



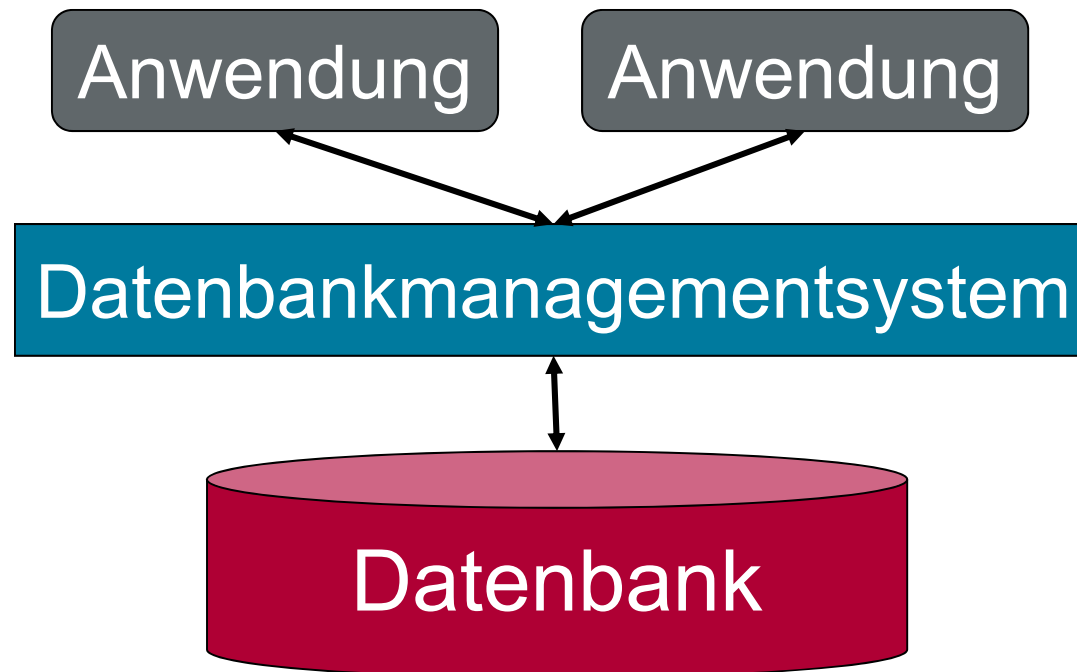
**Hasso
Plattner
Institut**

IT Systems Engineering | Universität Potsdam

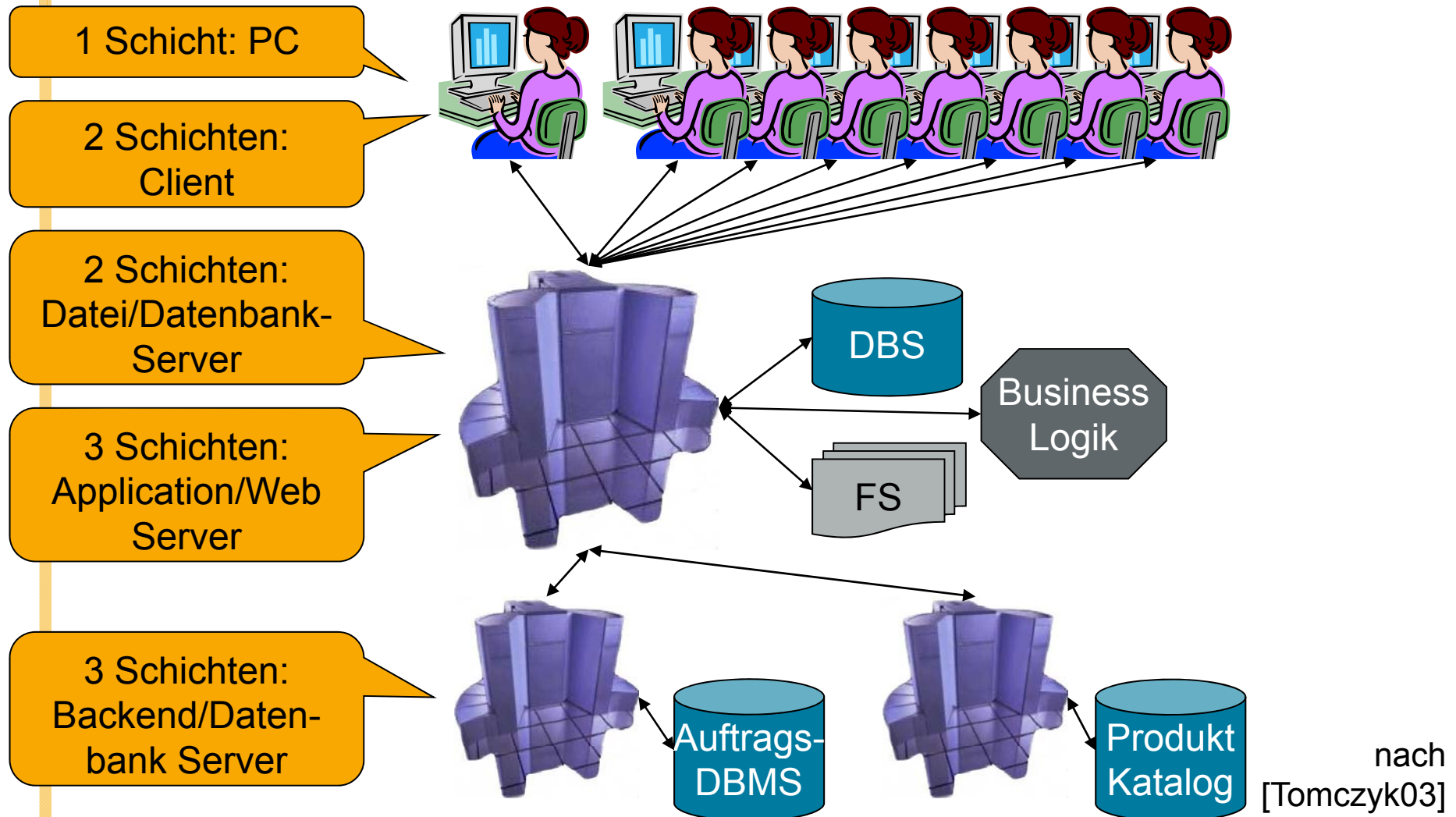
Datenbanksysteme 1
Organisatorisches und Einführung

19.4.2010

Felix Naumann

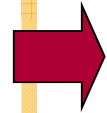


Mehr-Schichtenarchitekturen

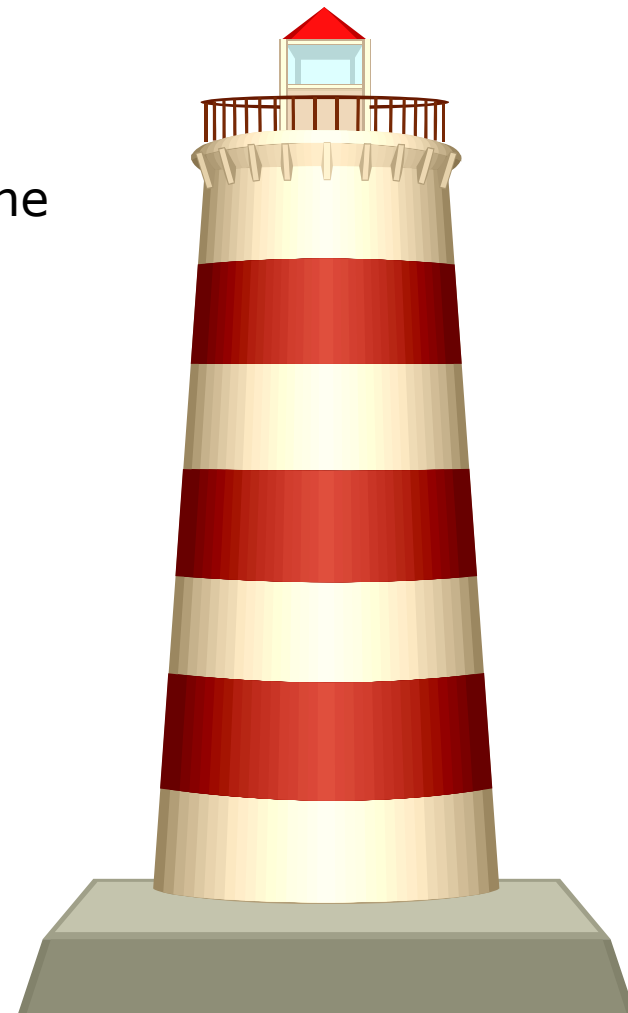


nach [Tomczyk03]

4



- Vorstellung der Arbeitsgruppe
- Organisatorisches
- Datenbanken und Informationssysteme
- Das Semester an einem Beispiel
- Ausblick auf das Semester



Information Systems Team

5

project **ViQTOR**



Paul Führung



Katrin Heinrich

DQ Annotation & Assessment



Prof. Felix Naumann

Data Fusion



Silke Trißi

Duplicate Detection



Uwe Draisbach



Dustin Lange

Schufa

Entity Search

Information Integration

project **HumMer**

Cloud Computing



Johannes Lorey

IBM

Data Profiling & Cleaning

Information Quality

Data Profiling



Armin Roth

project **System P**

Peer Data Management Systems

Matching

Data Integration for Life Science Data Sources

project **PoSR**

Service-Oriented Systems

project **Aladin**



Christoph Böhm

ETL Management



Tobias Vogel

Forschungskolleg



Mohammed AbuJarour

Ontologies, Profiling



Frank Kaufer



Jana Bauckmann



Alexander Albrecht

Data Profiling for Schema Management

Other courses in this semester

6

Lectures

- DBS I
- Information Integration

Seminars

- Bachelor: Beauty is our Business
- Bachelor: Mobile Application Development
- Master: Large-scale Data Analysis in the Cloud
- Master: Entity-centric Information Retrieval
- Master: Similarity Search Algorithms
- Forschungsseminar

Bachelorprojects

- ETL-Prozess-Management für BMW Financial Services
- IBM Midas: Extreme Web Data Integration for Government Data



Extending the Database Relational Model to Capture More Meaning

E. F. Codd
IBM Research Laboratory

During the last three or four years several investigators have been exploring "semantic models" for the data so that database design can become more systematic and the database system itself can behave more intelligently. Two major thrusts are clear:

- (1) the search for meaningful units that are as small as possible—atomic semantics;
- (2) the search for meaningful units that are larger than the usual n -ary relation—molecular semantics.

