
Datenbanken, Data Warehouse **Data Mining**

Bernd Blümel

**Ein Teil der Folien mit
freundlicher Genehmigung von
Hermann Johannes, FH-SWF**

1. Datenbanken

- Einleitung**
- Entity–Relationship–Modell**
- Normalisierung**
- Tabellenstruktur**

2. Data Warehouse

3. Data Mining

Einleitung

- **Problem: Treffen von Entscheidungen**

"I know half the money I spend on advertising is wasted, but I can never find out which half."

- **große Datenmengen vorhanden**

"We are drowning from data but starving for information."

- **gesucht: Wissen
bzw. Informationen**

Datenaggregation



Schlechtes Beispiel

Obsthändler Krämers erster Datenbankentwurf

Auftragsnr.	Datum	Kunde	Artikelnr.	Bezeichnung	Menge
1	1.1.99	1 Schmitt, Bonn	134	Coxorange	4 Kisten
1	1.1.99	1 Schmitt, Bonn	135	Kiwi	4 Kisten
2	1.1.99	2 Müller, Köln	140	Butterbime	2 Kisten
2	1.1.99	2 Müller, Köln	160	Kürbis, rot	2 Stück
2	1.1.99	2 Müller, Köln	160	Kürbis, gelb	10 Stück
3	2.2.99	1 Schmitz, Bonn	103	Johannisbeeren	5 Kilo
3	2.2.99	1 Schmitz, Bonn	134	Coxorange	12 Kisten
3	2.2.99	1 Schmitz, Bonn	135	Kiwi	2 Kisten
4	2.2.99	45 Lehmann, Jülich	30	Bananen	12 Kilo
4	2.2.99	45 Lehmann, Jülich	27	Ananas	60 Stück

Nachteile

- **Einfüge–Anomalie**
- **Lösch–Anomalie**
- **Änderungs–Anomalie**

Neu und Besser

Kundentabelle

Primärschlüssel

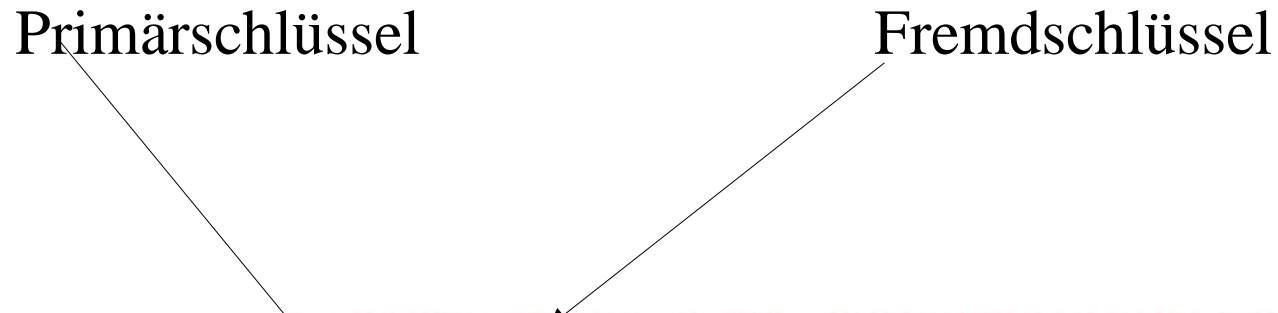
KKURZ	FIRMA	ZUSATZ	STRASSE	LKZ	PLZ	ORT	KTELEFON	KINFO
100	Versicherungspartner		Ringstr. 42	D	28719	Bremen	0421-123456	Versicherung
101	Computer Courier		Heerstr. 101	D	28307	Bremen	0421-452478	PC-Zeitschrift
102	MOTRON AG		Buggenacher 26	CH	6043	Adligenswill	0041-41- 316971	Consulting
103	Eutez Finanz		Luegallee 7	D	40545	Düsseldorf	0211-571618	Consulting
104	Deutsche Unterhaltung		Postfach 448360	D	28283	Bremen	0421-4585	Elektronik
105	Busy Software		BOX 27969	USA		San Diego	CA 9212	PC-Zeitschrift
106	Ringwerke AG	Diagnostica	Postfach 1140	D	35001	Marburg	06421-39- 4492	Chemie
107	Therm AG	Chem. Treatments	Zehntenstr. 54	CH	4133	Pratteln	0041-061- 818100	Chemie
108	Finkelstein		25 Washington St.	USA		Chicago	IL 60602	SQL-Zeitschrift
109	JUFI GmbH		Stolzen Str. 36	D	28207	Bremen	0421- 498114/15	Verpackung
110	Bacher Elektronik AG		Spinnereistr. 10	CH	9008	St. Gallen	0041-71- 263151	Elektronik
111	Alemann GmbH		Kreienweg 41	D	26127	Oldenburg	0441-34789	Elektronik
112	Donakil GmbH	Vertrieb	Karlsdamm 11	D	26123	Oldenburg	0441-43339	Elektronik

