



Einführung ins Modul

Technische Informatik
Bachelor

Datenbanksysteme

- Organisation der Veranstaltung
 - Vorlesungen und Übungen
 - Aufgaben
 - Klausur
- Kommunikation – Moodle
- Inhalte
- Literatur





Organisation

- Übungen wöchentlich: 2 Termine mit gleichen Inhalten
- 4 bewertete Übungen (gesamt 30 Punkte)
 - Gruppenarbeit (max. 4 Personen) bis Semesterende !!!

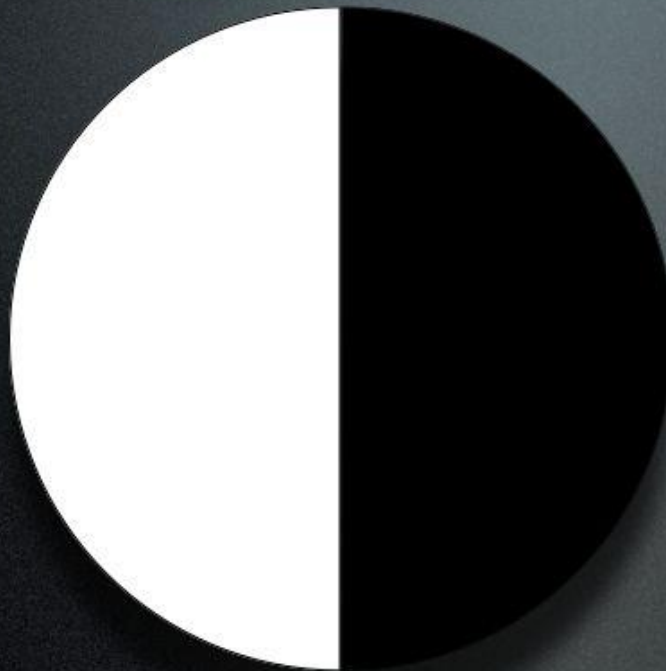
Klausur mit insgesamt 70 Punkten


- Die Gesamtnote resultiert aus den **insgesamt** erreichten Punkten (aus Klausur **und** den Übungen).
- **Übung erfolgreich**: bei min. **15 Punkten** erreicht





Your options according to Yoda.



-  Do.
-  Do not.
-  Try.

<http://imgur.com/fmGNs>





Termine – Aufgabenabgaben und Klausur:

- Übungen ohne Abgaben in den ersten 4 Wochen → SQL- und Datenbank-Training
- In ca. 3 Wochen 1. Aufgabe „Modell & Anfragen“
- Ca. 3 Wochen später 2. Aufgabe „SQL Anfragen“
- Ca. 6 Wochen später 3. Aufgabe „Datenintegration“
- Ca. 11 Wochen später 4. Aufgabe „Analyse und Wert“
- In ca. 15 Wochen Klausur → Termin tba





- Ankündigungen, Unterlagen, Übungsabgaben
 - Anmelden unter: <http://lms.beuth-hochschule.de>
 - Fachbereich VI → Technische Informatik Bachelor → Datenbanksysteme
 - Bitte Eintrag mit Nachname, Vorname, da Punkte- und Notenlisten daraus erzeugt werden
 - Aktivierung im gesamten Semester
 - Nach Nachklausur evtl. Deaktivierung
 - Kursname: <https://lms.beuth-hochschule.de/moodle/course/view.php?id=11612>
Passwort: dbs17! (Das Ausrufezeichen gehört dazu!)



Sprechzeit

- Nach Vereinbarung

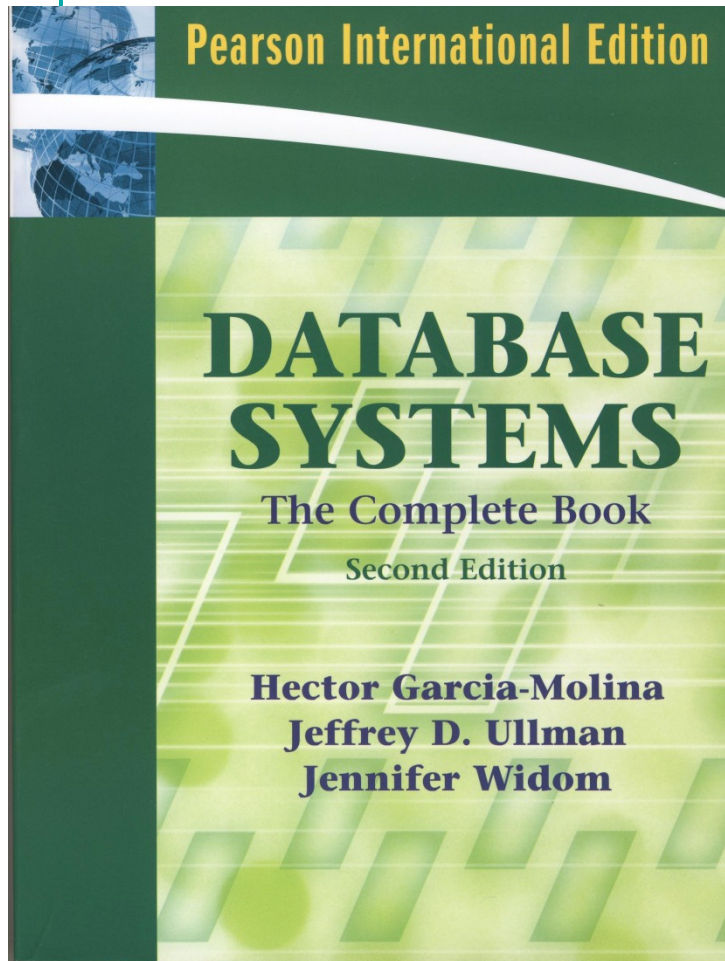
Mail

- aeverding@beuth-hochschule.de
- GnuPG-Key ID AC3FDFD4

Fingerprint 3811 DC50 C4D6 82A2 E78B

1119 C276 3073 AC3F DFD4





ISBN-13: 978-0-13-135428-9

ISBN-10: 0-13-135428-0

Database Systems –
The Complete Book

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

Viele Exemplare in Bibliothek

Mehrere Exemplare:

Universitätsbibliothek
im VOLKSWAGENHAUS
Fasanenstr. 88
10623 Berlin



Auch empfehlenswert:

- /EN02/ Elmasri, R. / Navathe, S.: Grundlagen von Datenbanksystemen. Ausgabe Grundstudium, Pearson Studium, 4. Auflage, 2008
- /Kudr07 Kudraß, T. (Hrsg.): Taschenbuch Datenbanken. Fachbuchverlag Leipzig, 2007
- /SSH08/ Saake, G. / Sattler, K. / Heuer, A.: Datenbanken: Konzepte und Sprachen. MITP-Verlag, 5. Auflage, 2013
- /KE01/ Kemper, A. / Eickler A.: Datenbanksysteme: Eine Einführung. 2013.



- Prof. Felix Naumann, HPI:
<http://www.tele-task.de/tagging/videolist/70/series/>
- Prof. Jens Dittrich, Uni Saarbrücken:
<https://www.youtube.com/user/jensdit>





- In jedem Vorlesungsblock werden Sie eine Multiple Choice Aufgabe entwerfen.
 - Von den Klausurpunkten werden auch Punkte aus Multiple Choice Aufgaben stammen, die Sie entworfen haben.
- Bitte nehmen Sie ein Blatt und entwerfen Sie eine multiple Choice Aufgabe mit einer Frage und 3 Antworten
- Davon sollte mindestens eine korrekt mindestens eine falsch sein.
- Im Anschluss tauschen Sie das Blatt mit Ihrem Nachbarn
 - dieser soll die richtigen Aussagen markieren
 - Diskutieren Sie über die Lösungen
- Am Ende der Vorlesungen sammle ich die MC Aufgaben ein.





