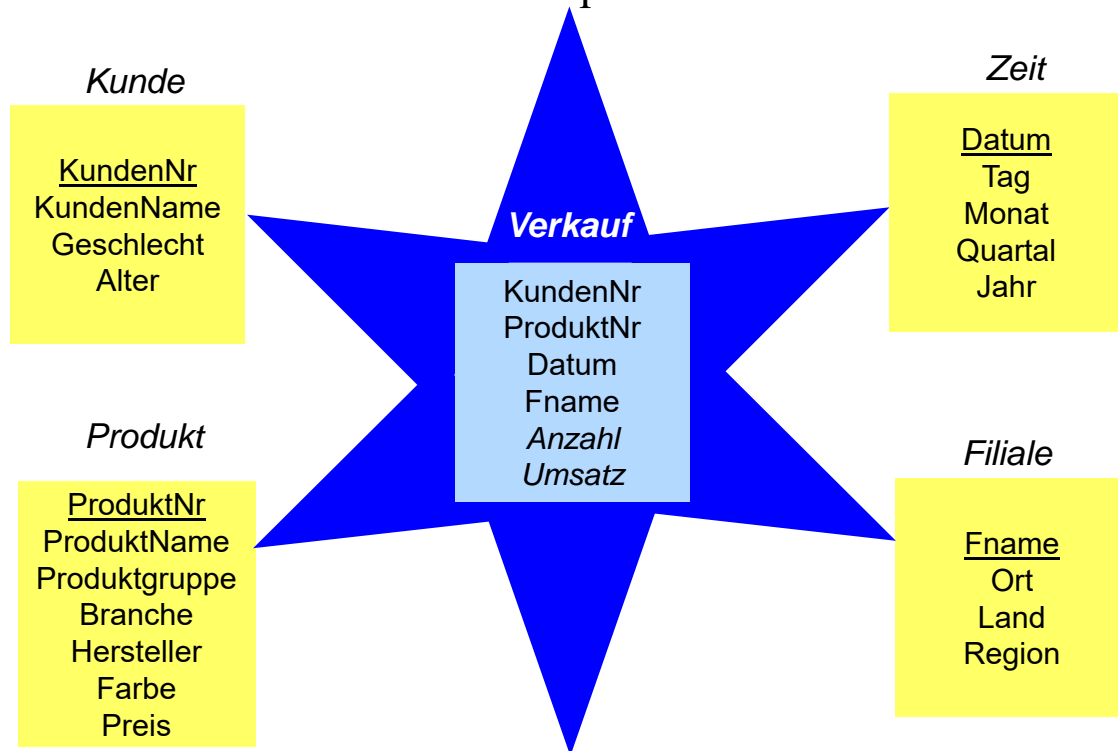
- Operationen zum Wechsel der Dimensionsebenen
  - Drill-Down
  - Roll-Up





# Star-Schema

- zentrale **Faktentabelle** sowie 1 Tabelle pro Dimension



## Anfragen

### Beispielanfrage:

*Welche Auto-Hersteller wurden von weiblichen Kunden in Sachsen im 1. Quartal favorisiert?*

```
select p.Hersteller, sum (v.Anzahl)
from Verkauf v, Filialen f, Produkt p, Zeit z, Kunden k
where z.Quartal = 1 and k.Geschlecht = 'W' and
      p.Produkttyp = 'Auto' and f.Land = 'Sachsen' and
      v.Datum = z.Datum and v.ProduktNr = p.ProduktNr and
      v.Filiale = f.FName and v.KundenNr = k.KundenNr
group by p.Hersteller
order by 2 desc;
```

### ■ Star-Join

- sternförmiger Join der (relevanten) Dimensionstabellen mit der Faktentabelle
- Einschränkung der Dimensionen
- Verdichtung der Kennzahlen durch Gruppierung und Aggregation