In this paper we propose extensions to the relational model to support certain atomic and molecular semantics. These extensions represent a synthesis of many ideas from the published work in semantic operations.

Key Words and Phrases: relation, relational database, relational model, relational schema, database, data model, database schema, data semantics, semantic model, knowledge representation, knowledge base, conceptual model, conceptual schema, entity model

CR Categories: 3.70, 3.73, 4.22, 4.26, 4.33, 4.34, 4.39

1. INTRODUCTION

The relational model for formatted databases [5] was conceived ten years ago, primarily as a tool to free users from the frustrations of having to deal with the clutter of storage representation details. This implementation, independence coupled with the power of the algebraic operators on n -ary relations and the open questions concerning dependencies (functional, multivalued, and join) within and between relations have stimulated research in database management (see [30]). The relational model has also provided an architectural focus for the design of databases and some general-purpose database management systems such as MACAIMS [13], PRTV [38], RDMS(GM) [41], MAGNUM [19], INGRESS [37], QBE [46], and System R [2].

During the last few years numerous investigations have been aimed at capturing

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

A version of this work was presented at the 1979 International Conference on Management of Data (SIGMOD), Boston, Mass., May 25-June 1, 1979.

Author's address: IBM Research Laboratory K01/282, 5600 Cottle Road, San Jose, CA 95133.

© 1979 ACM 0362-5915/79/1200-0387 \$06.75

ACM Transactions on Database Systems, Vol. 4, No. 4, December 1979, Pages 387-434.

Beauty is our Business

„Wenn wir uns klarmachen, daß der Kampf gegen Chaos, Durcheinander und unbeherrschte Kompliziertheit eine der größten Herausforderungen der Informatik ist, müssen wir zugestehen: Beauty is our Business.“ Edsger W. Dijkstra, 1978

Termine und Themenvergabe

Termin	Thema
19.4.2010	Einführung
26.4.2010	Wissenschaftliche Texte Lesen
10.5.2010	Literaturkritik / Diskussion
24.5.2010	Vortragstechniken
7.6.2010	Vortrag 1: Mariposa (pdf)
	Vortrag 2: Trio (pdf)
14.6.2010	Einführung in LaTeX
21.6.2010	Vortrag 3: Fagins Algorithmus (pdf)
	Vortrag 5: Enough Already in SQL (pdf)
5.7.2010	Vortrag 4: Sorted Neighborhood (pdf)
	Vortrag 6: Data Mining (pdf)
12.7.2010	Vorstellung der Gliederungen & Tipps zur Ausarbeitung
?.?.2010	Abgabe der Ausarbeitungen

Projektseminar

Mobile Application Development (MAD)



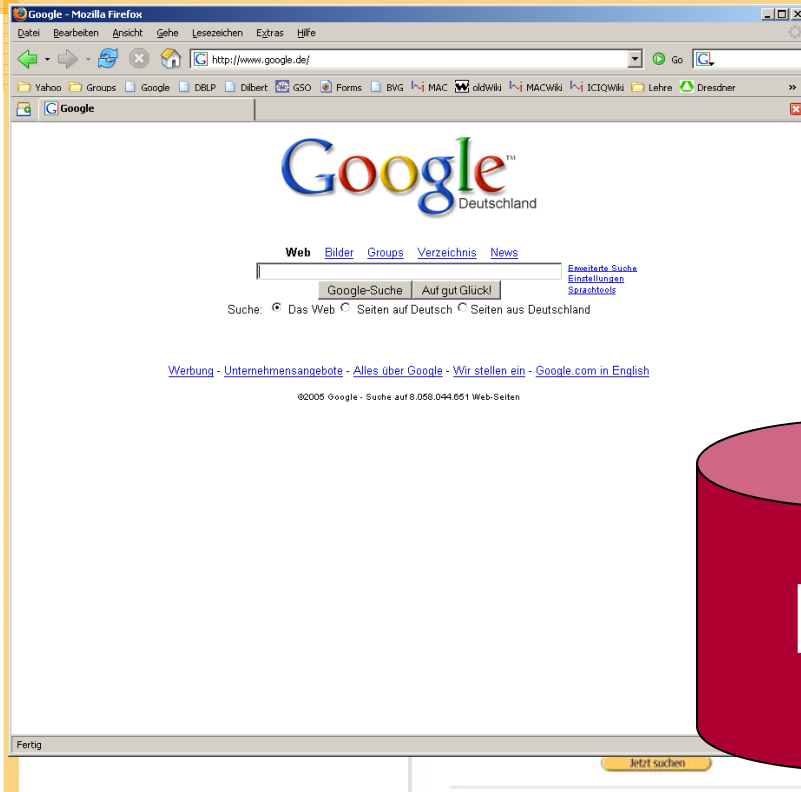
8

- (Datenbank-)Themen
 - **Data & information retrieval** (camera, audio, GPS, compass, speed, personal data, ...)
 - **Peer-to-peer communication** (Google Talk)
 - **Data management** (local vs. remote, caching, power efficiency)
 - **Stream processing** (web data, location data, ...)
 - **Mobile databases** (SQLite)
 - **Visualisierung** von Daten (Grafiken, Animationen), tbc.
- Ablauf
 - Jedes Team entwickelt eine Android App zu einem Thema
 - Präsentation zum Semesterende
 - 8 Teams à 2 Studenten à 1 G1 Dev Phone
 - Wöchentliche Veranstaltungen (Techniken, Programmierung, Beispiele, Übungen)
 - http://code.google.com/android/adc/gallery_winners.html
- Start: Heute 11 Uhr in A.1-1 !



Was sind Informationssysteme?

9



DB

```

ID  RNGTPCHI  standard; RNA; ROD; 1016 BP.
XX
DT  01-AUG-1991 (Rel. 28, Created)
DT  04-MAR-2000 (Rel. 63, Last updated, Version 2)
XX
DE  Rat GTP cyclohydrolase I mRNA, complete cds.
XX
KW  GTP cyclohydrolase I.
XX
OS  Rattus norvegicus (Norway rat)
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC  Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
XX
RN  [1]
RP  1-1016
RX  MEDLINE; 91093270.
RX  PUBMED; 1985963.
RA  Hatakeyama K., Inoue Y., Harada T., Kagamiyama H.;
RT  "Cloning and sequencing of cDNA encoding rat GTP cyclohydrolase I: The
RT  first enzyme of the tetrahydrobiopterin biosynthetic pathway";
RL  J. Biol. Chem. 266(2):765-769(1991).
XX
FT  CDS
      128..853
      /codon_start=1
      /db_xref="GOA:P22288"
      /db_xref="SWISS-PROT:P22288"
      /EC_number="3.5.4.16"
      /gene="GTP cyclohydrolase I"
      /product="GTP cyclohydrolase I"
      /protein_id="AAA41299.1"
      /translation="MEKPRGVRCTNGFPERELPRPGASRPAEKSRPPEAKGQPADAWK
      AGRPRSEEDNELNLPNLAAAYSSILRSLGEDPQRQGLLKTUPRAATAMQFFTKGYQETI
      SDVLND&IIFDEHDHDMVIVKIDIMFSMCEHHLVFPVGRVHIGYLPNKQVLGSLKLARIV
      EIYSRRLQVQERLTKQIAVAITEALQPA&GVGVVIEATHMCMVMRQVMNSKRTVSTML
      GVFREDPKTREFEFLTIR"
Sequence 1016 BP: 236 A; 279 C; 291 G; 210 T; 0 other:
gacttgaac  etcattcgtt  gcagaactcc  tgtcccgggt  acagccacag  gteaagccgc
ccggctaagc  cgagcccgag  cgcttgtag  caacctaggg  tgtctcggga  gcaatcgcc
cgggtccatg  gagaagccgc  ggggtgtaag  gtgcaccaat  ggggtccccg  agcgggagct
...
catcaggagc  tgaacttcg  tgtgcgagcc  cgggtttgca  gacccccgt  gaggccagc
ttatctgtct  cgattgtaca  ttccagttcc  agttggtata  ctgtccaact  ttatttcca
ccatgaattg  tatttaataa  ttatttatag  agatgtcaaa  taaaggtgat  caactt
1016
  
```

Molecule type
Name
Date of creation and last update
Free text description
Keywords describing the molecule
Organism
Article the sequence was published in
Structural annotation (coding sequence)
Link to functional annotation of resulting protein
Translated protein sequence
Sequence of bases