IHRE AUFGABE FÜR DAS SEMESTER





- Siehe http://en.wikipedia.org/wiki/AOL_search_data_leak
- Weiterführendes Material im Ordner AOL_Search

Overview [\[edit source\]](#)

On August 4, 2006, AOL Research, headed by Dr. Abdur Chowdhury, released a compressed text file on one of its websites containing twenty million search [keywords](#) for over 650,000 users over a 3-month period, intended for research purposes. AOL pulled the file from public access by the 7th, but not before it had been mirrored and distributed on the Internet.

AOL themselves did not identify users in the report; however, [personally identifiable information](#) was present in many of the queries, and as the queries were attributed by AOL to particular user accounts, identified numerically, an individual could be identified and matched to their account and search history by such information.^[1] *The New York Times* was able to locate an individual from the released and anonymized search records by cross referencing them with phonebook listings.^[2] Consequently, the ethical implications of using this data for research are under debate.^{[3][4]}

AOL acknowledged it was a mistake and removed the data. However, the removal was too late; the data was redistributed by others, and can still be downloaded from mirror sites.^{[5][6]}

In January 2007, Business 2.0 Magazine on CNNMoney ranked the release of the search data #57 in a segment called "101 Dumbest Moments in Business."^[7]

Die Daten



```
SELECT *
FROM `chunk`
WHERE querytext LIKE '%$ average %'
```

1 Result

2 Profiler

3 Messages

4 Table Data

5 Info

6 History

(Read Only)

Form

Grid

Limit rows

First row: 0

of rows: 1000

Refresh

uniqueID	AnonId	QueryText		Last_Changed_On	ItemRank	ClickURL
122026	294499	alaskas average temperature	27B	2006-05-26 17:57:31	0	
122027	294499	alaska's average temperature	28B	2006-05-26 17:57:44	1	http://www.alaska.com
122028	294499	alaska's average temperature in the winter	42B	2006-05-26 17:59:21	1	http://www.farmersalmanac.com
122029	294499	alasks average winter temperature	33B	2006-05-26 18:00:31	0	
122030	294499	alaska average winter temperature	33B	2006-05-26 18:00:35	1	http://www.farmersalmanac.com
122031	294499	alaska average winter temperature	33B	2006-05-26 18:00:35	4	http://www.alaska.com
122034	294499	what is the average tempture of alaska in the w...	52B	2006-05-26 18:03:54	0	
122035	294499	what is the average temperature of alaska in th...	55B	2006-05-26 18:03:58	1	http://www.alaska.com
122036	294499	what is the average temperature of alaska in th...	55B	2006-05-26 18:03:58	6	http://instaar.colorado.edu
122037	294499	what is the average temperature of alaska in th...	55B	2006-05-26 18:03:58	10	http://www.worldviewofglobalwarming.org
166588	393765	wal mart stock average price per share	38B	2006-04-01 08:57:13	0	
166589	393765	walmart stock average price per share	37B	2006-04-01 08:58:48	0	
270474	672368	what is the average merit raise	31B	2006-05-31 13:31:41	4	http://www.salary.com
360987	881500	what is the average cost of beating a dui wrap ...	72B	2006-03-02 15:15:36	0	
361235	881500	what is the average home price in ironwood mich...	51B	2006-04-28 22:10:04	2	http://www.city-data.com
361236	881500	what is the average home price in ironwood mich...	51B	2006-04-28 22:10:04	10	http://mattsonworks.com
361237	881500	what is the average home price in ironwood mich...	51B	2006-04-28 22:10:04	7	http://www.epodunk.com
361238	881500	what is the average home price in ironwood mich...	51B	2006-04-28 22:10:04	10	http://mattsonworks.com
361239	881500	what is the average home price in ironwood mich...	51B	2006-04-28 22:10:04	9	http://www.kristafracke.com
361240	881500	what is the average home price in ironwood mich...	51B	2006-04-28 22:10:04	8	http://www.neighborhoodscout.com
381526	934546	batting average report	22B	2006-05-31 22:13:37	1	http://www.baseballprospectus.com
381527	934546	little league batting average sheet	35B	2006-05-31 22:14:40	0	

- Dates: 01 March, 2006 - 31 May, 2006
- Normalized queries:
 - 36,389,567 lines of data
 - 21,011,340 instances of new queries (w/ or w/o click-through)
 - 7,887,022 requests for "next page" of results
 - 19,442,629 user click-through events
 - 16,946,938 queries w/o user click-through
 - 10,154,742 unique (normalized) queries
 - 657,426 unique user ID's
- http://www.gregsadetsky.com/aol-data/U500k_README.txt





Ihre Aufgaben (Details folgen)



■ **Aufgabe 1 Anfragen & Modellierung“**

Denken Sie mal darüber nach, welche Anfragen Sie an die AOL Daten stellen möchten. Bitte modellieren Sie ein logisches und physisches Schema zur Beantwortung dieser Anfragen.

■ **Aufgabe 2 „SQL und Abfrageausführung“**

Bitte formulieren Sie für Ihre Analyseideen aus 1.) die SQL Anfragen. Sie verstehen auch Möglichkeiten der Abfrageausführung bzw. Optimierung.

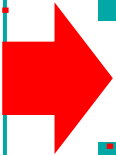
■ **Aufgabe 3 „Datenintegration“**

Zur Ausführung der Ausführung fehlen Ihnen noch externe Daten, z.B. aus dem Internet Archive, DMOZ oder Freebase.org. Bitte ergänzen Sie Ihr Schema und die Datenbasis.

■ **Aufgabe 4 „Analyse, Erkenntnisgewinn und Wert“**

Stellen Sie in 10 Minuten die wichtigsten Erkenntnisse aus den Daten vor. Bewerten Sie den Erkenntnisgewinn, z.B. gegenüber Ihren Kommilitonen oder der Literatur! Welche Erkenntnisse hätten einen kommerziellen Wert?



- 
- Was sind Datenbanken?
 - Motivation, Historie, Datenunabhängigkeit, Einsatzgebiete
 - Datenbankentwurf im ER-Modell & Relationaler Datenbankentwurf
 - Entities, Relationships, Kardinalitäten, Diagramme
 - Relationales Modell, ER -> Relational, Transformationseigenschaften
 - Relationale Algebra & SQL
 - Kriterien für Anfragesprachen, Operatoren, Transformationen
 - SQL DDL, SQL DML, SELECT ... FROM ... WHERE ...
 - Optimierungen
 - Datenintegration & Transaktionsverwaltung
 - Zugriff von Java: JDBC, Cursor,
 - Integration: Normalformen, Textdaten, ETL
 - Transaktionen: Mehrbenutzerbetrieb, Serialisierbarkeit, Sperrprotokolle, Fehlerbehandlung, Isolationsebenen in SQL
 - Ausblick
 - Wert von Daten / Technikfolgenabschätzung / Ethik





INFORMATIONSMANAGEMENT UND DATENBANKEN





- Digitale Repräsentation von
 - Dingen
 - Entitäten
 - Wissen
 - Information
- in der wirklichen Welt.