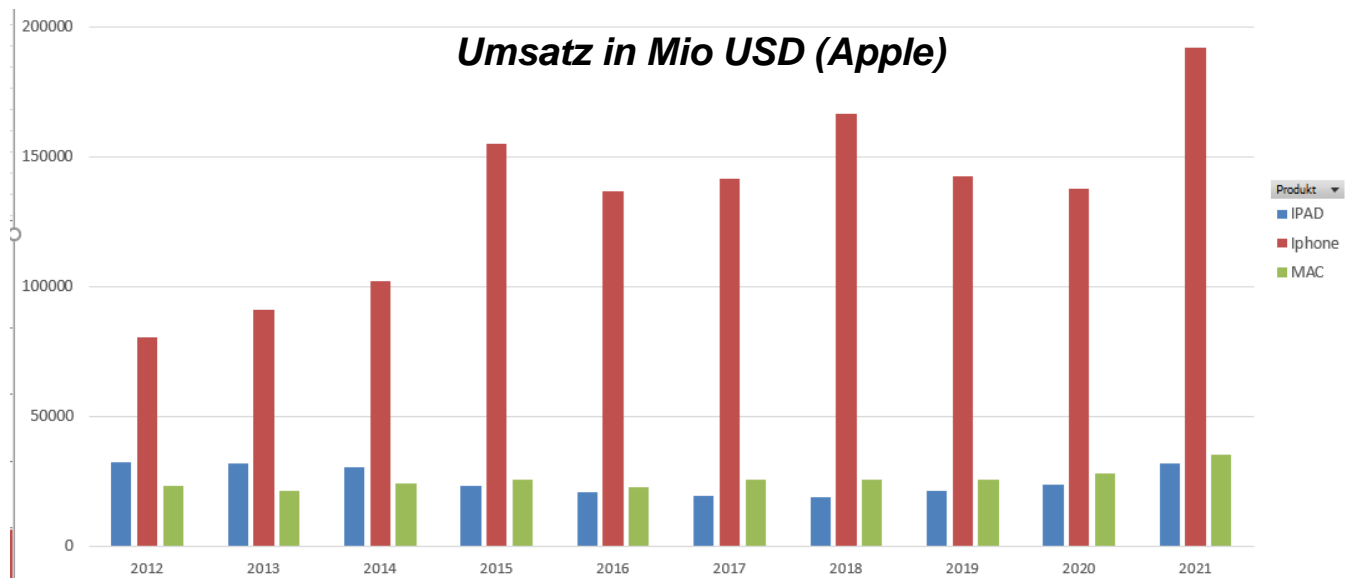
# Analysewerkzeuge

- (Ad Hoc-) Query-Tools
- Reporting-Werkzeuge, Berichte mit flexiblen Formatierungsmöglichkeiten
- OLAP-Tools
  - OLAP-Unterstützung in Spreadsheet-Tools bzw. im Web-Browser
  - oft Datendarstellung als Pivot-Tabellen (Kreuztabellen)
  - interaktive mehrdimensionale Analyse und Navigation (Drill Down, Roll Up, ...)
  - Gruppierungen, statistische Berechnungen,
  - unterschiedlichste Visualisierungen
- Tools/Verfahren für Data Mining und maschinelles Lernen