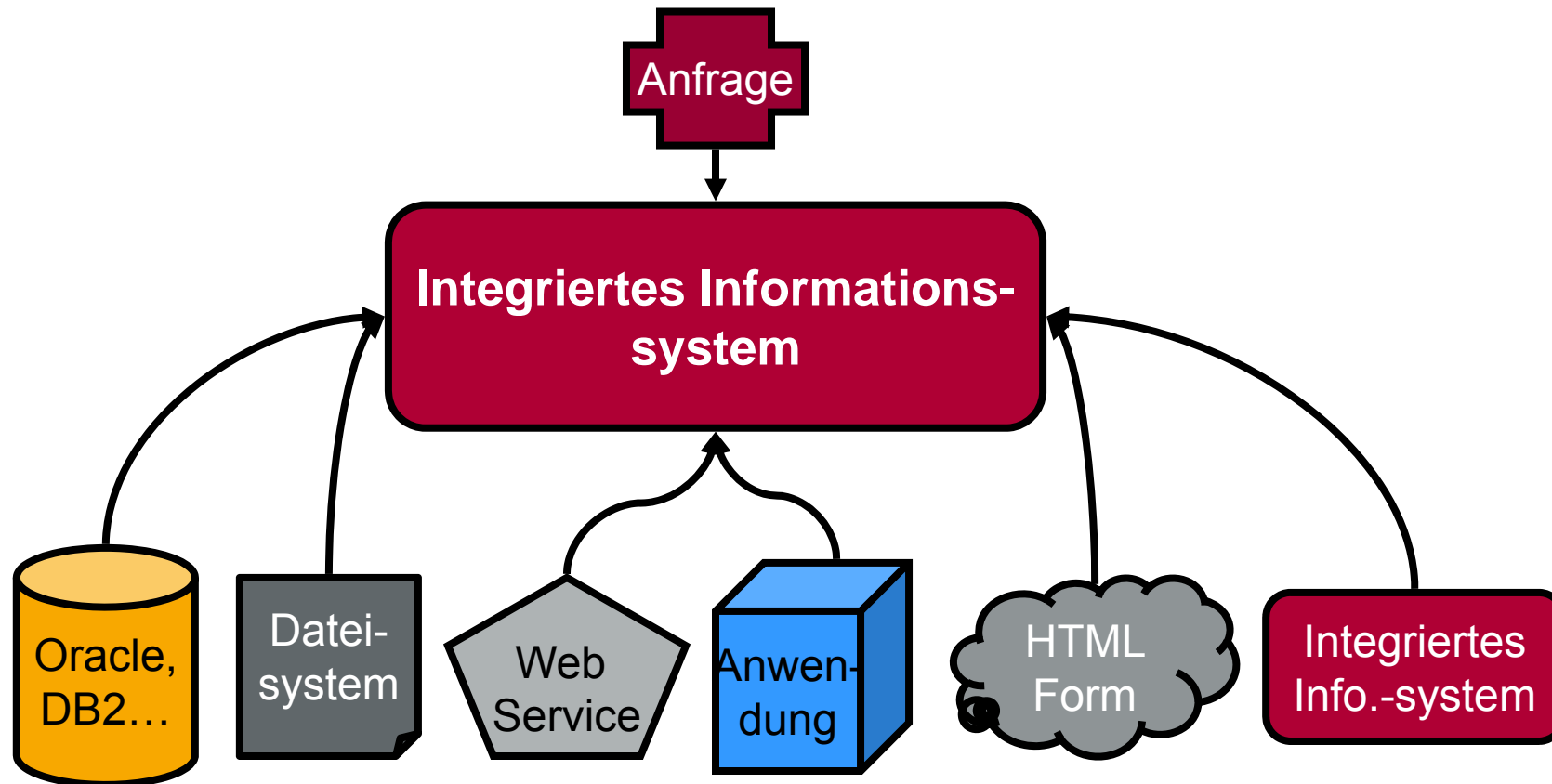
```

<buch>
  <isbn>0-201-318051</isbn>
  <titel>XML und Datenbanken</titel>
  <autor>Klettke/Meyer</autor>
</buch>
  
```



Integrierte Informationssysteme

10



Schematische und Daten-Heterogenität

11

Variante 1

Männer	
Vorname	Nachname
Felix	Naumann
Jens	Bleiholder

Frauen

Vorname	Nachname
Melanie	Weis
Jana	Bauckmann

Variante 2

Personen			
Vorname	Nachname	Männl.	Weibl.
Felix	Naumann	Ja	Nein
Jens	Bleiholder	Ja	Nein
Melanie	Weis	Nein	Ja
Jana	Bauckmann	Nein	Ja

Variante 3

Personen		
Vorname	Nachname	Geschlecht
Felix	Naumann	Männlich
Jens	Bleiholder	Männlich
Melanie	Weis	Weiblich
Jana	Bauckmann	Weiblich

Schematische und Daten-Heterogenität

12

Variante 1

Männer	
Vorname	Nachname
Felix	Naumann
Jens	Bleiholder

Frauen	
Vorname	Nachname
Melanie	Weis
Jana	Bauckmann

Variante 2

Personen			
FirstNa	Name	male	femal
Felix	Naumann	Ja	Nein
Jnes	Bleiho.	Ja	Nein
Melanie	Weiß	Nein	Ja
Jana	baukman	Nein	Ja

Variante 3

Personen		
VN	NN	SEX
F.	Naumann	Männlich
J.	Bleiholder	Männlich
M.	Weis	Weiblich
J.	Bauckmann	Weiblich

Schematische und Daten-Heterogenität

13

Variante 1

*■□■□■	
†□○■	♣■□○■
♣■●×	♣○◆○■
☺■●	×■□●○■

♣○◆■	
†□○■	♣■□○■
♣■●○■	♣■×
☺■○	♣&○■

Variante 2

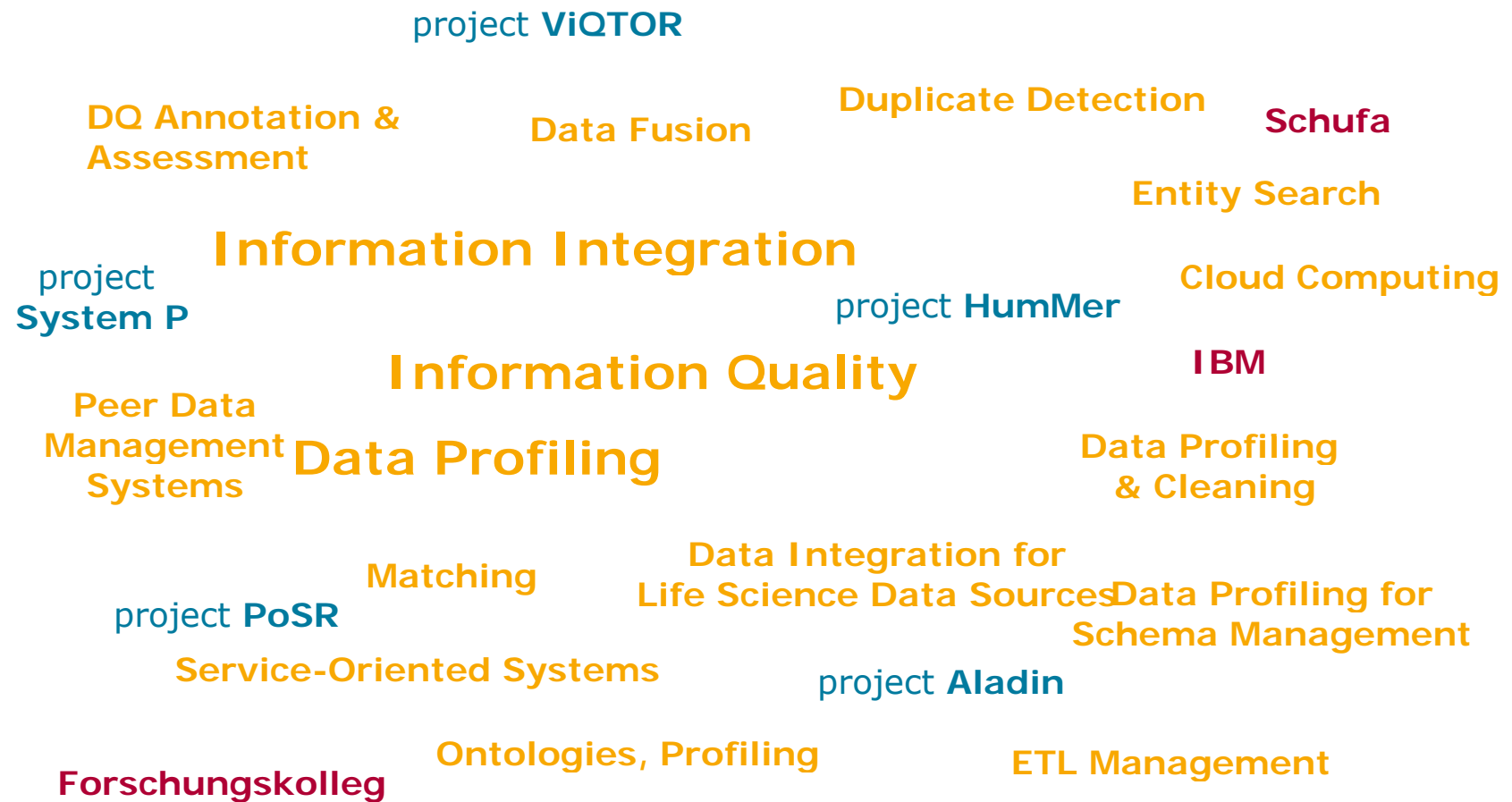
♣■□○■			
†□□■	♣■○■	■●	●
♣■●×	♣○◆○■	📁	📁
☺■●	♣●○■	📁	📁
♣■●○■	♣■×	📁	📁
☺■○	&○■	📁	📁

Variante 3

♣■□○■		
■○■	♣■○■	♣●■×◆
♣■●×	♣○◆○■	♣■●×■
☺■●	♣●○■	♣■●×■
♣■●○■	♣■×	♣○●×■
☺■○	♣○◆■&	♣○●×■

Weitere Forschungsthemen

14



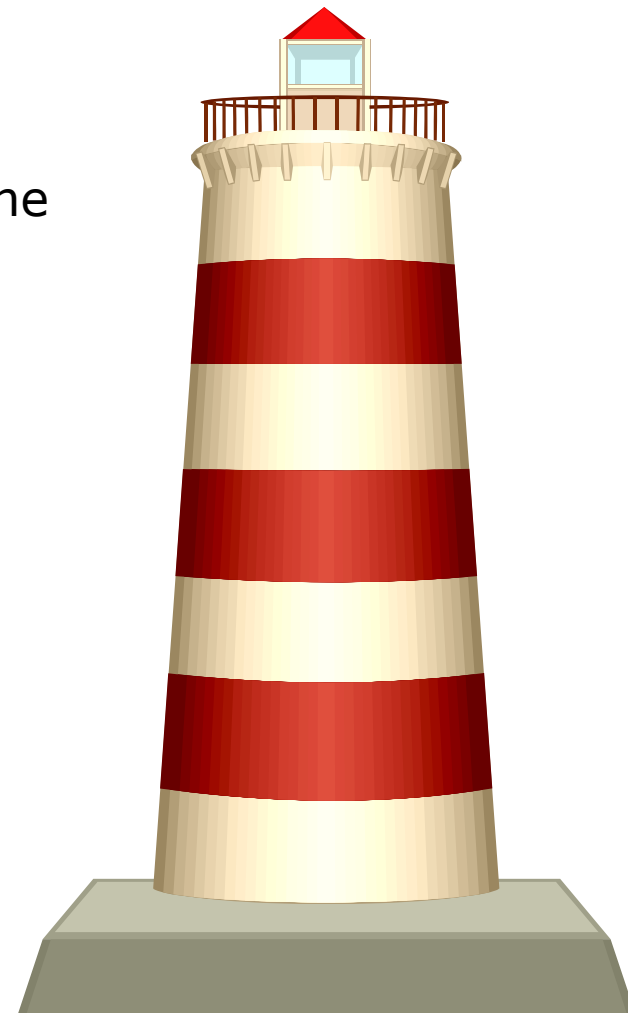
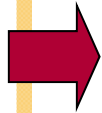
Vorstellung – Hörer

15

- Welches Semester?
 - 2. bzw 4. Semester
- Erasmus o.ä.?
- Berlin/Brandenburg oder anderswo?