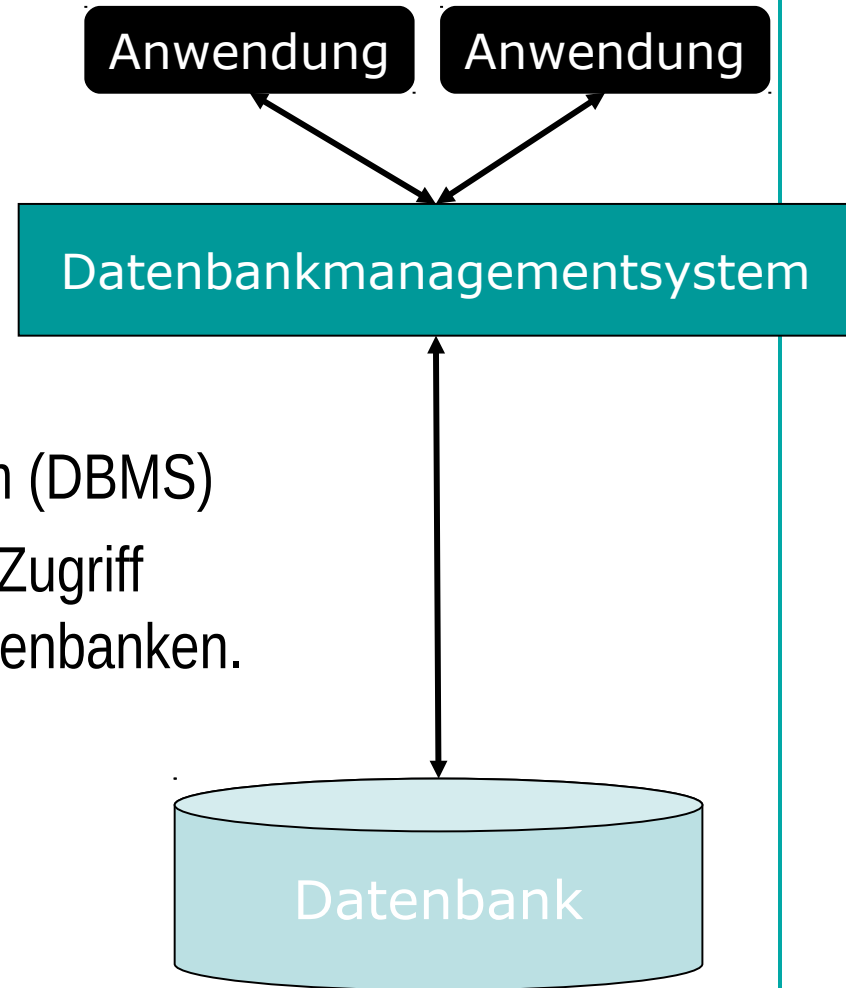
- Kernfragen:
 - Welche Daten speichere ich?
 - Wie speichere ich die Daten?
 - Wie frage ich Daten ab?
 - Wie erledige ich all dies effizient und sicher?