neu und besser

Auftragstabelle

Primärschlüssel

Fremdschlüssel



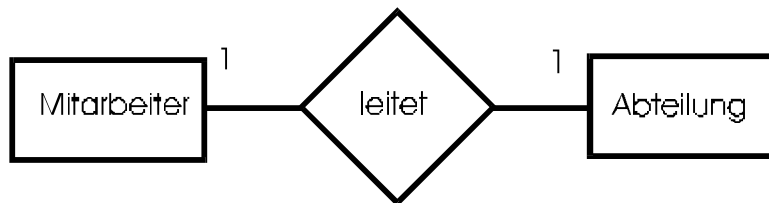
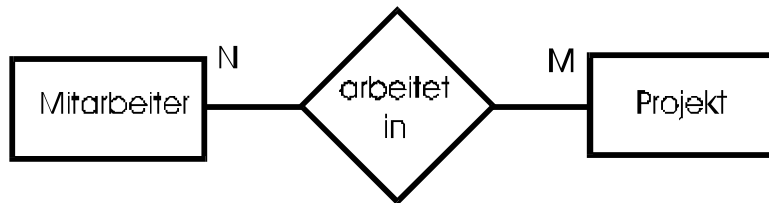
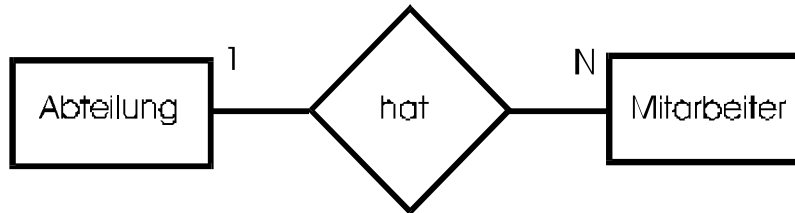
AKURZ	KKURZ	ADATUM	FERTIGIST	FERTIGSOLL	ATEXT
109	109	29.04.96	01.05.96	01.05.96	Serienbrief-Einweis.
111	100	11.05.98		30.11.98	Mitarbeiterschulung
112	110	10.01.96	13.04.97	28.02.97	Offertenverwaltung
113	110	05.02.97	04.02.98	15.12.97	Kunden-Verwaltung
114	101	16.05.98		12.07.98	ORACLE-Rezension
115	106	19.06.97	02.09.97	30.07.97	Datenbank-Entwurf DIAG

Entity-Relationship Modellierung

- **Entities**
 - **Kunden**
 - **Auftrag**
 - **Angebot**
- **Beziehung zwischen Entities**
 - **Kunde fragt an Angebot**
 - **Auftrag resultiert aus Angebot**
- **Attribute**

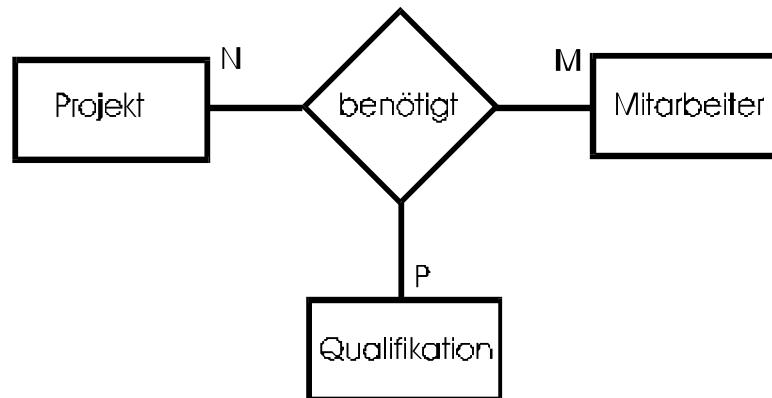
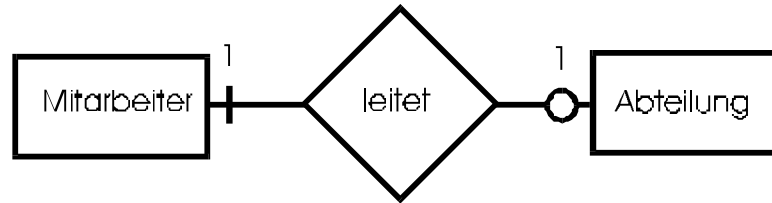
Beziehungen