- Berührung mit Datenbanken?
 - Wie?

- Vorstellung der Arbeitsgruppe
- Organisatorisches
- Datenbanken und Informationssysteme
- Das Semester an einem Beispiel
- Ausblick auf das Semester



Termine

17

- Montags 9:15 - 10:45 (müde)
 - Hörsaal 1
- Freitags 11:00 - 12:45 (hungrig)
 - Hörsaal 2
- Erste Vorlesung
 - 19.4.2009
- Letzte Vorlesung
 - 23.7.2009
- Relevante Feiertag(e)
 - Pfingstmontag am 24. Juni
- Klausur
 - Nach Semester
 - 180(!) Minuten
- Übungen
 - Ungefähr 1 SWS
 - Verteilt auf Doppelstunden im Semester
 - Drei parallele Übungsgruppen
- Übungsleitung: Silke Trißl
- Tutoren-Team
 - Fabian Tschirschnitz
 - Christoph Thiele

Freitag 23.4. erst um 17 Uhr!

Übung: Termine

18

- 3 identische Übungstermine pro Woche
 - Mi 13:30-15:00 Uhr in A-2.2
 - Mi 15:15-16:45 Uhr in A-1.2
 - Fr 11:00 – 12:30 Uhr in HS 2 (statt Vorlesung)

- Anmeldung
 - E-Mail – bis 23.4. – an dbs1-2010@hpi.uni-potsdam.de mit Reihenfolge der drei Termine
 - First come, first served!
 - Wir ordnen die Übungstermine fest zu und veröffentlichen das Ergebnis unter \\fs3\lehrveranstaltungen\FG_Naumann\DBSI_naumann

Übung: Aufgabenblätter

19

- Insgesamt 6 Aufgabenblätter (Übungen)
 - „Theoretische“ Übungen
 - ◇ Auf Papier
 - Praktische Übungen
 - ◇ DB2 im Studentenpool
 - ◇ Ergebnisse auf Papier
- Bearbeitung der Aufgabenblätter in **2er Gruppen**
- Prüfungszulassung:
 - Maximal ein Aufgabenblatt mit mind. 25% der Punkte bewertet
 - Alle weiteren Aufgabenblätter mit mind. 50% der Punkte bewertet
- Abgabe
 - Per E-Mail an dbs1-2010@hpi.uni-potsdam.de
 - **Pro Aufgabe** eine Datei mit Dateiname:
blatt<aufgabenblattNr>aufgabe<aufgabenNr><Namen>.pdf

Übung: Hilfe bei Problemen

20

- Sprechstunde Silke (in den Übungswochen)
 - Mittwochs

- Mailverteiler für Fragen zur Übung: `fragen-dbs1-2010@hpi...`
 - Hilfe zur Selbsthilfe: Fragen gegenseitig beantworten
 - Mitarbeiter lesen (und antworten) auch mit
 - Admins tragen alle DBS1-Studenten auf der Liste ein, sobald die Teilnehmer der VL feststehen

- Folien
 - Vor der VL im WWW
 - ASAP
- Anregungen zur Verbesserung:
 - Gebrauch der Folien
 - Infos im WWW
 - Übungsbetrieb
 - ...
- Fragen bitte jederzeit!
 - In der Vorlesung
 - Sprechstunde
 - ◇ Dienstags 15:00 – 16:00
 - ◇ Raum A-1.13
 - ◇ Am liebsten mit Anmeldung
 - Email: naumann@hpi.uni-potsdam.de

Feedback

22

- gut gelaunt und engagiert
- vermittelt positive Stimmung
- aber redet ein wenig zu schnell und zu viel

Die Bewertung bzw. Korrigierung der Übungsaufgaben ist sehr schön, durch das genaue Korrigieren erhält man einen guten Feedback.

Es ist beruhigend, wenn man merkt, wenn der Dozent bei einem bestimmten Problem auch nicht 100% sicher ist.

Vote
Naumann for
Best-Prof-
WS-06/07

Die Vorlesung war sehr angenehm, auch theoretische Teile wurden so dargestellt, dass man nicht gleich eingepennt ist.

Bitte kein DB2
mehr! Bitte bitte!

Dozent könnte
sein Sprechtempo
ab und zu etwas
herabsetzen

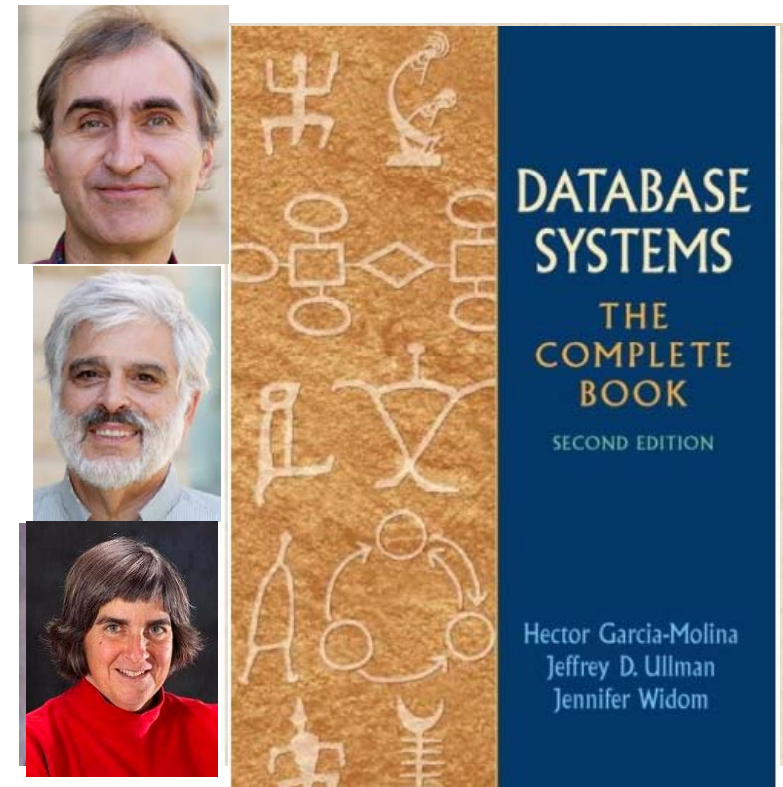
Es wäre schön wenn Sie in
Zukunft langsamer sprechen
könnten bzw. manchmal eine
Kunstpause einschieben könnten

Database Systems - The Complete Book

- Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom: Pearson Education International, 2002.

Ebenfalls empfehlenswert

- Datenbanksysteme
Alfons Kemper, André Eickler
ISBN: 3486273922
- Datenbanken:
Implementierungstechniken
Gunter Saake, Andreas Heuer, Kai-
Uwe Sattler



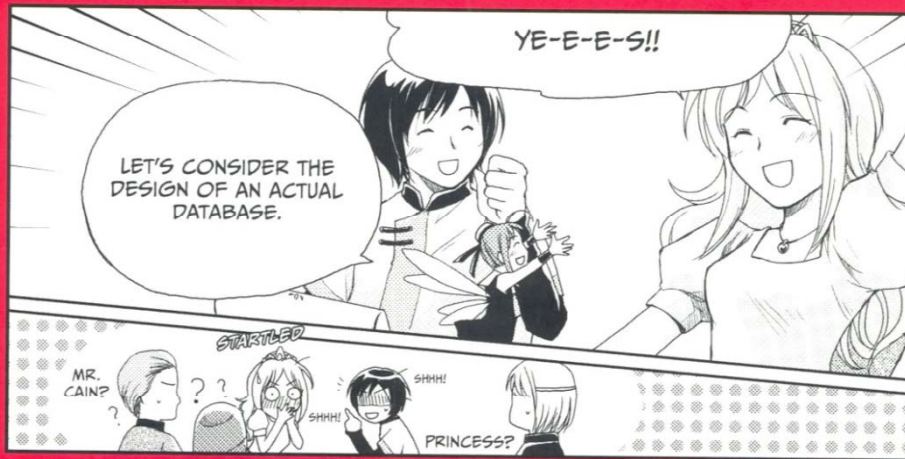
Und viele andere mehr...