} Unterstützung
durch
Datenbanksysteme



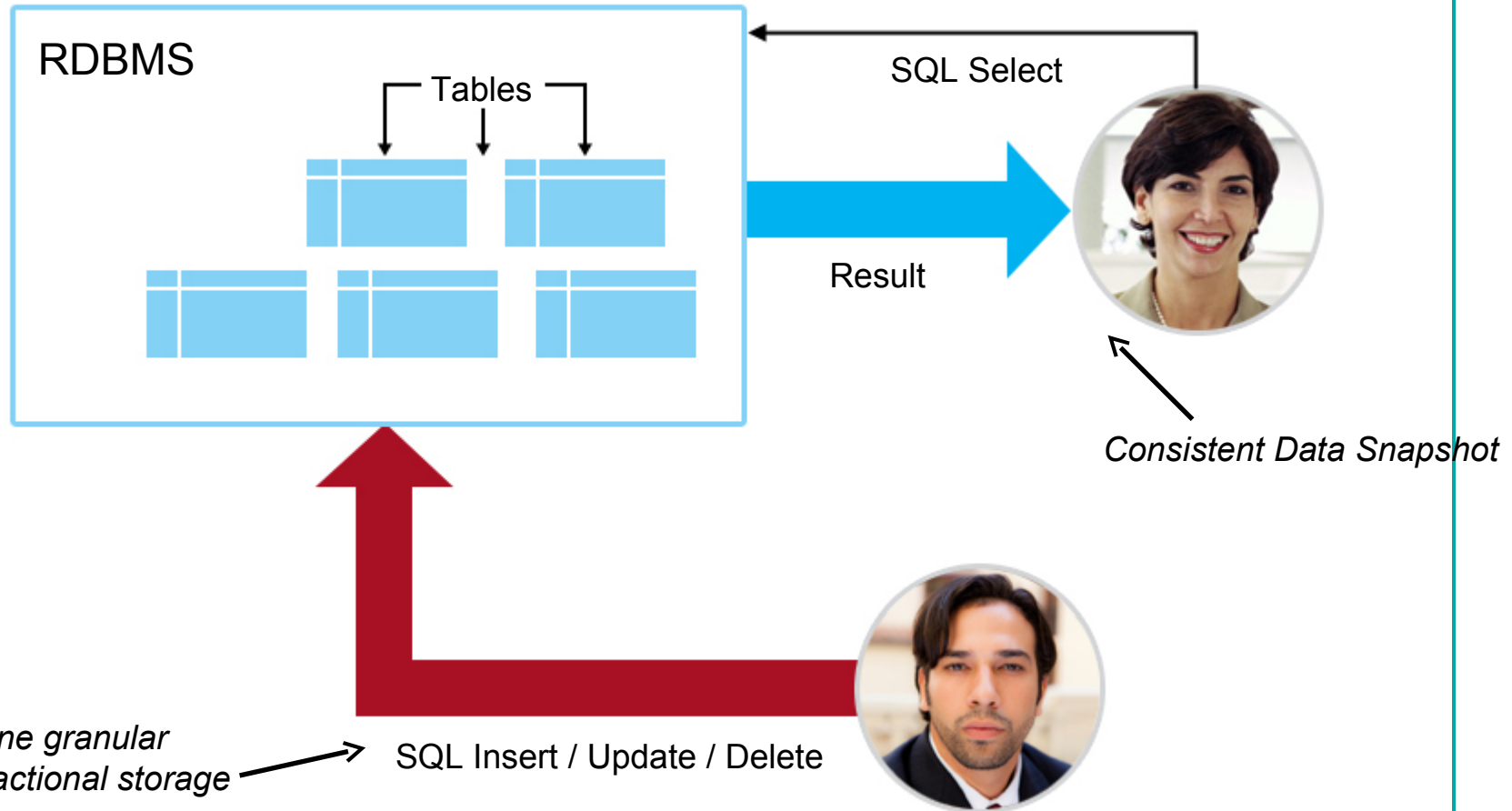


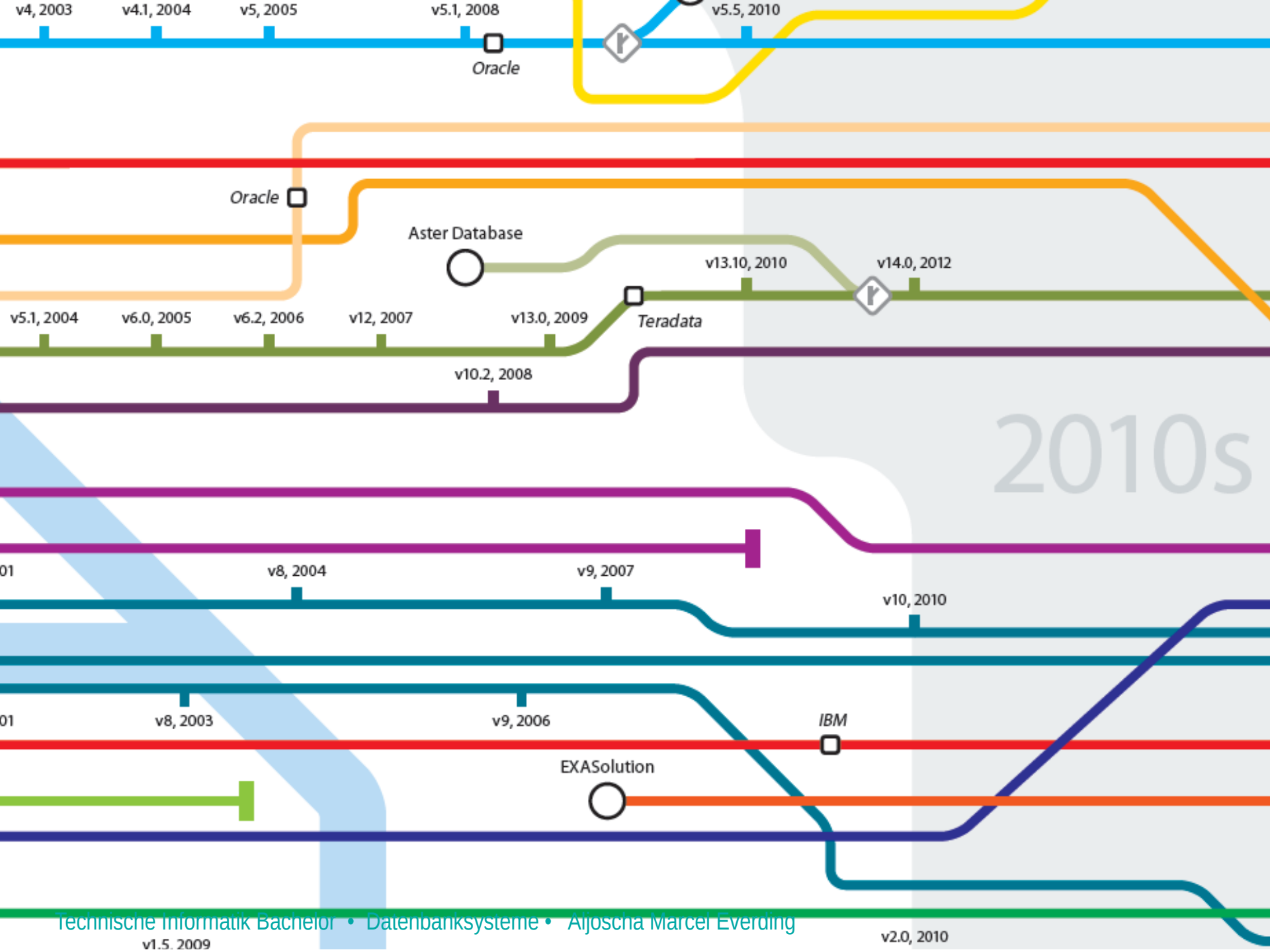
- Bestandteile
 - Datenbank
 - Die Daten selbst
 - + Metadaten
(Beschreibung der Daten)
 - Datenbankmanagementsystem (DBMS)
 - Softwarekomponente zum Zugriff
auf eine oder mehrere Datenbanken.
 - Server-basiert
 - (Anwendungen)



- 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

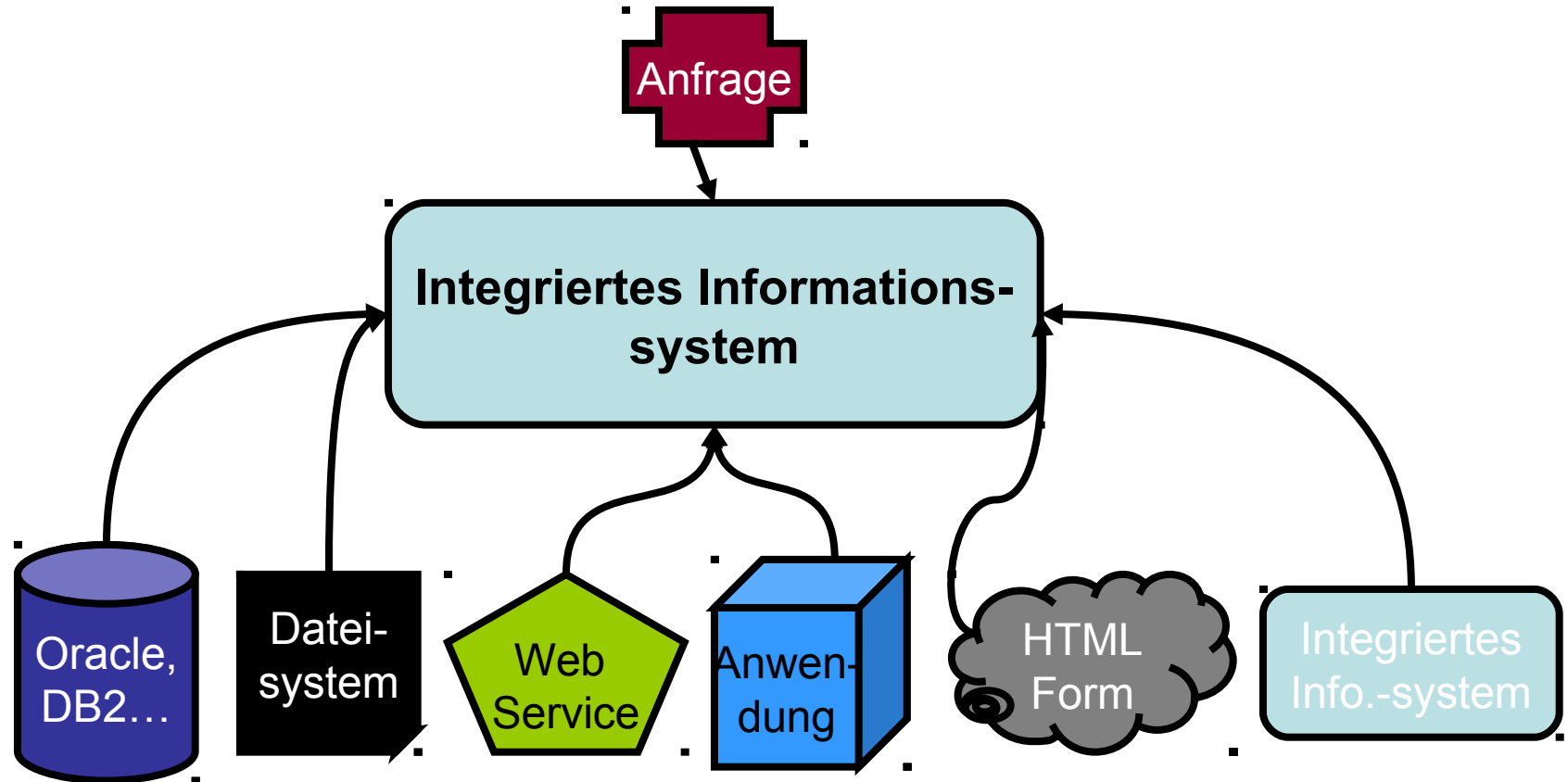






- DBMS
 - OLTP (Online Transaction Processing)
 - Oracle, DB2, SQL-Server
 - Informix, Sybase
 - Teradata
 - PostgreSQL, Interbase
 - MySQL, ...
 - SQLite
 - Apache Derby, HSQLDB
 - ...
- Alle großen Softwaresysteme nutzen ein, mehrere oder sehr viele DBMS.
 - SAP
 - Siebel
 - SABRE
 - AMADEUS
 - SWIFT
 - ...