Kardinalitäten



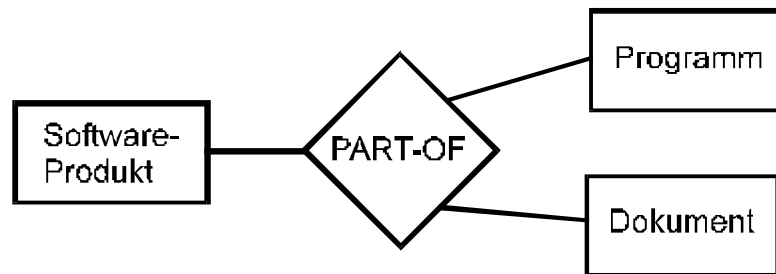
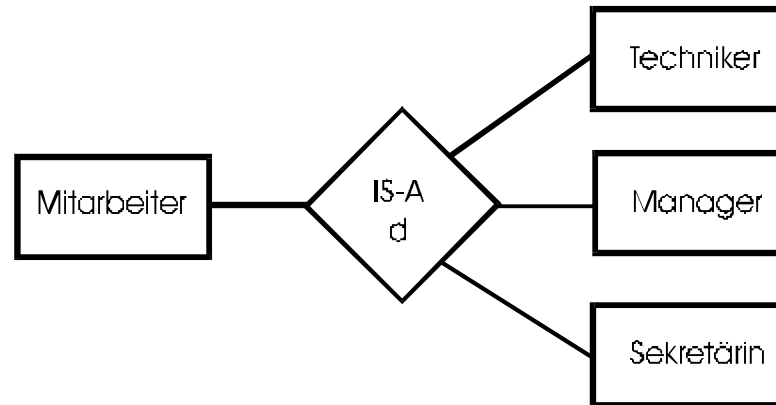
Beziehungen