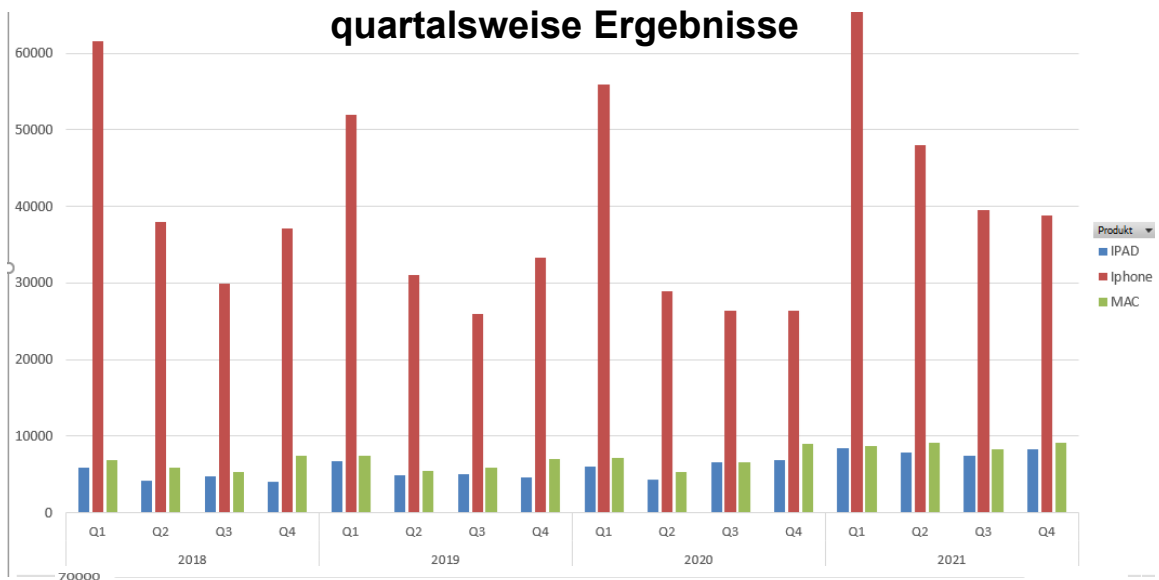


## Beispiel: OLAP-Ausgabe (Excel)

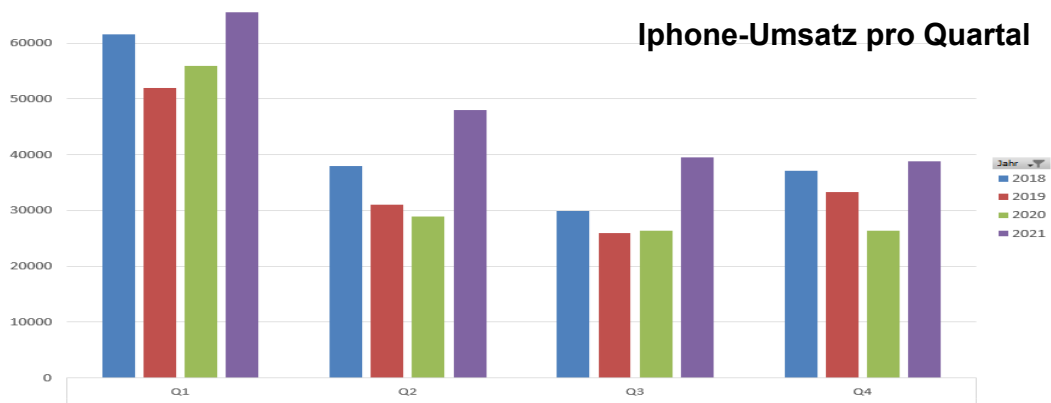
Summe von Umsatz Mio\$	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Gesamtergebnis
<b>IPAD</b>	32.424	31.980	30.283	23.227	20.628	19.222	18.805	21.280	23.719	31.862	<b>253.430</b>
<b>Iphone</b>	80.477	91.279	101.991	155.041	136.700	141.319	166.699	142.381	137.781	191.973	<b>1.345.641</b>
<b>MAC</b>	23.221	21.483	24.079	25.471	22.831	25.850	25.484	25.740	28.127	35.190	<b>257.476</b>
<b>Gesamtergebnis</b>	<b>136.122</b>	<b>144.742</b>	<b>156.353</b>	<b>203.739</b>	<b>180.159</b>	<b>186.391</b>	<b>210.988</b>	<b>189.401</b>	<b>189.627</b>	<b>259.025</b>	<b>1.856.547</b>