Wer? Und wieviele Daten?



processes 20 PB a day (2008)



9 PB of user data +
>50 TB/day (11/2011)



36 PB of user data +
80-90 TB/day
(6/2010)



S3: 449B objects, peak
290k request/second
(7/2011)



LHC: ~15 PB a year
(at full capacity)

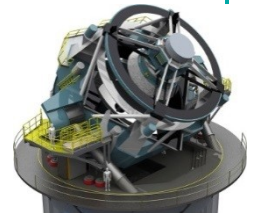
640K ought to
be enough for
anybody.



150 PB on 50k+ servers
running 15k apps



Wayback Machine: 3
PB + 100 TB/month
(3/2009)



LSST: 6-10 PB a year
(~2015)

Source: Jimmy Lin 2011





Das Rechenzentrum ist der Computer



I think there is a world market for about five computers.

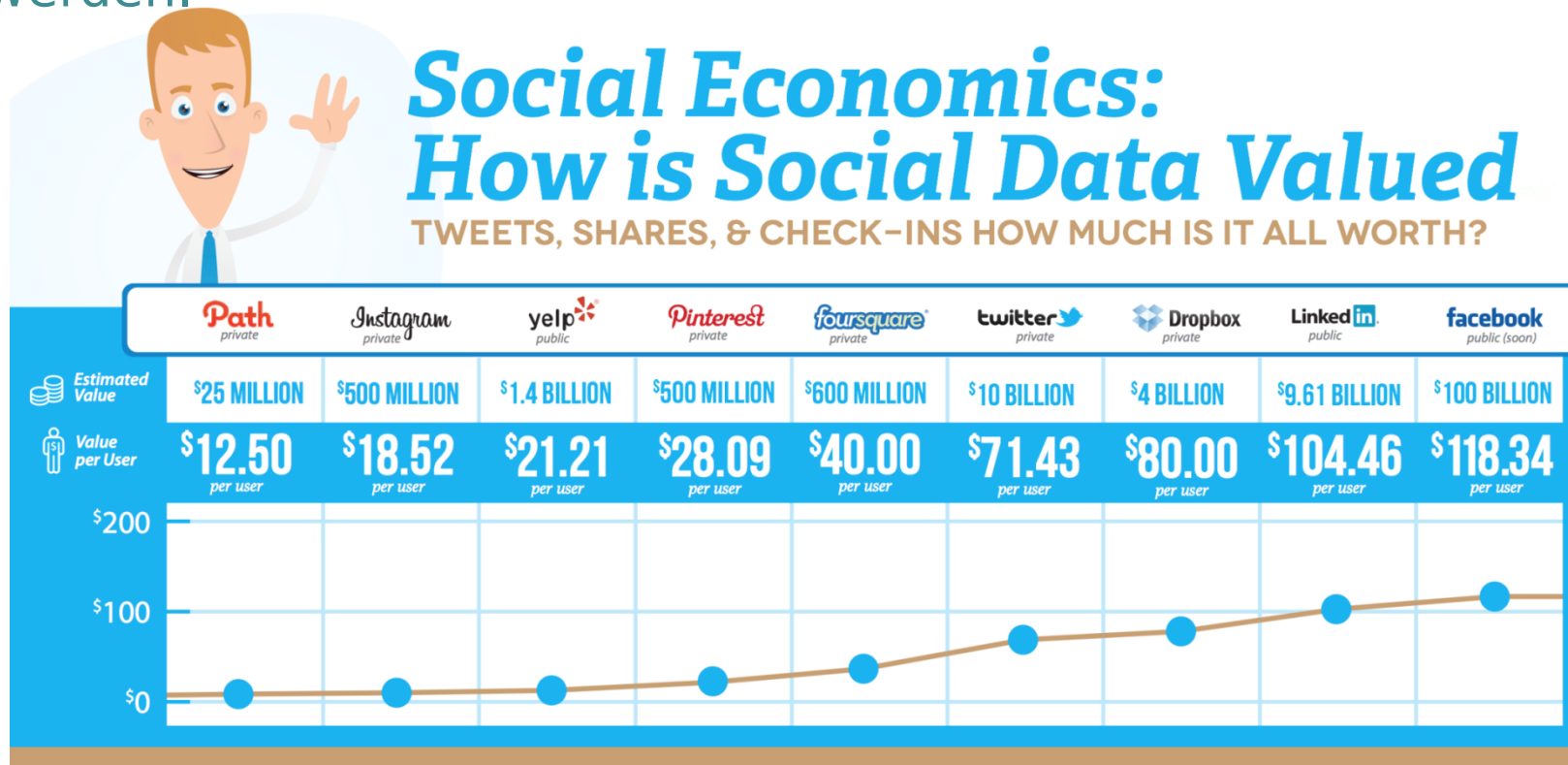




Datenwert



Menschen beschreiben Gefühle und Beobachtungen, Kaufintentionen, Buzz etc. in textueller Form, durch Ihre "Klicks" oder Transaktionen. Diese Daten sind für häufig exklusiv nur dem Erheber zugänglich. Sie werden besonders wertvoll wenn Sie mit weiteren exklusiven Daten gekreuzt werden.

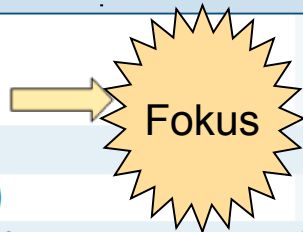


<http://www.readwriteweb.com/enterprise/2012/04/infographic-how-much-is-that-t.php>

Last visited 6/4/2012

CIOs investieren in ‚Intelligence‘, ‚Analytics‘ und ‚Mobile Technologies‘.

CIO technologies	Ranking of technologies CIOs selected as one of their top 3 priorities in 2012				
Ranking	2012	2011	2010	2009	2008
Analytics and business intelligence	1	5	5	1	1
Mobile technologies	2	3	6	12	12
Cloud computing (SaaS, IaaS, PaaS)	3	1	2	16	*
Collaboration technologies (workflow)	4	8	11	5	8
Virtualization	5	2	1	3	3
Legacy modernization	6	7	15	4	4
IT management	7	4	10	*	*
Customer relationship management	8	18	*	*	*
ERP applications	9	13	14	2	2
Security	10	12	9	8	5
Social media/Web 2.0	11	10	3	15	15