Existenz, tenär

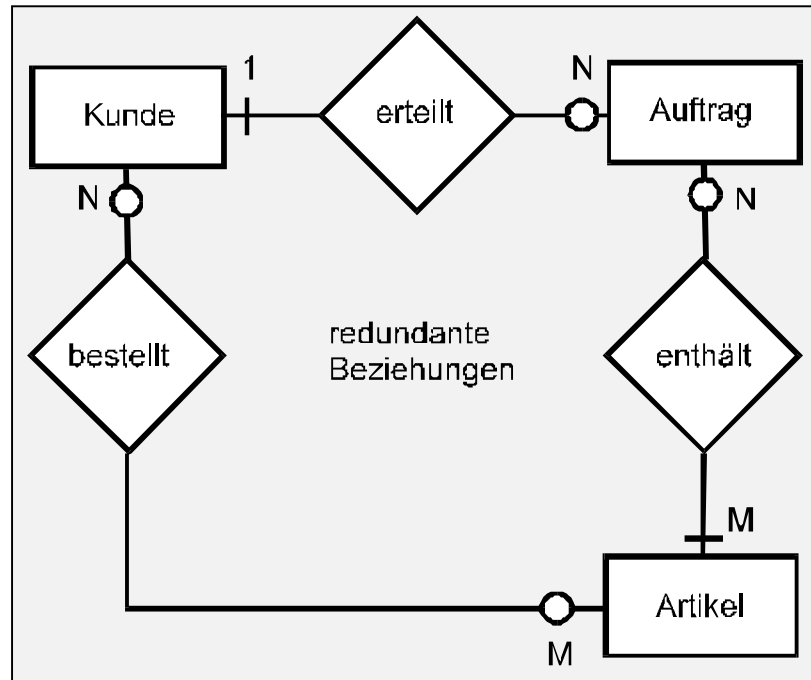


Beziehungen

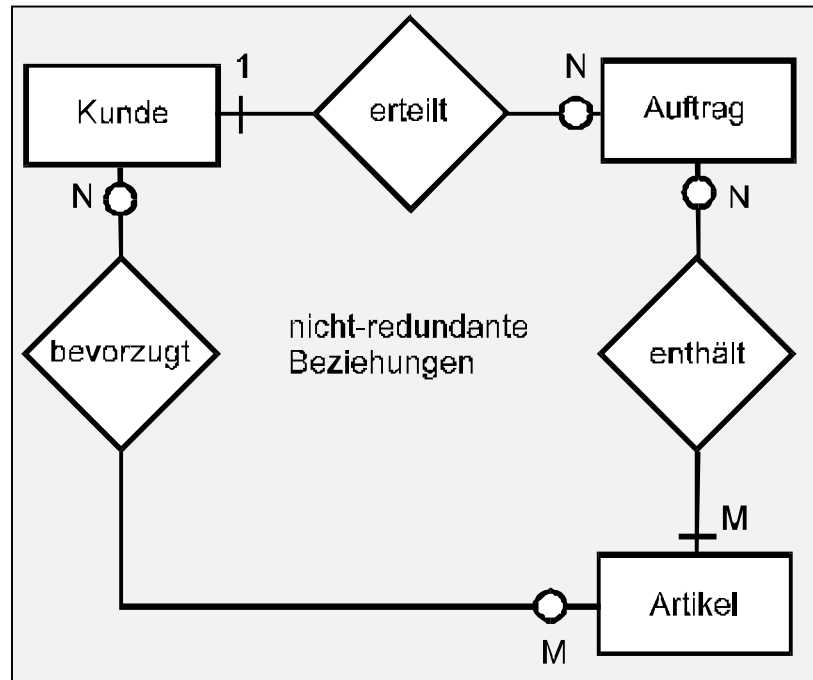
Generalisierung, Part-Of



Redundante Beziehung

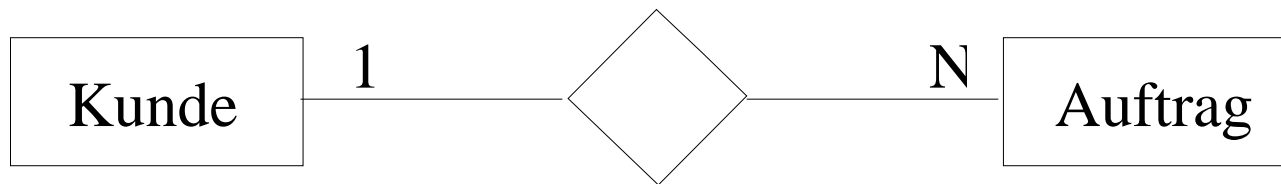


Nicht-redundante Beziehung



Beziehungen und Tabellenstruktur

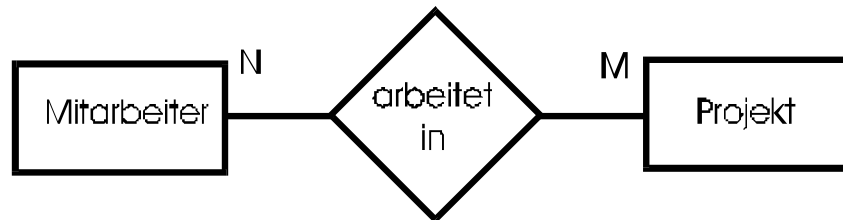
- **1 zu N: Primärschlüssel der 1-Entity wird Fremdschlüssel der N-Entity**



- **Kundennummer wird Fremdschlüssel in Auftrag**

Beziehungen und Tabellenstruktur

- N zu M: Es ergibt sich eine zusätzliche Tabelle



- Tabelle Mitarbeiter: M-Nummer, M-Name, etc
- Tabelle Projekt: P-Nummer, P-Name, etc
- MitarbeiterInProjekt: *M-Nummer*, *P-Nummer*, Arbeitszeit prozentual

Normalisierung

- Keine Tabelle in Tabelle (z.B. Auftragsnummer in Kundentabelle)
- Jedes Nicht–Schlüsselattribut ist von allen Schlüsselattributen abhängig (Tabelle Position mit ArtNr, AufNr, Menge, Preis)
- Keine Abhängigkeiten unter Nichtschlüsselattributen (Tabelle Artikel mit ArtNr, ArtBez, Preis, LagNr, LagOrt, LagStr)
- Keine berechneten Felder (Tabelle Rechnung mit RechNr, RechDat, AufNr, NettoWert, MWSt, BruttoWert)

SQL

- **SELECT * FROM KUNDE;**
- **SELECT KKURZ, FIRMA FROM KUNDE ORDER BY FIRMA;**
- **SELECT KKURZ, FIRMA, LKZ, ORT FROM KUNDE WHERE KKURZ > 105 AND LKZ = "D";**
- **SELECT AUFTRAG.AKURZ, AUFTRAG.KKURZ, KUNDE.FIRMA, AUFTRAG.ASTNR FROM AUFTRAG, KUNDE WHERE AUFTRAG.KKURZ = KUNDE.KKURZ AND ASTNR IN (1, 2);**