Weitere Literatur

24

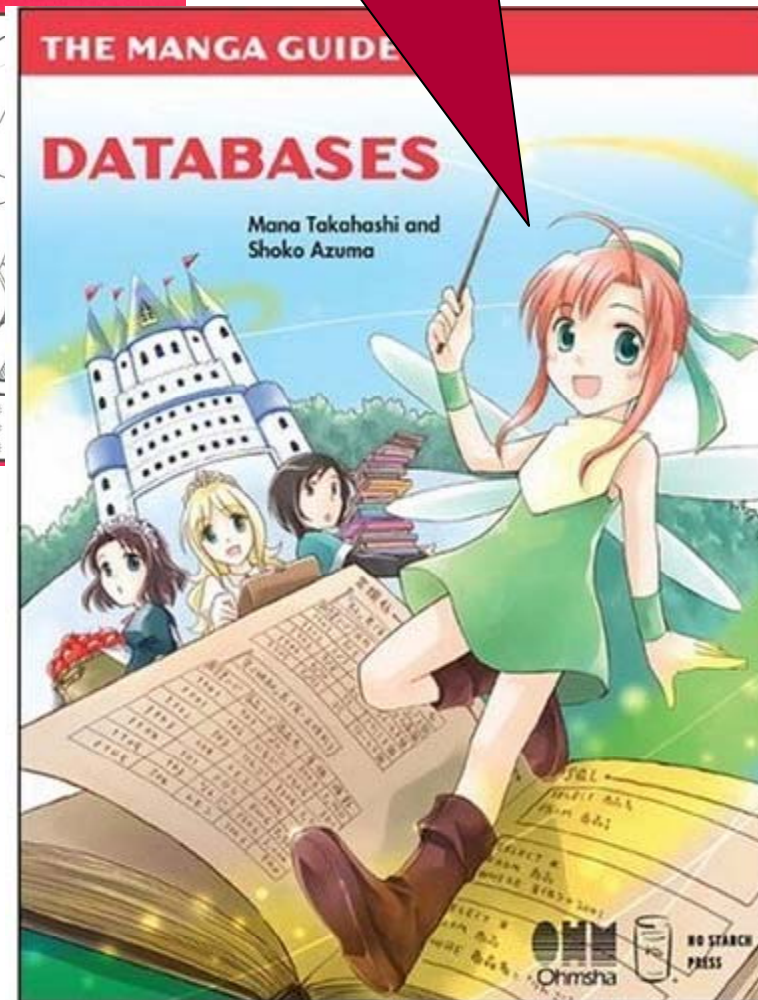
UNLEASH THE POWER OF THE DATABASE!

Magical database fairy

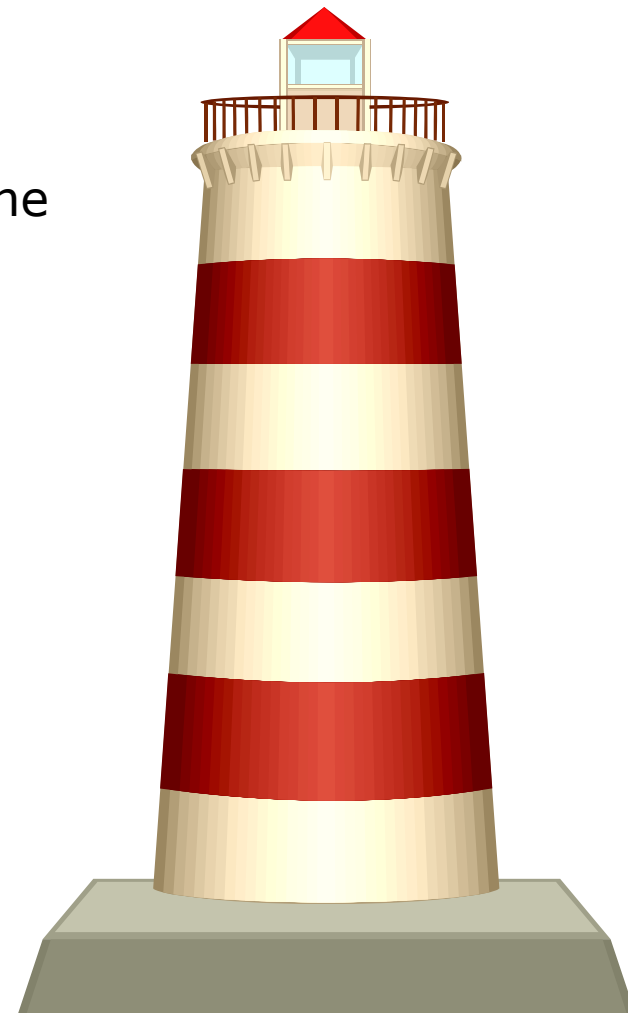
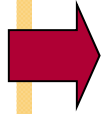


PRINCESS RURUNA AND CAIN HAVE A PROBLEM: THEIR FRUIT-SELLING EMPIRE IS A TANGLE OF CONFLICTING AND DUPLICATED DATA, AND SORTING THE MELONS FROM THE APPLES AND STRAWBERRIES IS CAUSING REAL DIFFICULTIES. BUT WHAT CAN THEY DO?

WHY, BUILD A RELATIONAL DATABASE OF COURSE, WITH THE HELP OF TICO, THE MAGICAL DATABASE FAIRY. FOLLOW ALONG IN *THE MANGA GUIDE TO DATABASES* AS TICO TEACHES RURUNA AND CAIN HOW TO BUILD A DATABASE TO MANAGE THEIR KINGDOM'S SALES, MERCHANDISE, AND EXPORTS. YOU'LL LEARN HOW DATABASES WORK AND THE MEANING OF TERMS LIKE SCHEMAS, KEYS, NORMALIZATION, AND TRANSACTIONS.



- Vorstellung der Arbeitsgruppe
- Organisatorisches
- Datenbanken und Informationssysteme
- Das Semester an einem Beispiel
- Ausblick auf das Semester



Was sind Daten?

26

Digitale Repräsentation von

- Dingen
- Entitäten
- Wissen
- Information

in/aus der wirklichen Welt.

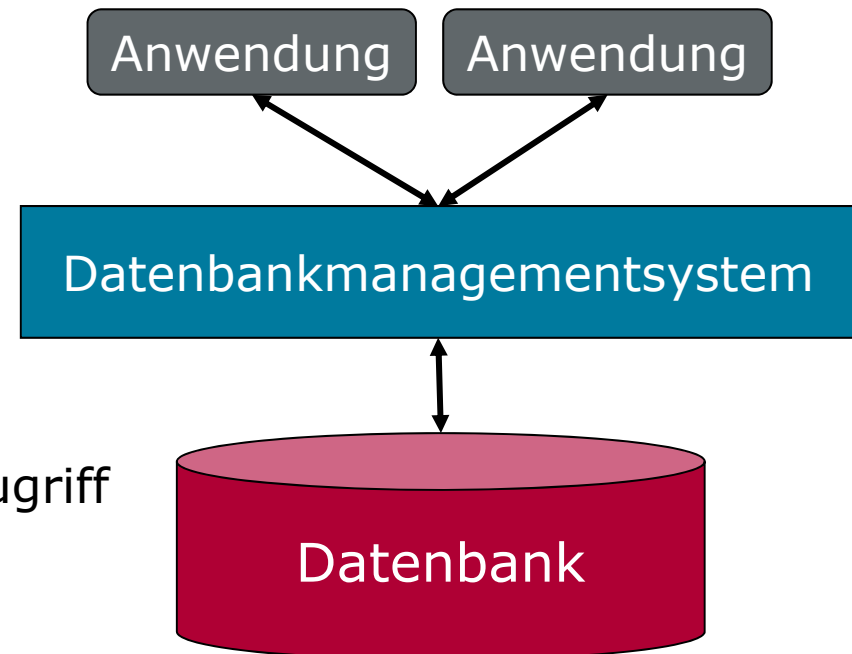
Kernfragen:

- Welche Daten speichere ich?
- Wie speichere ich die Daten?
- Wie frage ich Daten ab?
- Wie erledige ich all dies effizient und sicher?

} Unterstützung durch
Datenbanksystem

Bestandteile

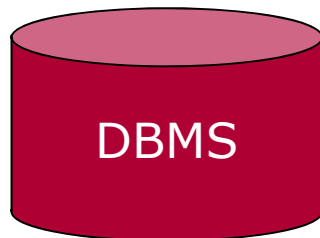
- Datenbank
 - Die Daten selbst
 - + Metadaten (Beschreibung der Daten)
- Datenbankmanagementsystem (DBMS)
 - Softwarekomponente zum Zugriff auf eine oder mehrere Datenbanken.
 - Server-basiert
- (Anwendungen)



DBMS Beispiele

28

- DBMS
 - OLTP (Online Transaction Processing)
 - Oracle, DB2, SQL-Server
 - Informix, Sybase
 - Teradata
 - PostgreSQL, Interbase
 - MySQL, ...
- Data Warehouses
 - OLAP (Online Analytical Processing)
- Alle großen Softwaresysteme nutzen ein, mehrere oder sehr viele DBMS.
 - SAP
 - Siebel
 - SABRE
 - Banken, Versicherungen, Telekoms, usw.



ORACLE

tamino

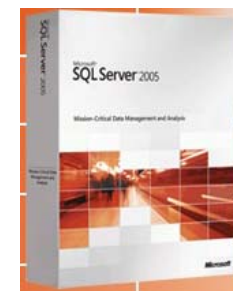
SYBASE



DB2 9 for Linux UNIX and Windows

pureXML™ and storage compression

Information Management software



DBMS Aufgaben

29

- Unterstützung des Datenmodells
- Bereitstellung einer Anfragesprache (DDL & DML)
- Effiziente Anfragebearbeitung
- Robustheit
 - Wahrung der Datenintegrität (Konsistenz etc.)
 - Abfangen von Systemfehlern
- Speicherverwaltung (RAM & Disk)
- Transaktionsmanagement
 - Auch im Mehr-Benutzer-Betrieb
- Nutzerverwaltung & Zugangskontrolle

Beispiele für Informationssysteme

30

■ Dateisystem

- Informationseinheit:
Flat file
- Anfrage:
File search, RegEx
- Struktur:
Flach oder hierarchisch

■ Beispiele

- NTFS
- NIS
- FTP Zugriff

■ Einsatzgebiete

- WWW (HTML Dateien)
- Desktop-Anwendungen
(Textverarbeitung, etc.)