*Not an option that year

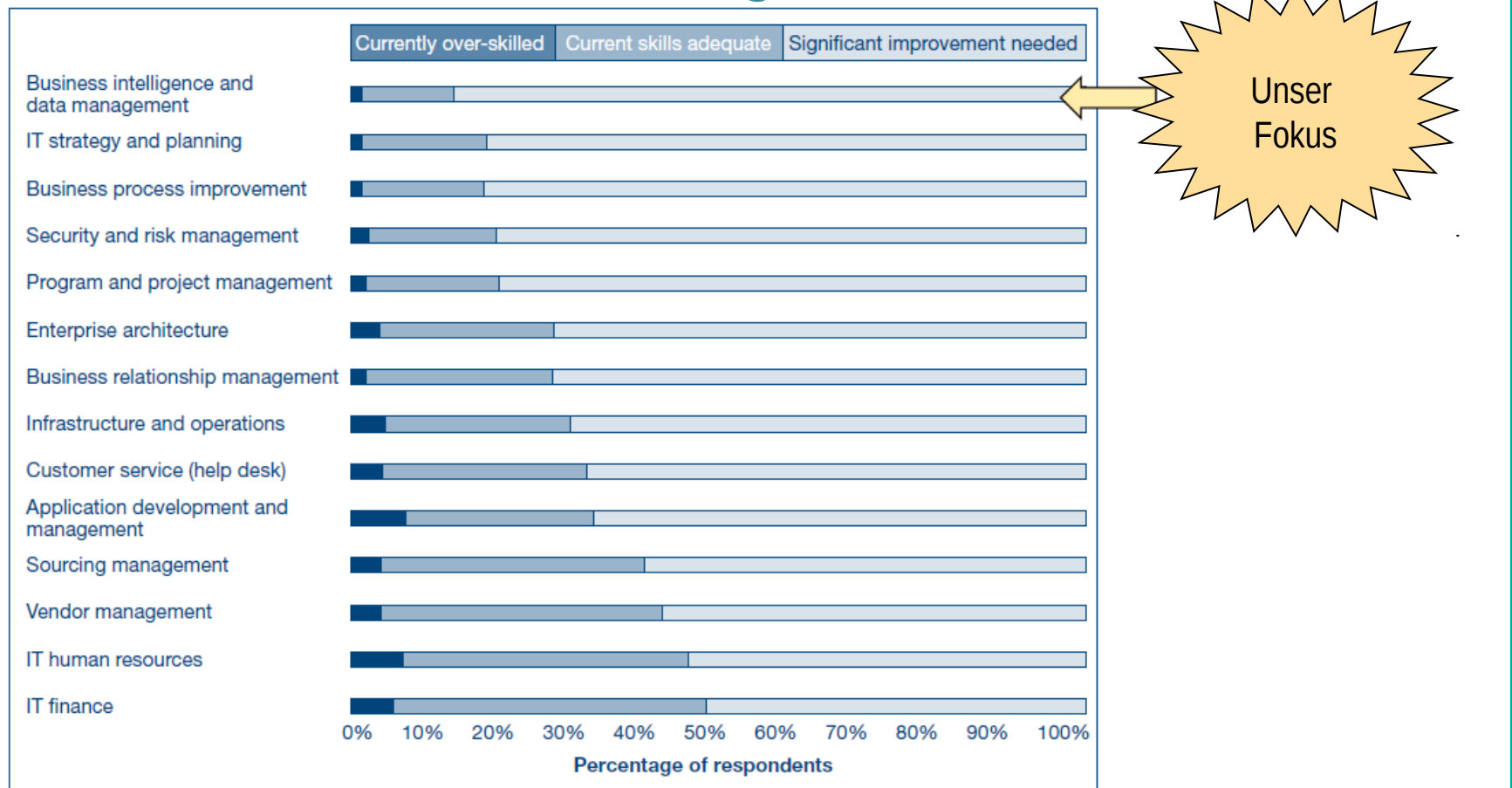
The 2012 Gartner CIO Agenda Report,
Page 3 http://imagesrv.gartner.com/cio/pdf/cio_agenda_insights.pdf



Wanted: Data-driven Entrepreneurs



Allein bis zum Jahre 2018 benötigt die USA 190,000 Wirtschaftsinformatiker mit analytischen Fähigkeiten und betriebswirtschaftlichem Hintergrund.



The 2012 Gartner CIO Agenda Report, Page 5 http://imagesrv.gartner.com/cio/pdf/cio_agenda_insights.pdf
Jacques Bughin et.al. McKinsey Quarterly: Sizing the potential of 'Big Data'. October 2011

Welche Skills benötigt die Industrie? SQL, R, Datenbereinigung ...

		Analytics	Biology	Datamart	Finance	Finance	Healthcare	Healthcare	Healthcare	Insurance	Marketing	News Aggre	Retail	Retail	Social New	Social New	Visualization	Web	Web	Analytics	Analytics	Analytics	Finance	Healthcare	Media	Retail	Finance	Insurance	Retail	Retail	Sports	Web Security		
Process	Discovery	Locating Data	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
		Field Definitions	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
	Wrangle	Data Integration	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
		Parsing Semi-Structured	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
		Advanced Aggregation and Filtering	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
	Profile	Data Quality	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
		Verifying Assumptions	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
	Model	Feature Selection	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
		Scale	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
		Advanced Analytics	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
Report	Communicating Assumptions						x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
	Static Reports		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
Workflow	Data Migration	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
	Operationalizing Workflows		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
Tools	Database	SQL	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
		Hadoop/Hive/Pig	x	x	x			x						x		x	x										x							
		MongoDB						x	x	x								x																
		CustomDB	x					x	x	x										x														
	Scripting	Java	x	x			x		x	x	x			x	x	x		x	x	x														
		Perl												x	x	x																		
		Python	x		x	x	x	x	x	x	x	x				x	x	x																
		Clojure																																
		Visual Basic			x																													
	Modeling	R	x		x											x	x	x			x	x												
		Matlab				x					x										x													
		SAS	x																															
		Excel		x	x	x	x									x	x																	

Fig. 1. Respondents, Challenges and Tools. The matrix displays interviewees (grouped by archetype and sector) and their corresponding challenges and tools. *Hackers* faced the most diverse set of challenges, corresponding to the diversity of their workflows and toolset. *Application users* and *scripters* typically relied on the IT team to perform certain tasks and therefore did not perceive them as challenges.

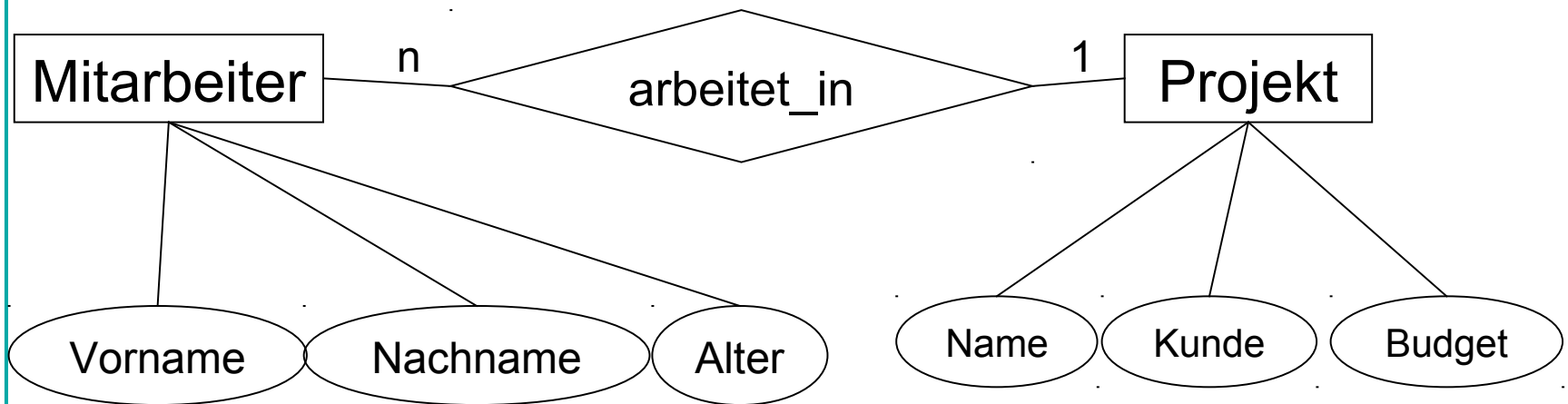
Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, Jeffrey Heer: Enterprise Data Analysis and Visualization: An Interview Study. IEEE Trans. Vis. Comput. Graph. 18(12): 2917-2926 (2012)
<http://db.cs.berkeley.edu/papers/vast12-interview.pdf>



DAS SEMESTER AN EINEM BEISPIEL



- Mitarbeiter und Projekte





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



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
...