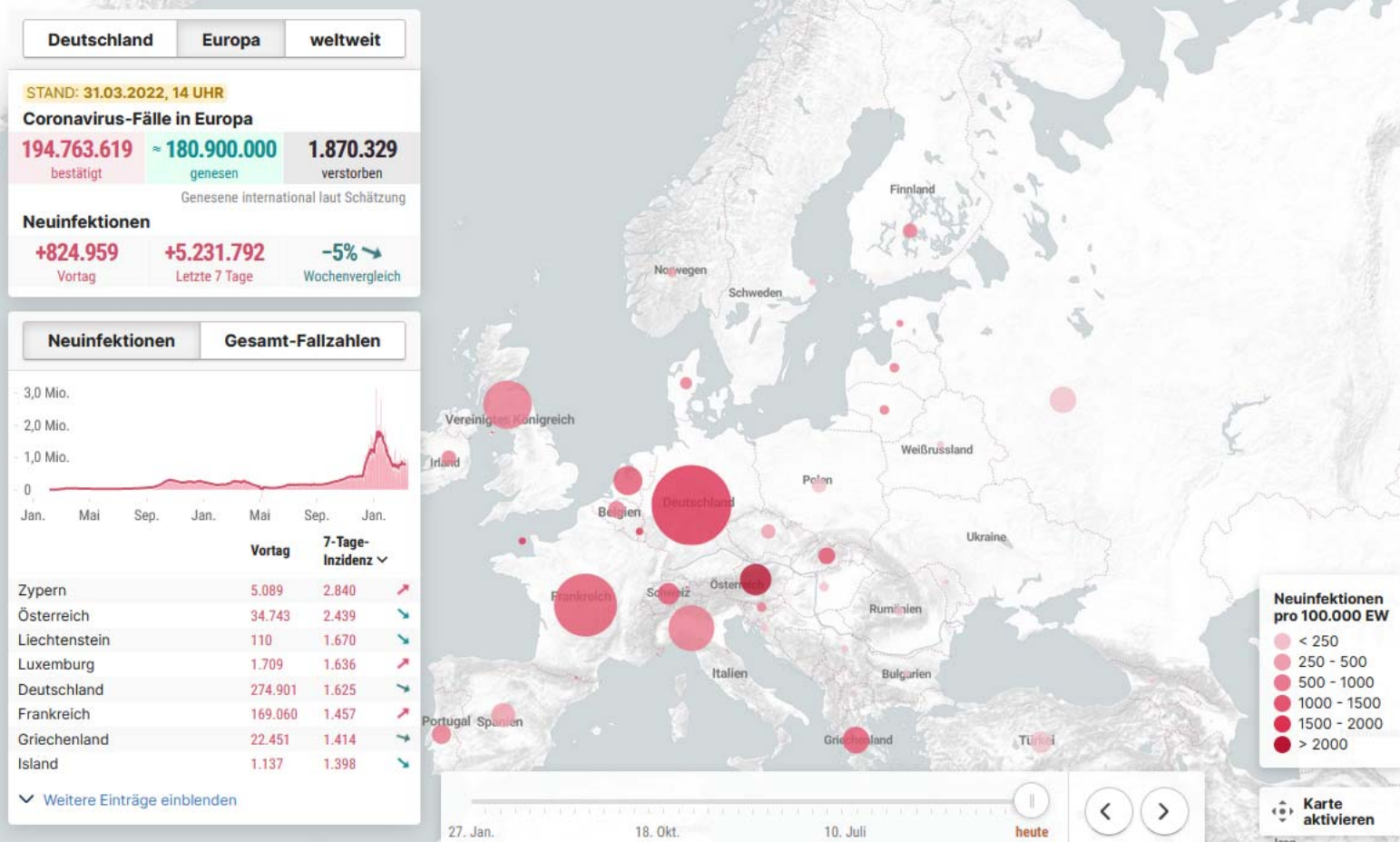
## quartalsweise Ergebnisse



## Iphone-Umsatz pro Quartal

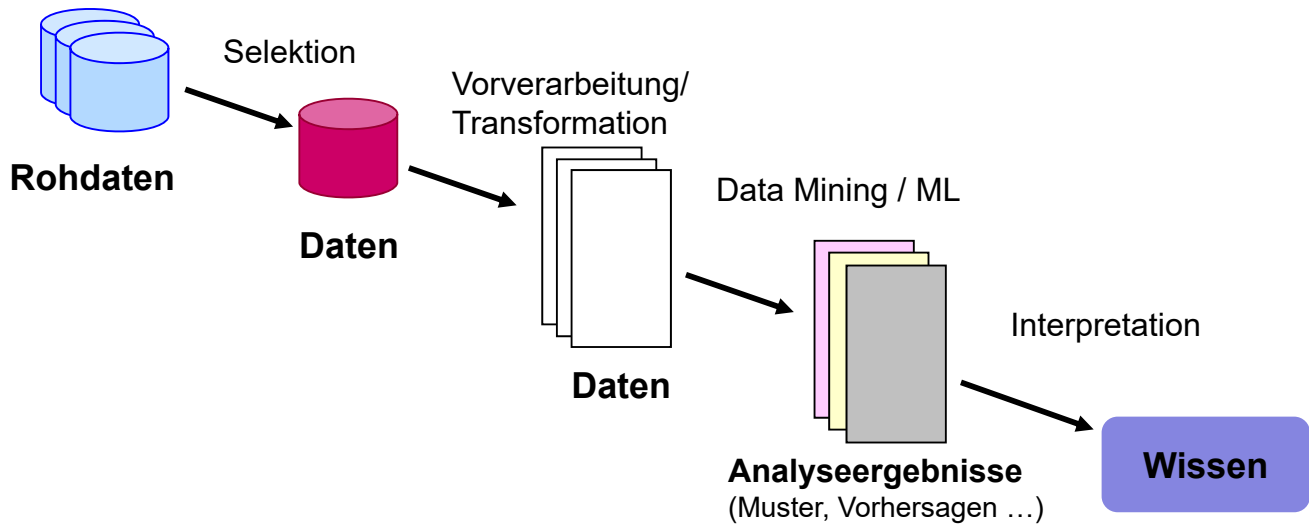


## Ergebnisvisualisierung (Corona-Fälle)



# Knowledge Discovery

- (semi-) automatische Extraktion von Wissen aus Daten
- Kombination von Verfahren zu Datenbanken, Statistik (Data Mining) und KI (maschinelles Lernen)

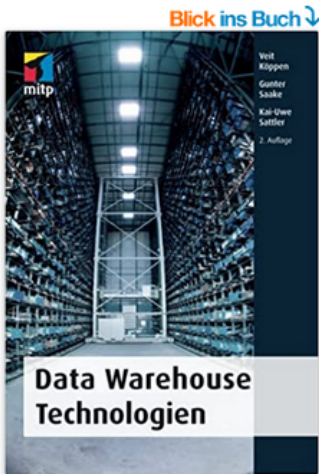


## Techniken des Data Mining/ML

- Einsatz statistischer, wissens- und lernbasierter Methoden zur Datenanalyse
  - Auffinden von Korrelationen, Mustern und Trends in Daten, Vorhersagen
- Clusteranalyse
  - Objekte werden aufgrund von Ähnlichkeiten in Klassen eingeteilt (Segmentierung)
  - Bsp.: ähnliche Kunden, ähnliche Website-Nutzer ...
- Assoziationsregeln
  - Warenkorbanalyse (z.B. Kunde kauft A und B => Kunde kauft C)
  - Nutzung für Kaufvorhersagen / Recommendations, ...
- Klassifikation
  - Zuordnung von Objekten zu Gruppen/Klassen mit gemeinsamen Eigenschaften bzw. Vorhersage von Attributwerten
  - Verwendung von Stichproben (Trainingsdaten)