Beispiele für Informationssysteme

31

Datei

- Informationseinheit:
Zeile /Token

- Anfrage:
Parser

- Struktur:
Flach

Beispiele

- Komma-delimited files
- Annotated files

Einsatzgebiete

- SwissProt

```

ID   RNGTHPCHI    standard; RNA; ROD; 1016 BP.
XX
DT   01-AUG-1991 (Rel. 28, Created)
DT   04-MAR-2000 (Rel. 63, Last updated, Version 2)
XX
DE   Rat GTP cyclohydrolase I mRNA, complete cds.
XX
KW   GTP cyclohydrolase I.
XX
OS   Rattus norvegicus (Norway rat)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
XX
RN   [1]
RP   1-1016
RX   MEDLINE; 91093270.
RX   PUBMED; 1985963.
RA   Hatakeyama K., Inoue Y., Harada T., Kagamiyama H.;
RT   "Cloning and sequencing of cDNA encoding rat GTP cyclohydrolase I: The
RT   first enzyme of the tetrahydrobiopterin biosynthetic pathway";
RL   J. Biol. Chem. 266(2):765-769(1991).
XX
FT   CDS                128..853
FT                       /codon_start=1
FT                       /db_xref="GOA:P22288"
FT                       /db_xref="SWISS-PROT:P22288"
FT                       /EC_number="3.5.4.16"
FT                       /gene="GTP cyclohydrolase I"
FT                       /product="GTP cyclohydrolase I"
FT                       /protein_id="AAA41299.1"
FT                       /translation="MEKPRGVRCITNGFPERELPRPGASRPAEKSRPEAKGAQPADAWK
FT   AGRPRSEEDNELNLPNLAAYSSILRSLGEDPQRQGLLKTWPRAATAMQFFTKGYQETI
FT   SDVLNDAIFDEHDENVIVKIDIDFMSCEHHLVFPVGRVHIGYLPNKQVLGSLKLARIV
FT   EIYSRRLQVQERLTKQIAVAITEALQPAQVGVVIEATHMCMVMRQVQKMNKSTVTSTML
FT   GVFREDPKTREEFLTLIRS"
SQ   Sequence 1016 BP; 236 A; 279 C; 291 G; 210 T; 0 other;
gacttcgaac ctcattcggg gcagaactcc tgtcccgggtg acagccacag gtcacggcgg      60
cgggtaagc cgagccgcag cgcttggtag caccttaggg tgtctcgga gcaatcgcgc      120
cgggtccatg gagaagccgc ggggtgtaag gtcaccaat ggggtccccg agcggggagct      180
...
catcaggagc tgaacttccg tgtgcgagcc cgggtttgca gacccccgct gaggccagcg      900
ttatctgtct cgattgtaca ttccagttcc agttgggtata cttgtcaact ttatttctca      960
ccatgaattg tatttaataa ttatttatag agatgtcaaa taaaggtgat caact          1016
//

```

Molecule type
Name
Date of creation and last update
Free text description
Keywords describing the molecule
Organism
Article the sequence was published in
Structural annotation (coding sequence)
Link to functional annotation of resulting protein
Translated protein sequence
Sequence of bases

Beispiele für Informationssysteme

32

- Markup Datei
 - Informationseinheit:
Tagged text
 - Anfrage:
Parser, Anfragesprache
 - Struktur:
Flach, hierarchisch oder
graph-basiert
- Beispiele
 - XML
 - HTML
- Einsatzgebiete
 - Web Services
 - Messages
 - Interoperable
Anwendungen

Beispiel: XML-Dokument

```
33 <?xml version="1.0" encoding="UTF-8"?>
<rechnung kundennummer="k333063143">
  <monatspreis>0,00</monatspreis>
  <einzelverbindungsachweis>
    <verbindung>
      <datum>26.2.</datum>
      <zeit>19:47</zeit>
      <nummer>200xxxx</nummer>
      <einzelpreis waehrung="Euro">0,66</einzelpreis>
    </verbindung>
    <verbindung>
      <datum>27.2.</datum>
      <zeit>19:06</zeit>
      <nummer>200xxxx</nummer>
      <einzelpreis waehrung="Euro">0.46</einzelpreis>
    </verbindung>
    <verbindungskosten_gesamt waehrung="Euro">2.19</verbindungskosten_gesamt>
  </einzelverbindungsachweis>
</rechnung>
```

Beispiele für Informationssysteme

34

■ Datenbank

- Anfrage:
Komplexe Sprache, z.B.
SQL
- Informationseinheit:
Tupel / Attribut, Objekt
- Struktur:
Relational, OO,
Hierarchisch

■ Beispiele

- Relationale DBMS
- OO DBMS
- Auch: XML DBMS

■ Einsatzgebiete

- Data Warehouses
- OLTP
- Banken/Versicherungen
- Telekom

Beispiele für Informationssysteme

35

■ HTML Formular

- Informationseinheit:
HTML Seite, Text
- Anfrage:
Suchworte, Formular
(inkl. Radiobutton,
dropdown-list, etc.)
- Struktur:
wie Markup Datei: Flach,
hierarchisch oder graph-
basiert
I.d.R. flach

■ Beispiele

- Einfache Suchformulare
- Komplexe
Anfrageformulare

■ Einsatzgebiete

- Suchmaschinen
- Reisedienste
- Kataloge

Dahinter: Fast immer ein DBMS!

Suche Bücher

HARRY POTTER und die Heiligtümer des Todes  Frühstück mit Harry [Jetzt portofrei vorbestellen](#)
am 27.10.2007 bis 10:30 Uhr geliefert

Erweiterte Suche Bücher

Je mehr Felder Sie ausfüllen, desto zielgerichteter können wir suchen. Es reicht jedoch aus, nur eines der Felder auszufüllen.

Autor/in:

Titel:

Schlagwörter:

ISBN: (10- oder 13-stellig, ohne Bindestriche)

Verlag:

Verfeinern Sie Ihre Suche, indem Sie nur nach bestimmten Buchformaten suchen lassen.

Nur gebraucht:

Format: Alle Formate

Ordnen nach: Topseller

Erscheinungsdatum: vor

Jeder Monat

2009

Suche in: deutsche Bücher englische Bücher

Beispiele für Informationssysteme

37

■ Web Service

- Informationseinheit:
XML Dokument
- Anfrage
XML Dokument
- Struktur:
Wie XML: Flach,
hierarchisch, graph-
basiert

■ Beispiele

- Einfach:
Temperaturdienst, etc
- Komplex: Reservierungen
(Schachtelung und
Verknüpfung von Web
Services)

■ Einsatzgebiete

- Intra-organisatorische
Workflows
- E-Marketplaces
- Datenaustausch

Dahinter: Fast immer ein DBMS!



Full Service List

Also accessible via XML Interfaces: [DISCO](#) [WS-Inspection](#) [RSS](#) See the [interfaces section](#) for more information.

List is ordered by submission time, with most recent services listed first.

Publisher	Style	Service Name	Description	Implementation
aandreu	RPC	Try It neuroFuzz Crypto Service	A simple demonstration of AES encryption and decryption via SOAP.	NuSOAP
XWebServices	DOC	Try It XWebBlog	XML/SOAP based Web Service which provides Content Management integration of a Weblog (Blog) to client applications.	MS .NET
StrikeIron	RPC	Try It LMSecurities	Latest stock quotes from Warsaw Stock Exchange	
kylehayes	RPC	Try It Amortization Calculator	Calculates amortization given principal, periodic interest rate, total number of payments	
jbardin	RPC	Try It Interconnect	Initiates calls between one or more phones	SOAP4R
StrikeIron	DOC	Try It U.S. Yellow Pages	Access to yellow pages listings for 17 million U.S. businesses	
7leapcom	RPC	Try It Conversão de IP para País	Converts an IP do it's country origin, retrieves not only country name, as well iso code, region, capital, and coin	NuSOAP
7leapcom	RPC	Try It Custo de Envio de Encomendas pelos CTT	Calculates shipping costs from Portugal via nacional CTT post offices	NuSOAP
agenteel	RPC	Try It Codigos Postales de Mexico	Codigos Postales de Mexico is a client/server implementation to lookup Postal Codes in Mexico.	gSOAP
Xignite	DOC	Try It XigniteCompensation	Biographical and Compensation information for more than 100,000 US executives.	MS .NET
Xignite	DOC	Try It XigniteWatchLists	Provide for the management of financial watchlists (stocks, currencies, rates...)	MS .NET
Xignite	DOC	Try It XigniteOFAC	Enables OFAC Compliance against the Specially Designed Nationals database	MS .NET

Beispiele für Informationssysteme

39

■ Anwendung

- Informationseinheit:
Java Objekt, Text
- Anfrage:
via Anwendungsschnittstelle oder GUI
- Struktur:
Objekt (Interface)
Display (GUI)

■ Beispiele

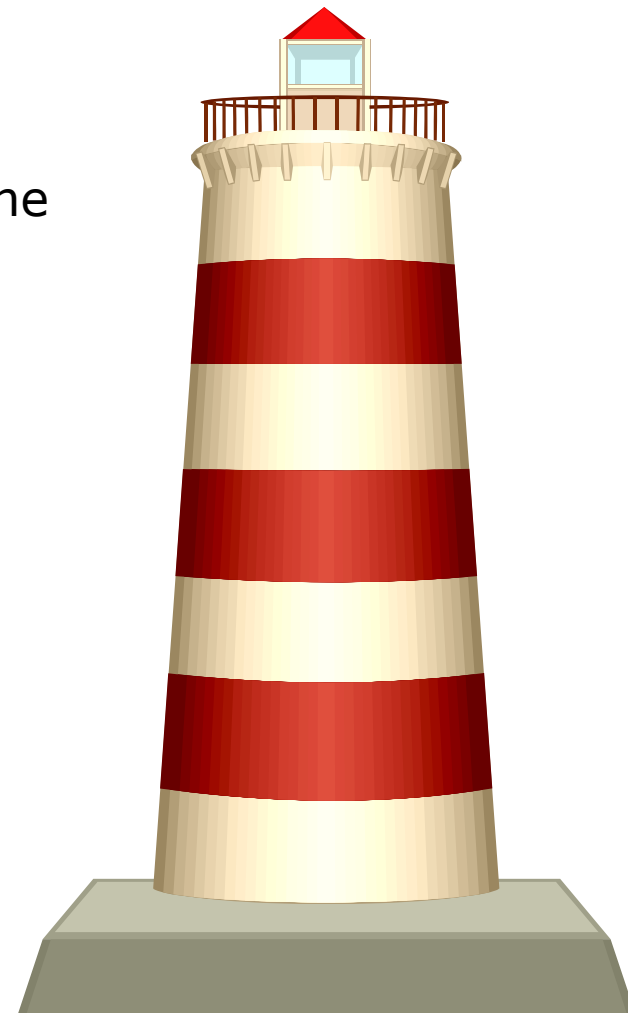
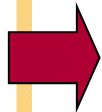
- Java, C++, etc
- Legacy Informationssysteme