Data Warehouse

B. Inmon:

Ein Data Warehouse ist eine

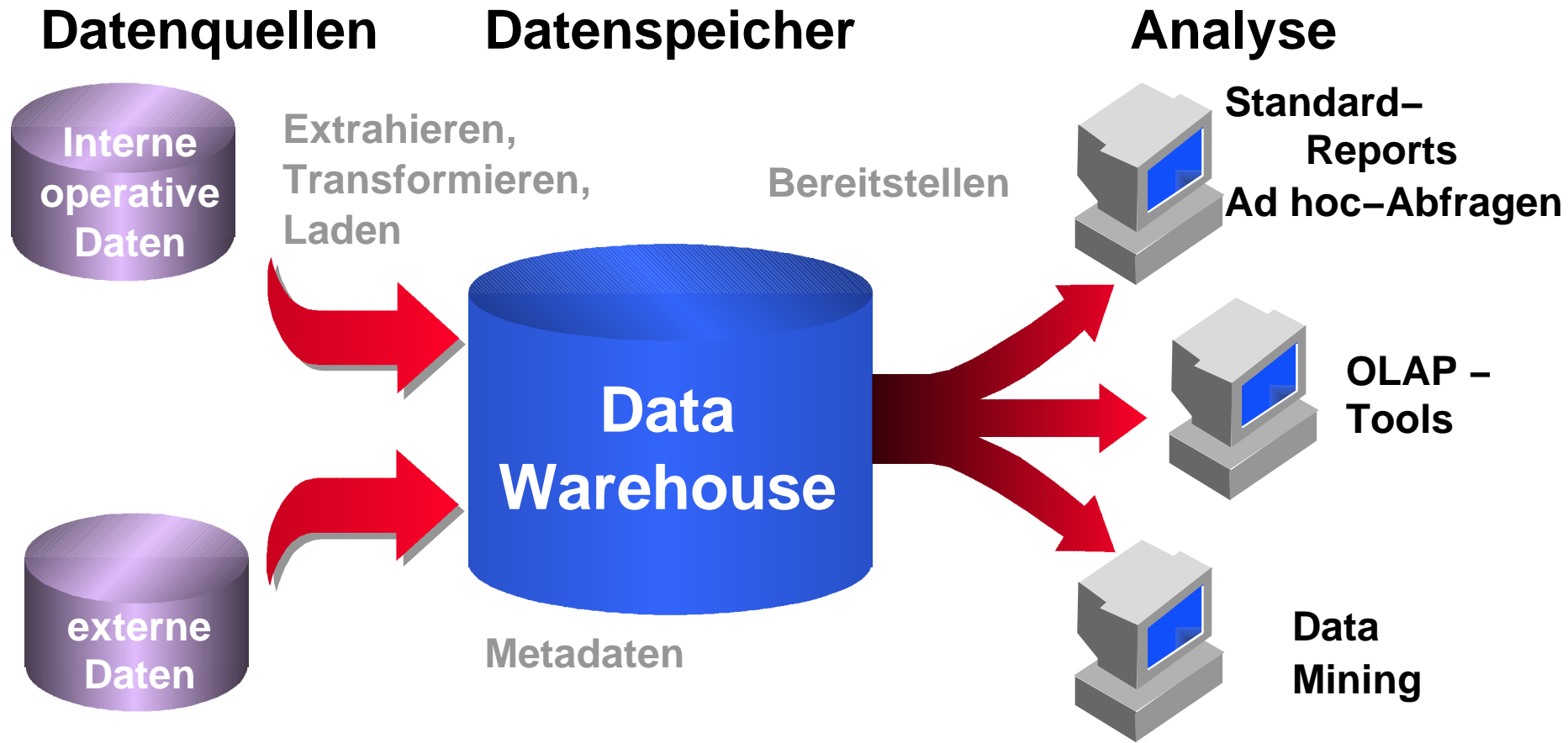
- integrierte,***
- themenorientierte,***
- zeitlich veränderliche,***
- nichtflüchtige***