  - Ansätze: Entscheidungsbaum-Verfahren, neuronale Netze, statistische Auswertungen







## Data Warehouse Technologien (mitp Professional)

Taschenbuch – 26. Mai 2014

von Veith Köppen (Autor), Kai-Uwe Sattler (Autor), Gunter Saake (Autor)

★★★★☆ 2 Sternebewertungen

Alle Formate und Editionen anzeigen

Kindle  
23,99 €

Lesen Sie mit unserer **kostenfreien App**

Taschenbuch  
29,99 €

6 Gebraucht ab 21,89 €  
21 Neu ab 29,99 €

- Architekturprinzipien von Data-Warehouse-Systemen
- Datenstrukturen und Algorithmen
- Anwendungsfeld Business Intelligence

Dieses Lehrbuch behandelt Konzepte und Techniken von Data-Warehouse-Systemen, die eine wesentliche Komponente in betrieblichen Entscheidungsprozessen darstellen. Im Mittelpunkt stehen dabei Architekturprinzipien sowie die Umsetzung des multidimensionalen Datenwürfels als zentrale

Kunden, die diesen Artikel angesehen haben, haben auch angesehen



Data-Warehouse-Systeme für Dummies  
Wolfgang Gerken  
★★★★☆ 17  
Taschenbuch  
26,99 €  
Erhalten Sie es bis Montag.