■ Einsatzgebiete

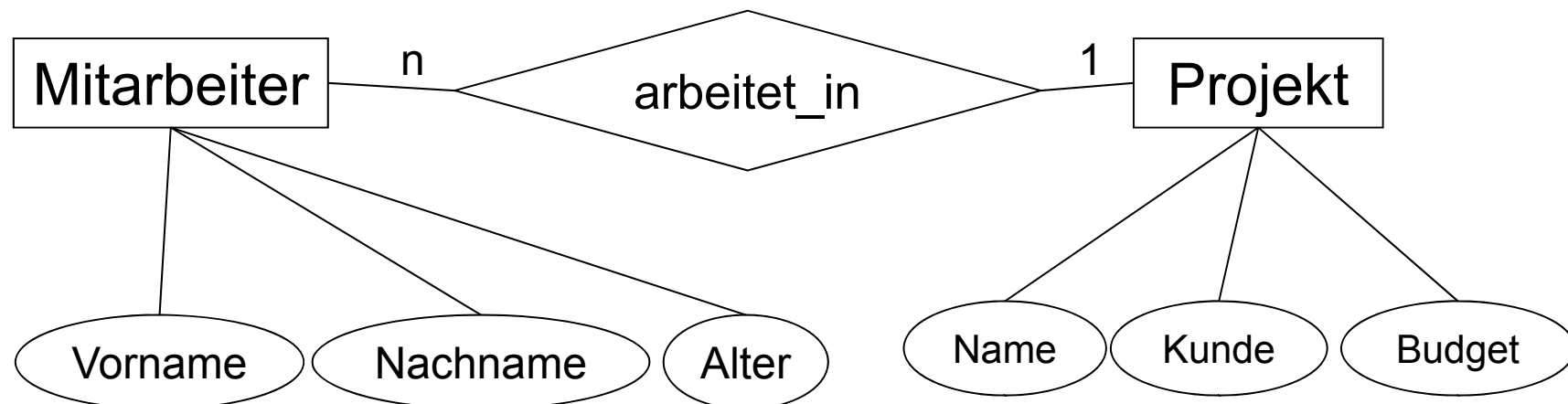
- Komplexe Analysen (Data Mining, Statistik)

Dahinter: Fast immer ein DBMS!

- Vorstellung der Arbeitsgruppe
- Organisatorisches
- Datenbanken und Informationssysteme
- Das Semester an einem Beispiel
- Ausblick auf das Semester



Mitarbeiter und Projekte



Relationales Modell – Relationen

42

mitarbeiter				
p_id	vorname	nachname	alter	proj_id

```
CREATE TABLE mitarbeiter(
    p_id          INTEGER,
    vorname       CHAR(25),
    nachname      CHAR(50),
    alter         INTEGER,
    proj_id       INTEGER,
    PRIMARY KEY(p_id),
    FOREIGN KEY proj_id
    REFERENCES projekte )
```

projekte			
proj_id	name	kunde	budget

Relationales Modell – Extension

43

mitarbeiter

p_id	vorname	nachname	alter	proj_id
1	Peter	Müller	32	1
2	Stefanie	Meier	34	1
5	Petra	Weger	28	2
7	Andreas	Zwickel	44	5
...	

projekte

proj_id	name	kunde	budget
1	DWH	BMW	400000
2	eCommerce	Metro	20000
5	SAP	RAG	50000
...

- Datentypen
- Schlüssel
- Fremdschlüssel
- Rechte
- Transaktionen

mitarbeiter

p_id	vorname	nachname	alter	proj_id
1	Peter	Müller	32	1
2	Stefanie	Meier	34	1
5	Petra	Weger	28	2
7	Andreas	Zwickel	44	5
...	

projekte

proj_id	name	kunde	budget
1	DWH	BMW	400000
2	eCommerce	Metro	20000
5	SAP	RAG	50000
...

Anfragen

45

- Deklarativ:
 - Nicht „Wie erzeuge ich ein Anfrageergebnis?“
 - Sondern „Was soll im Anfrageergebnis stehen?“
- Sprachlich
 - Nachnamen aller Personen, die in kleinen Projekten arbeiten
- Relationale Algebra
 - $\Pi_{m.nachn.} \sigma_{p.budget < 40000} (\sigma_{m.proj_id = p.proj_id}(\text{mit. } x \text{ proj.}))$
- SQL
 - **SELECT** m.nachname
FROM Mitarbeiter m, Projekte p
WHERE p.budget < 40000
AND m.proj_id = p.proj_id

Anfragebearbeitung – Problem

46

Anfragen

- Anfragen sind deklarativ.
- Anfragen müssen in eine ausführbare (prozedurale) Form transformiert werden.

Ziele

- QEP – prozeduraler Query Execution Plan
- Optimierung (Effizienz)
 - Schnell
 - Wenig Ressourcenverbrauch (CPU, I/O, RAM, Bandbreite)
 - Green IT: Stromverbrauch senken.

Anfragebearbeitung

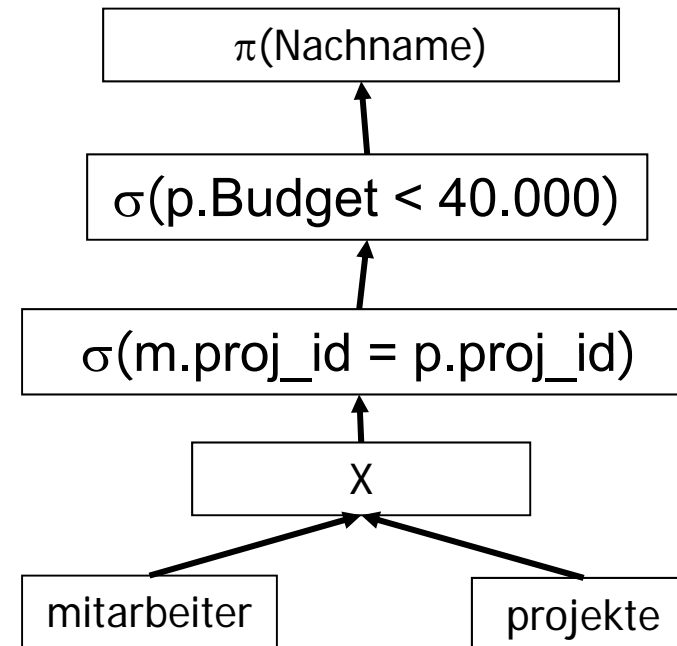
47

- Genereller Ablauf
 1. Parsen der Anfrage (Syntax)
 2. Überprüfen der Elemente (Semantik)
 3. Berechnung von möglichen Ausführungsplänen
 - ◇ Exponentiell viele
 4. Wahl des optimalen Ausführungsplans
 - ◇ Regelbasierter Optimierer
 - ◇ Kostenbasierter Optimierer
 5. Anfrageausführung

Anfragebearbeitung – Ausführung

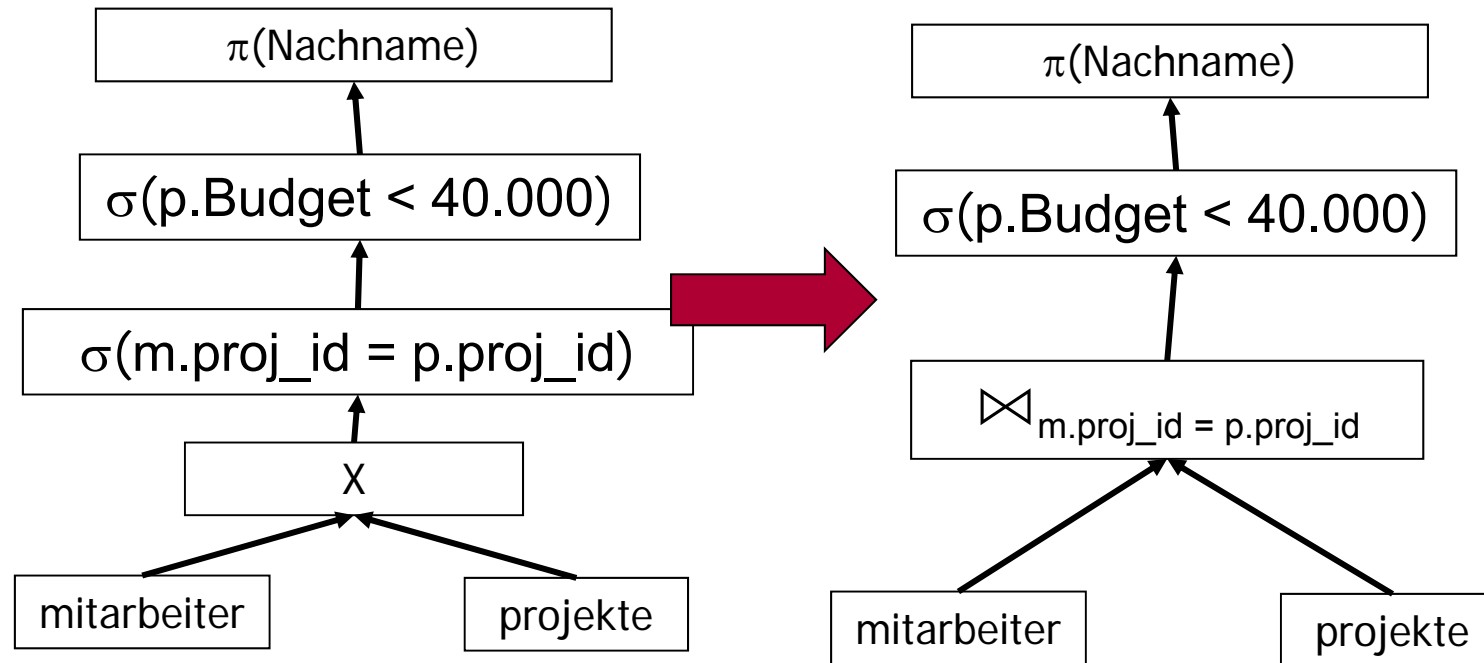
48

- Interpretation des Ausführungsplans als Baum
 1. Kartesisches Produkt
 2. Zwei Selektionen
 3. Projektion
- Sehr aufwendig!