***Datensammlung zur Unterstützung
von Management-Entscheidungen“***

Anforderungen an ein Data Warehouse

- **Verbesserter Zugang zu Informationen**
- **Globale, konsistente und standardisierte Sicht auf die Unternehmensdaten**
 - ⇒ alle Anwender sollen mit den gleichen Zahlen arbeiten
- **Self Service (Empowering) des Endbenutzers**
 - ⇒ Jeder Anwender soll seine eigenen Geschäftsberichte und Analysen erstellen;
alle Anwender sollen mit den gleichen Tools arbeiten
- **Einbeziehung externer Datenquellen**
- **Berücksichtigung der Historie**

Architektur eines Data Warehouse



Daten



Information



Wissen

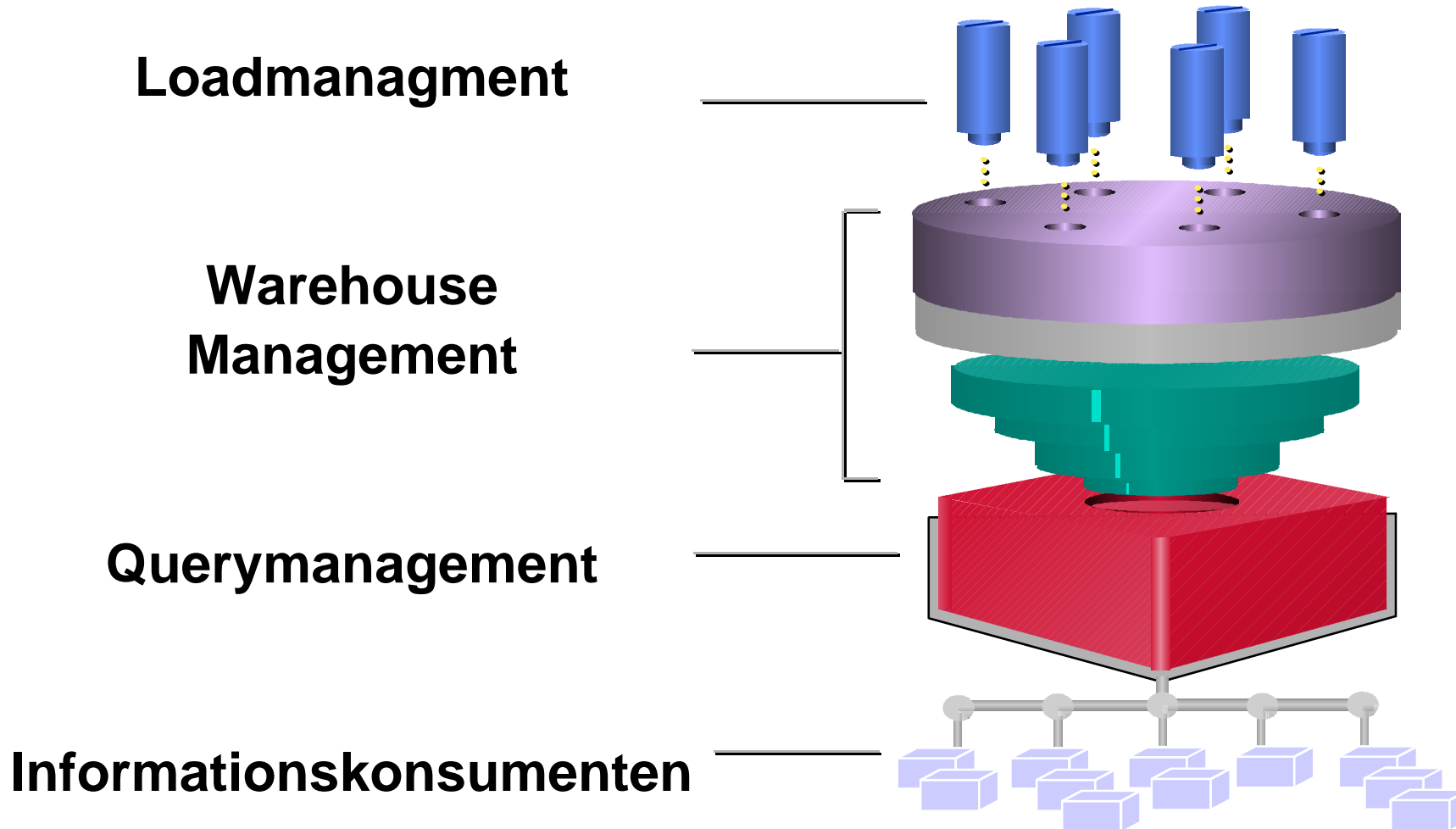
Nutzen eines Data Warehouse

- **Vereinheitlichung der Daten**
 - unterschiedliche operative Vor-Systeme
 - externe Daten
- **Transformation der Daten**
 - verdichten, filtern, konvertieren, normieren
- **Optimale Struktur für Analyse**
- **Korrektheit der Daten**
 - Konsistenz, Vollständigkeit, prüfbar
- **Halten der Historie**
- **Keine Belastung des operativen Systems**