Data-Warehouse-Systeme: Architektur, Entwicklung, ...  
Dr.-Ing. Andreas Bauer  
★★★★☆ 18  
Gebundene Ausgabe  
49,90 €



Data Warehouse Blueprints: Business Intelligence in der Praxis  
Dani Schneider  
★★★★☆ 6  
Gebundene Ausgabe  
42,00 €



Modellieren Business-Intelligence-Systemen  
Michael Har  
★★★★☆ 6  
Gebundene Ausgabe  
69,90 €

### Verwandte Produkte zu diesem Artikel

Gesponsert @

Neuerscheinung

**Hands-On Azure Data Platform**  
Building Scalable Enterprise-Grade Relational and...  
Sagar Lad  
★★★★☆ 1  
Taschenbuch  
28,88 € vprime

Neuerscheinung

**Mastering Tableau 2021**  
Implement advanced business intelligence techniques and ana...  
Marleen Meier  
★★★★☆ 23  
Taschenbuch  
40,64 € vprime

Neuerscheinung

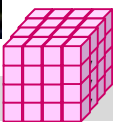
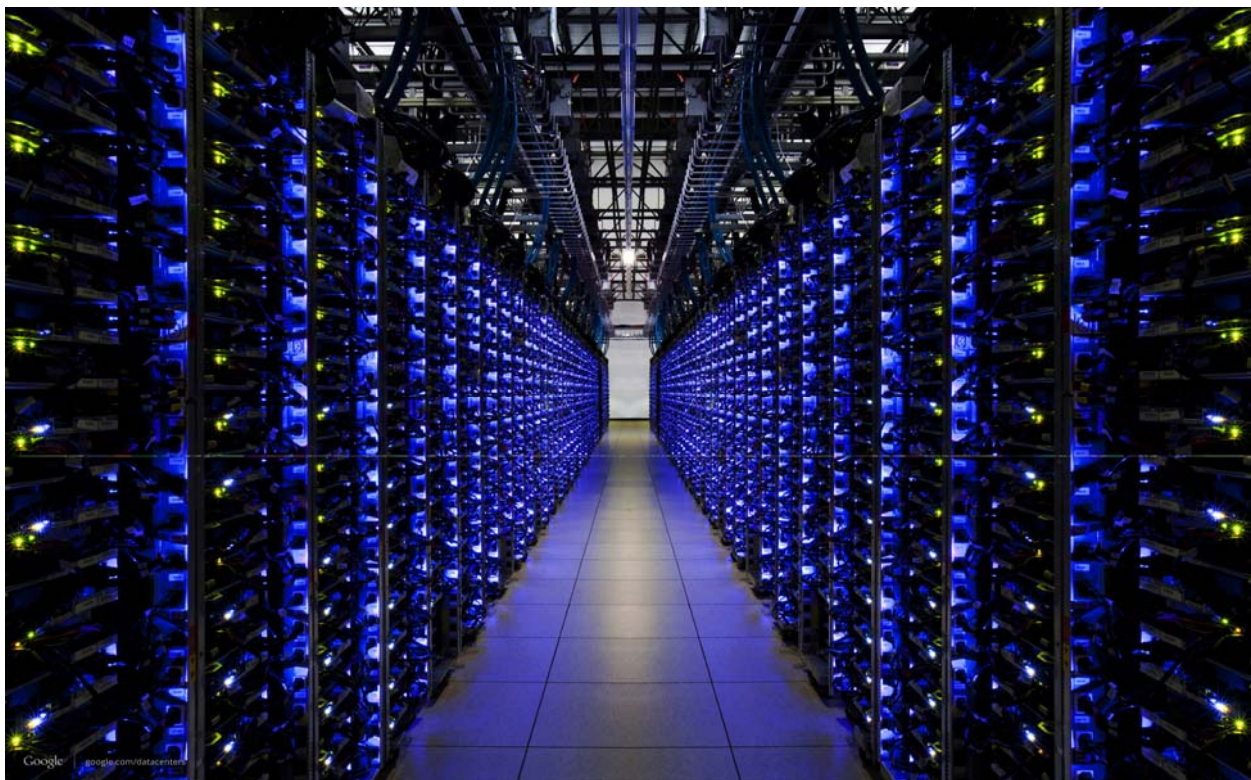
**Business Analytics with SAS Studio**  
Deliver Business Intelligence by Combining SQL ...  
Rajinder Kr. Chitoria  
Taschenbuch  
28,88 € vprime