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





- 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)





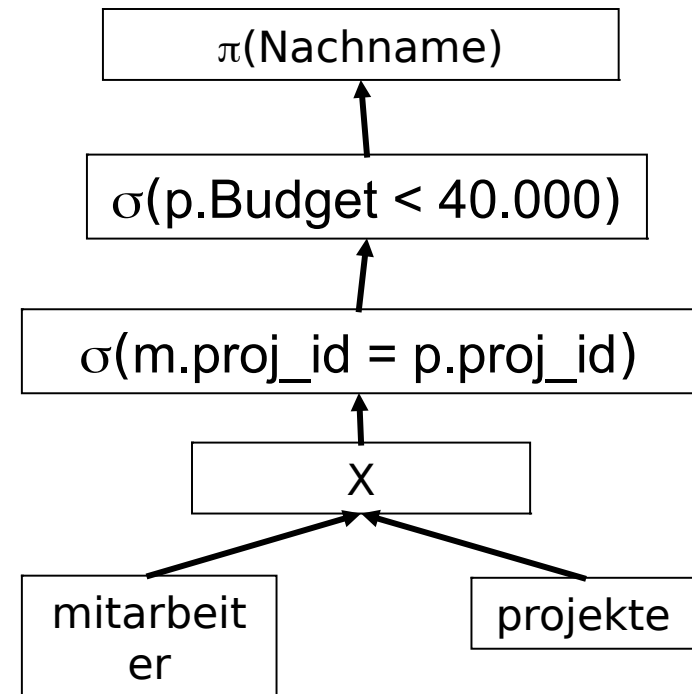
- Genereller Ablauf
  - Parsen der Anfrage (Syntax)
  - Überprüfen der Elemente (Semantik)
  - Berechnung von möglichen Ausführungsplänen
    - Exponentiell viele
  - Wahl des optimalen Ausführungsplans
    - Regelbasierter Optimierer
    - Kostenbasierter Optimierer
  - Anfrageausführung





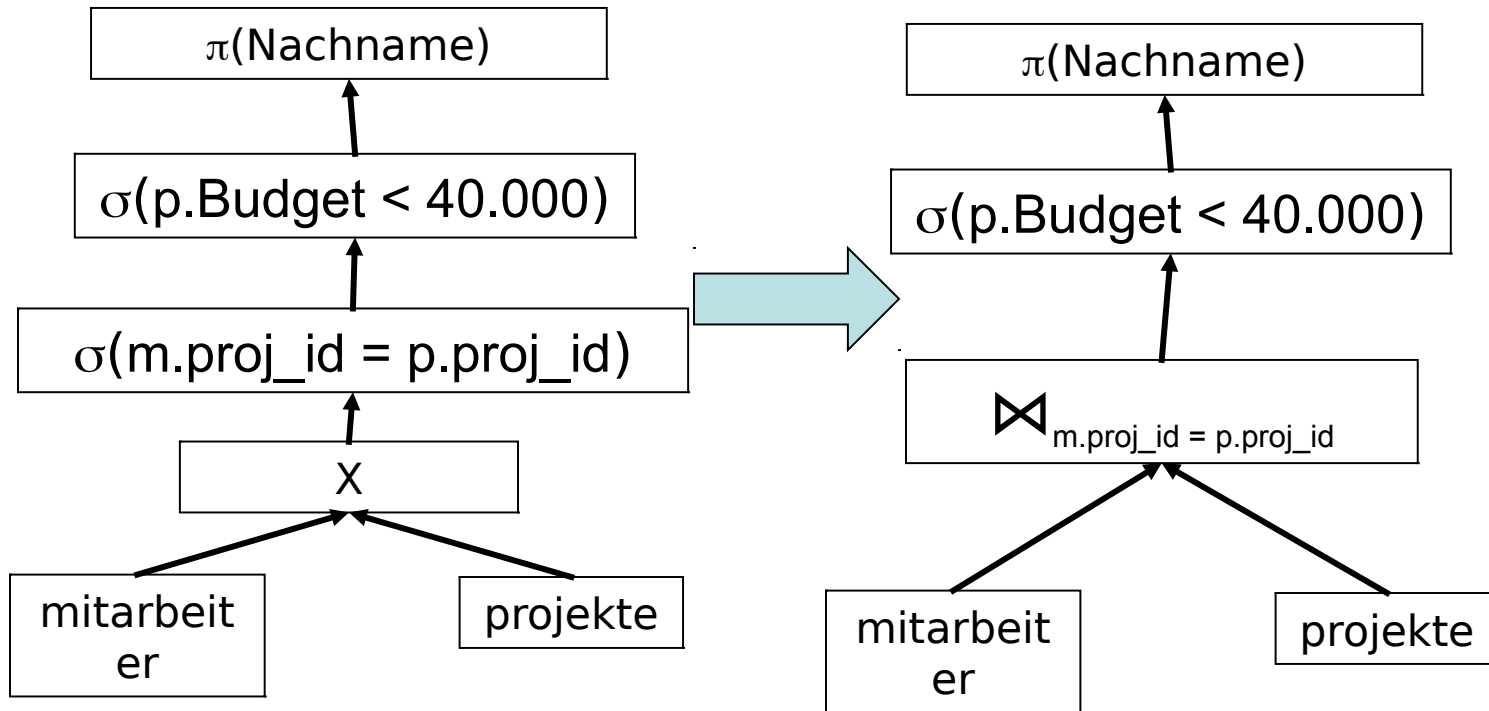


- Interpretation des Ausführungsplans als Baum
  - Kartesisches Produkt
  - Zwei Selektionen
  - Projektion
- Sehr aufwendig!