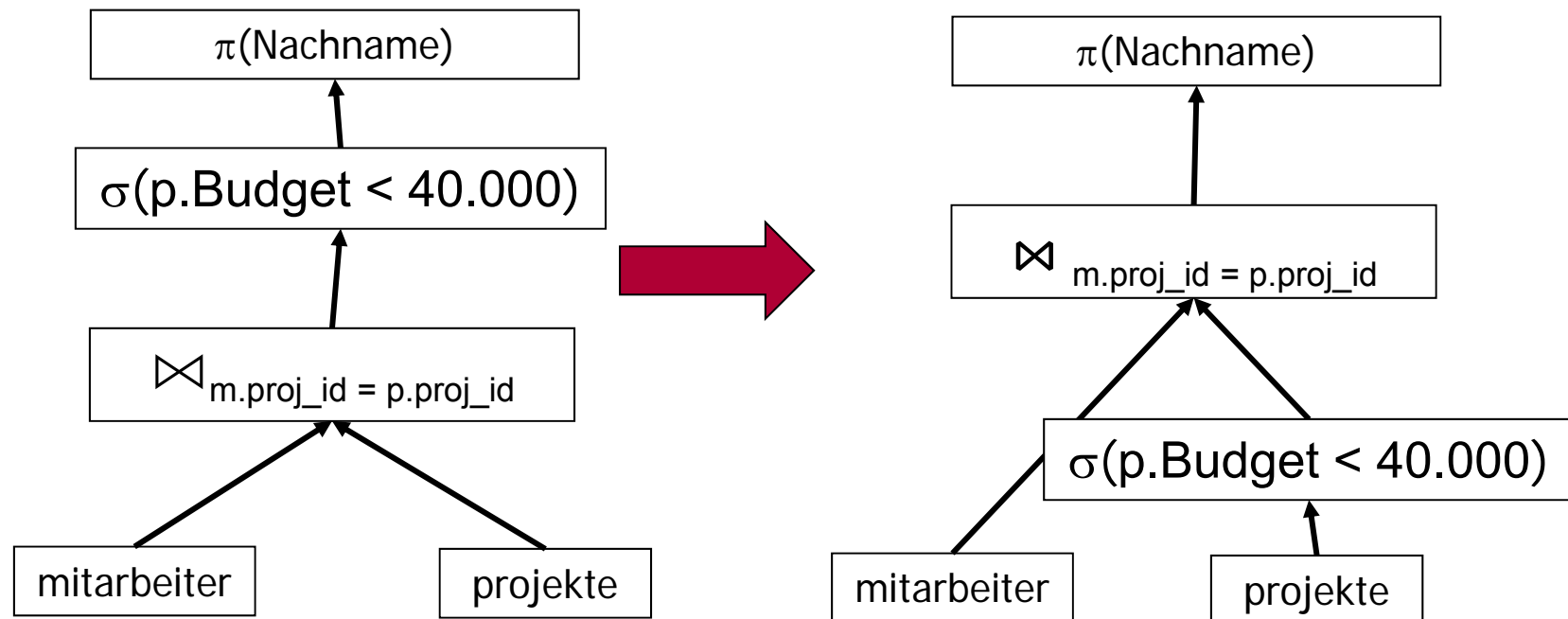
Anfragebearbeitung – Beispiel

49



Anfragebearbeitung – Beispiel

50



Sichten

51

- **CREATE VIEW** MA_MiniProjekte AS (
 SELECT m.nachname
 FROM Mitarbeiter m, Projekte p
 WHERE p.budget < 40000
 AND m.proj_id = p.proj_id)

- **SELECT * FROM** MA_MiniProjekte
- Probleme:
 - Anfrageplanung
 - Materialisierung
 - Updates durch Sichten hindurch

Transaktionsmanagement

52

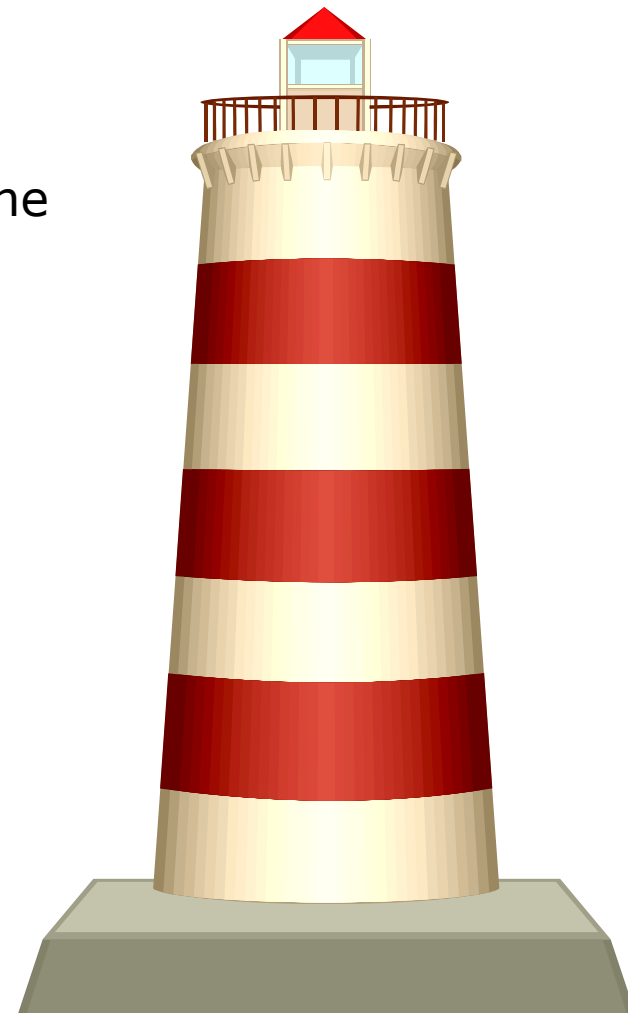
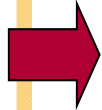
- Problem: Mehrbenutzerbetrieb
 - **INSERT INTO** mitarbeiter **VALUES** (Hans, Müller, 36, 5)
 - **DELETE FROM** projekte **WHERE** proj_id = 5

- Reihenfolge der Operationen ist nicht unabhängig
 - Serialisierbarkeit
 - Sperren

- XML und XML Schema
- Speicherung datenzentrierter XML-Dokumente
- Speicherung dokumentzentrierter XML-Dokumenten
- Speicherung von XML in bestehende Datenbank
- XPath und XQuery
- Indizierung



- Vorstellung der Arbeitsgruppe
- Organisatorisches
- Datenbanken und Informationssysteme
- Das Semester an einem Beispiel
- Ausblick auf das Semester



Kurzüberblick

55

- Einführung & Beispiel
 - ER-Modellierung →
 - Datenbanksysteme - Historie
 - Relationaler Datenbankentwurf →
 - Relationale Algebra →
 - SQL →
 - Anfragebearbeitung und -optimierung →
 - Transaktionsmanagement →
 - XML
 - Data Warehouses
- Übung 1: ER-Modellierung
 - Übung 2: Datenbankentwurf
 - Übung 3: Relationale Algebra
 - Übung 4: SQL
 - Übung 5: JDBC
 - Übung 6: Transaktionen
 - Übung: Klausurvorbereitung

Die Themen im Einzelnen

56

- Was sind Datenbanken?
 - Motivation, Historie, Datenunabhängigkeit, Einsatzgebiete
- Datenbankentwurf im ER-Modell
 - Entities, Relationships, Kardinalitäten, Diagramme
- Relationaler Datenbankentwurf
 - Relationales Modell, ER -> Relational, Normalformen, Transformationseigenschaften
- Relationale Algebra
 - Kriterien für Anfragesprachen, Operatoren, Transformationen
- SQL
 - SQL DDL, SQL DML, SELECT ... FROM ... WHERE ...

Die Themen

57

- Integrität und Trigger
 - Fremdschlüssel, weitere Bedingungen, Trigger
- Datenbankprogrammierung
 - JDBC
- Sichten und Zugriffskontrolle
 - Sichtenkonzept, Änderungen auf Sichten, Rechtevergabe
- Anfragebearbeitung und -optimierung
 - Anfragebäume, Kostenmodell, Transformationen, Dynamische Programmierung
- Transaktionsverwaltung
 - Mehrbenutzerbetrieb, Serialisierbarkeit, Sperrprotokolle, Fehlerbehandlung, Isolationsebenen in SQL
- XML
- Ausblicke
 - Data Warehouses, Data Mining, Informationsintegration, DBS II

Feedback

58

Fragen bitte jederzeit!

- In der VL
- Sprechstunde: Dienstags 15 - 16 Uhr
- Email: naumann@hpi.uni-potsdam.de
- Telefon: (0331) 5509 280

Anregungen zur Verbesserung!

The end