**BUSINESS INTELLIGENCE GUIDEBOOK**  
From Data Integration to Analytics (English Edition)  
Rick Sherman  
★★★★☆ 65  
Kindle-Ausgabe  
32,44 €

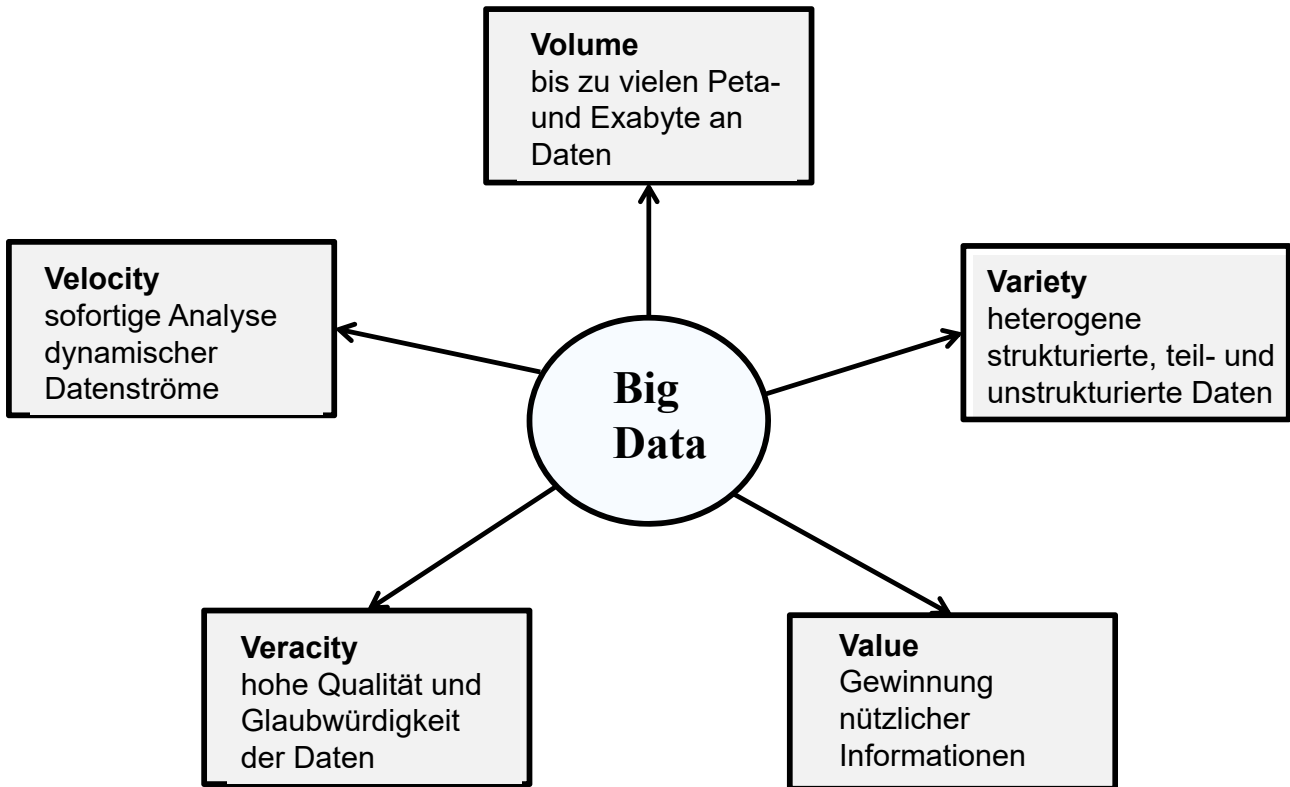
**Mastering Microsoft Power BI**  
Expert techniques for effective data analytics and bu...  
Brett Powell  
★★★★☆ 118  
Taschenbuch  
44,92 € vprime