Strategisches Vorgehen

- **Wo liegt der Nutzen für das Unternehmen ?**
- **Welche Veränderungen verlangen nach einer DW-Lösung ?**
- **Welche Unternehmensbereiche profitieren am meisten ?**
- **Zeitliche Rahmenbedingungen ?**
- **Beispielabfragen**
- **Kritische Erfolgsfaktoren ?**

Entwurf der Architektur



Trends und Perspektiven

Web-Integration

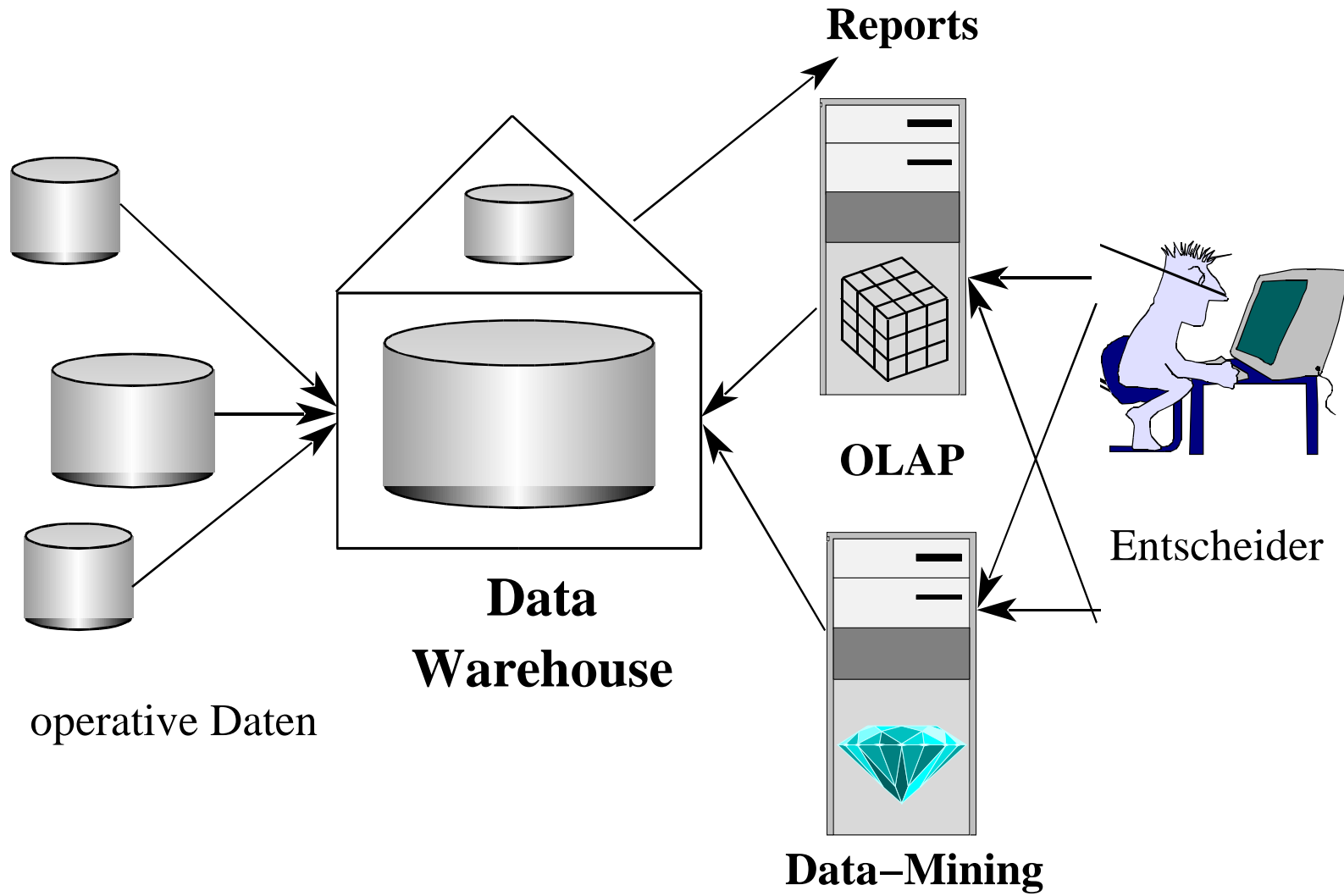
Wissensmanagement

Daten Information
Warehouse
Schöpfungskette
Wissen Aktion

Data Mining

Informations-Portal

Data Warehouse und Data Mining

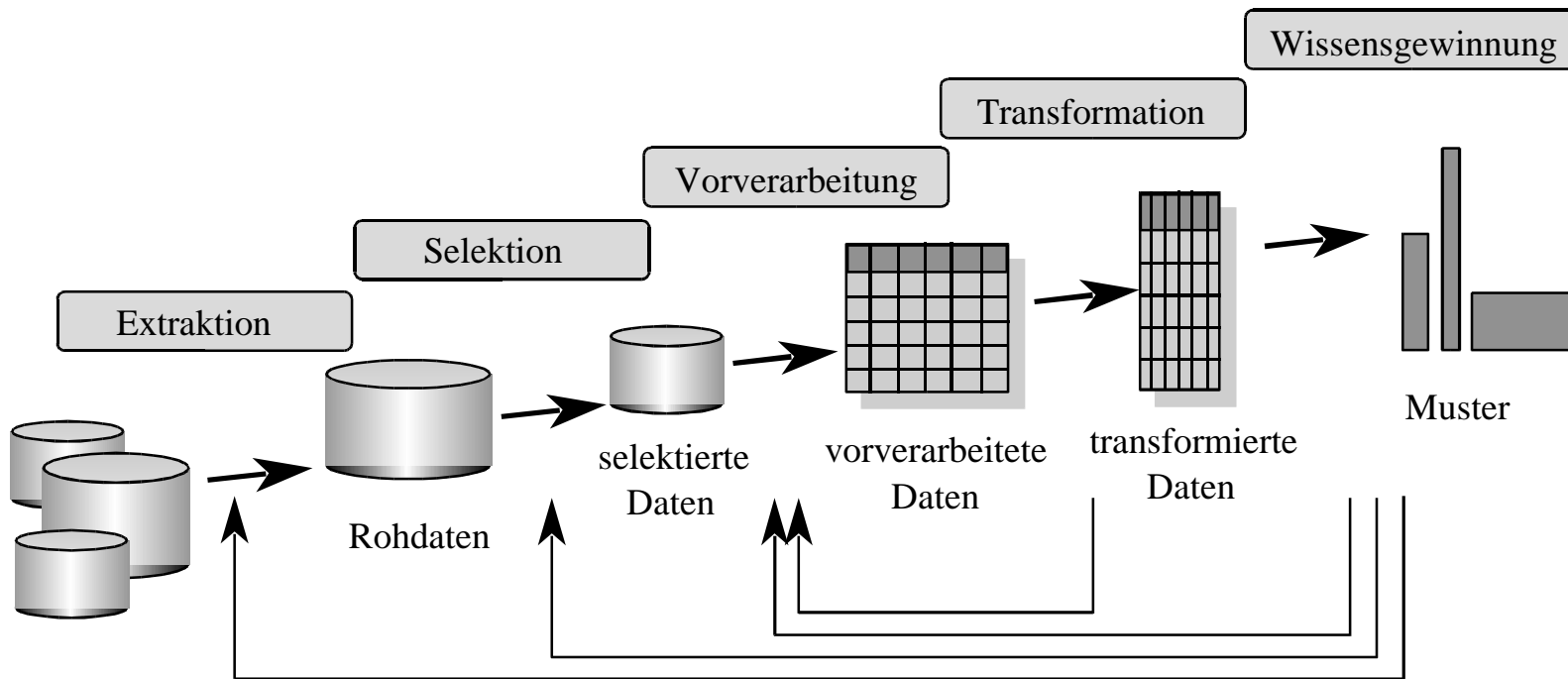


Data Mining

Definition

The process of extracting valid, previously unknown, and ultimately comprehensible information from large databases and using it to make crucial business decisions.

Data Mining Prozeß



Data Mining

Chancen und Risiken

- **Automatische Erkennung von Zusammenhängen (Vorhersage, Beschreibung)**
- **Völlig neue Erkenntnisse über das Business des Kunden**
- **Data Mining als Bestandteil einer Data Warehouse-Lösung**
- **Kosten, Erwartungen, Trivialitätenerkennung, Datenschutz**
- **Data Mining ist ein aufwendiger Prozeß**
- **Projekte beratungsintensiv**
- **Forschungsgebiet vs. Markt**