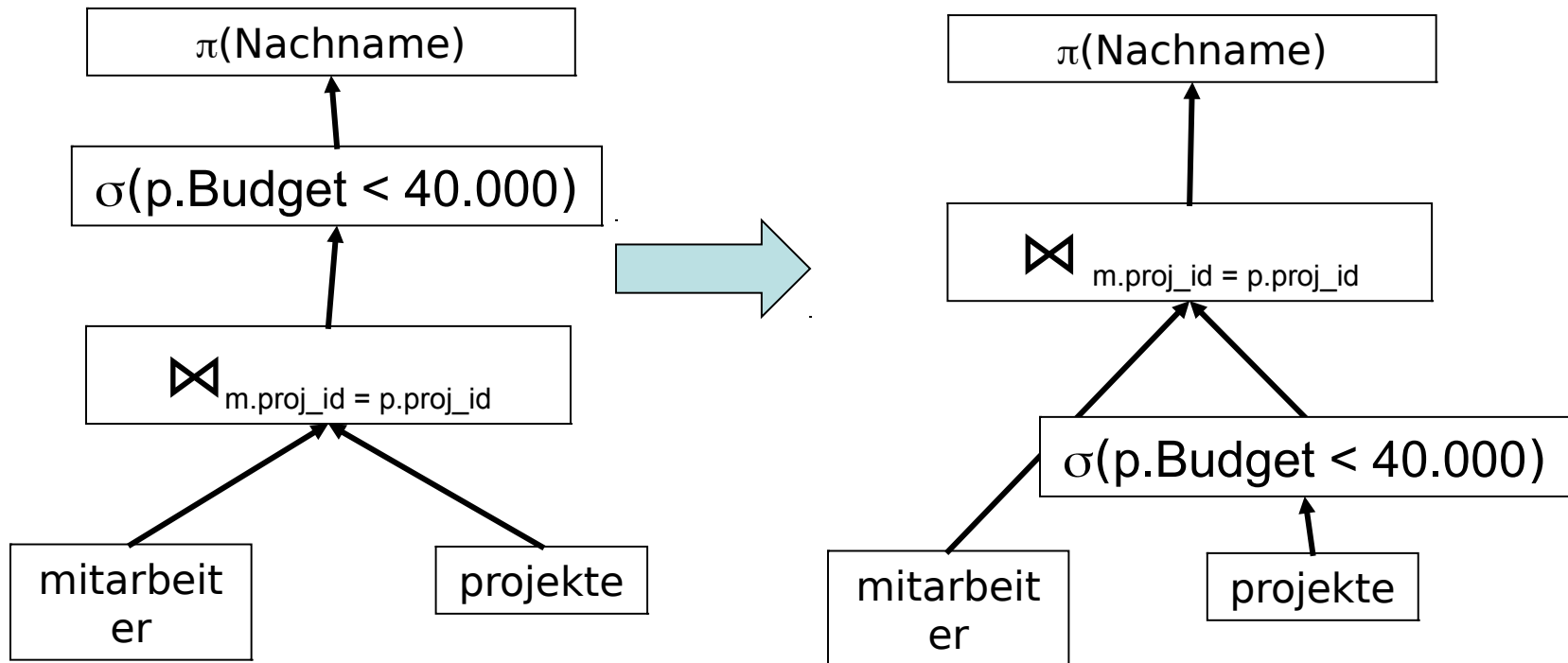


# Anfragenbearbeitung - Beispiel





# Anfragenbearbeitung - Beispiel



- ```
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



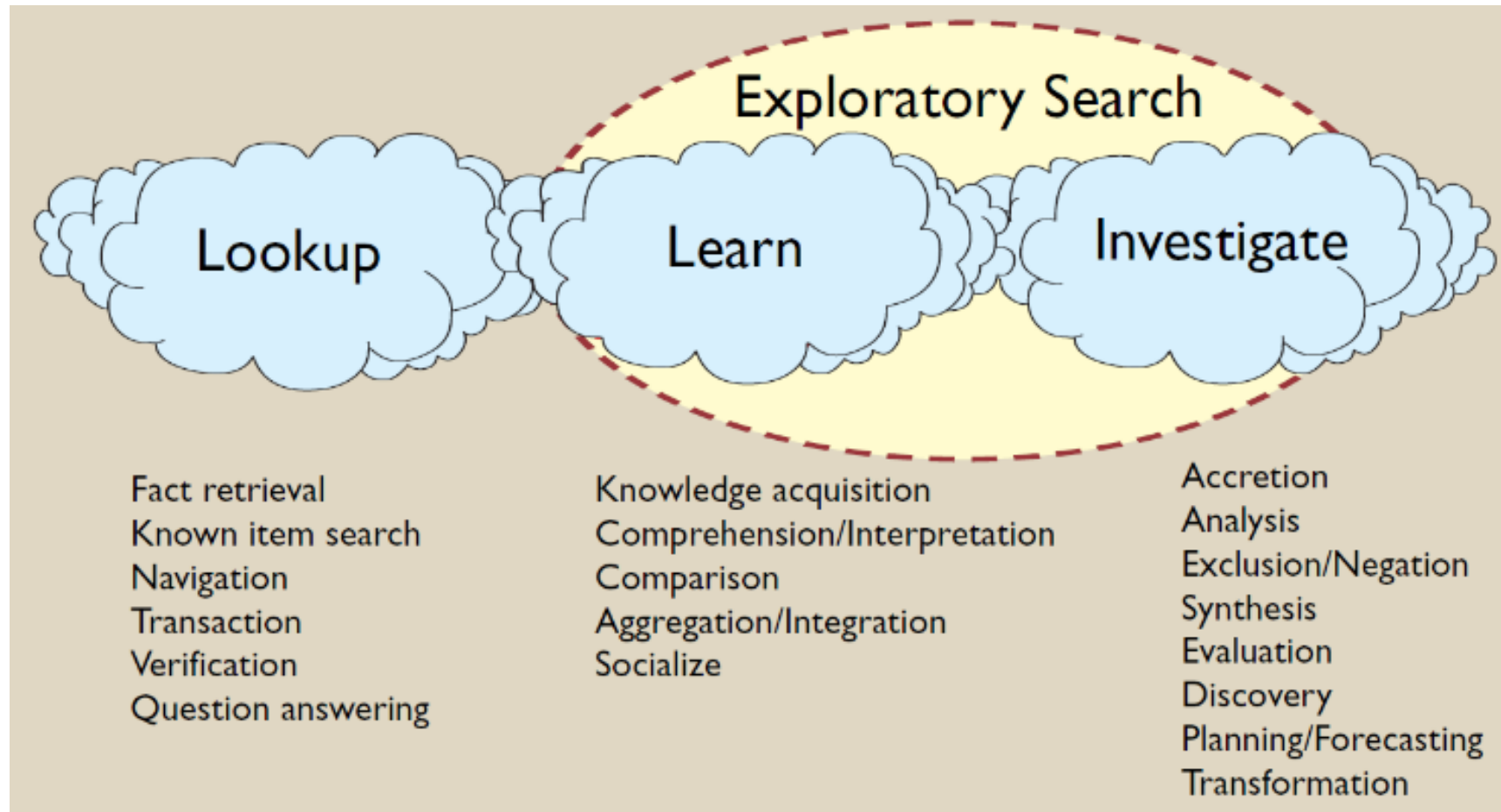


- Problem: Mehrbenutzerbetrieb
  - `INSERT INTO mitarbeiter VALUES (Hans, Müller, 36, 5)`
  - `DELETE FROM projekte WHERE proj_id = 5`
- Reihenfolge der Operationen ist nicht unabhängig
  - Serialisierbarkeit
  - Sperren





# Explorative Datenanalyse



Benjamin Bloom: Taxonomy of educational objectives: Handbook I: Cognitive Domain New York, Longmans, Green, 1956.

G. Marchionini. Exploratory search: from searching to understanding. Communications of the ACM, 49:41-46, Apr. 2006.



# ZUSAMMENFASSUNG





- Vorstellung DATEXIS.DE
- Organisatorisches
- Datenbanken und Informationssysteme
- „Das Semester“ an einem Beispiel
- Ausblick auf das Semester

In der nächsten Veranstaltung:

- ER-Modellierung  
(Kapitel 4 des Lehrbuches)
  - Motivation und Einbettung
  - Begriffe und Definitionen
  - ER-Diagramme
  - Modellierung von Nebenbedingungen
  - Schwache Entitytypen
  - Erweitertes ER-Modell
  - Designprinzipien
  - Sichtintegration







## ■ **Aufgabe 1 Anfragen & Modellierung“**

Denken Sie mal darüber nach, welche Anfragen Sie an die AOL Daten stellen möchten. Bitte Sie bitte ein logisches und physisches Schema zur Beantwortung dieser Anfragen.

## ■ **Aufgabe 2 „SQL und Abfrageausführung“**

Bitte formulieren Sie für Ihre Analyseideen aus 1.) die SQL Anfragen. Sie verstehen auch Möglichkeiten der Abfrageausführung bzw. Optimierung.

## ■ **Aufgabe 3 „Datenintegration“**

Zur Ausführung der Ausführung fehlen Ihnen noch externe Daten, z.B. aus dem Internet Archive, DMOZ oder Freebase.org. Bitte ergänzen Sie Ihr Schema und die Datenbasis.

## ■ **Aufgabe 4 „Analyse, Erkenntnisgewinn und Wert“**

Stellen Sie in 5 Minuten die wichtigsten Erkenntnisse aus den Daten vor. Bewerten Sie den Erkenntnisgewinn, z.B. gegenüber Ihren Kommilitonen oder der Literatur! Welche Erkenntnisse hätten einen kommerziellen Wert?