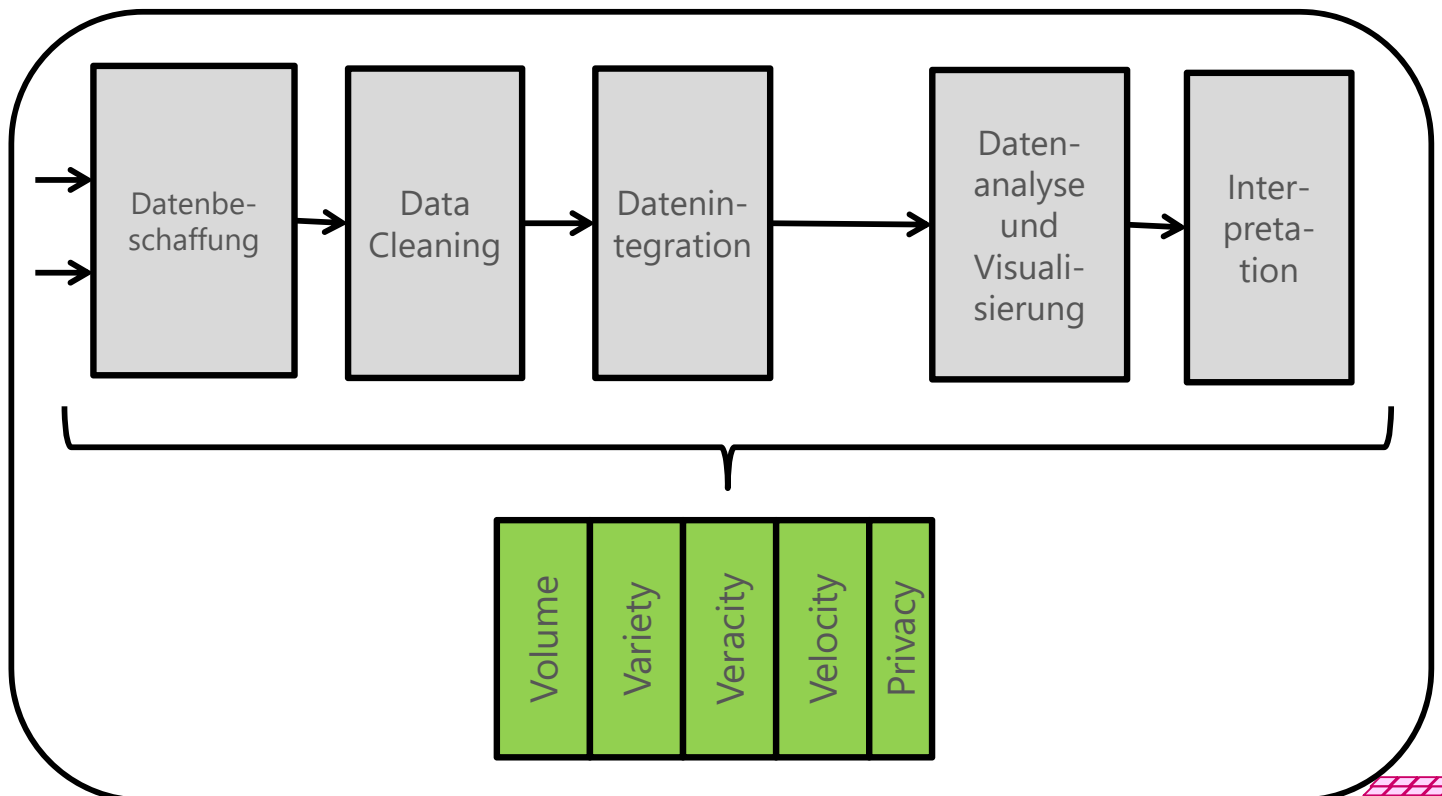
# Big Data



# Anforderungen für „Big Data“



# Big Data Analyse-Pipeline



# Zusammenfassung

- Data Warehousing: DB-Anfrageverarbeitung und Analysen auf integriertem Datenbestand für Decision Support (OLAP)
- riesige Datenvolumina
- Hauptschwierigkeit: Integration heterogener Datenbestände sowie Bereinigung von Primärdaten
- physische Datenintegration ermöglicht
  - aufwändige Datenbereinigung
  - effiziente Analyse auf großen Datenmengen
- mehrdimensionale Datenmodellierung und -organisation
- breites Spektrum an Auswertungs- und Analysemöglichkeiten
- Data Mining: selbständiges Aufspüren relevanter Muster in Daten
- Big Data: Datenanalysen auf großen Mengen auch unstrukturierter Daten

