

Prof. Dr.-Ing. Dr. h. c. Theo Härder
Universität Kaiserslautern
Fachbereich Informatik
haerder@informatik.uni-kl.de
<http://www.haerder.de/>

Vorlesung

Realisierung von Datenbanksystemen

SS 2005

Theo Härder

Vorlesung:

Di., 11.45 - 13.15 Uhr, 13-305

Do., 11.45 - 13.15 Uhr, 46-210

Beginn: Di., 26. 04. 2005

Ziele

- **Vermittlung von vertieften Kenntnissen**

zu Entwurf, Aufbau, Realisierung und Programmierung von DBS, insbesondere

- Architektur von zentralisierten und verteilten DBS (Schichtenmodelle),
- Speichersystem, Zugriffssystem und Datensystem
- Implementierungstechniken für Konzepte und Verfahren, mit denen die einzelnen Schichten eines DBS aufgebaut werden

- **Erarbeitung eines tiefergehenden Verständnisses**

- für das Zusammenspiel der DBS-Schichten und -Komponenten und der Wechselwirkung von Implementierungsverfahren
- zur Erklärung und Verbesserung des Leistungsverhaltens bei der Abwicklung von DB-Operationen
- für das Entwerfen, Aufbauen und Warten von Datenbanken sowie für die Programmierung von DB-Anwendungen

- **Voraussetzungen für Übernahme von Tätigkeiten:**

- DB-Systementwicklung und -implementierung
- Entwicklung und Einsatz von DB-gestützten Anwendungen
- Integration von kooperativen Informations- und Anwendungssystemen
- Installation, Administration und Tuning von DB- und Transaktionssystemen
- Systemverantwortlicher für Datenbanksysteme, insbesondere Unternehmens-, Datenbank-, Anwendungs- und Datensicherungsadministrator, Informationswirt, ...

ÜBERSICHT

1. Architekturmodell von DB-Systemen

- Abbildungshierarchie eines DBS (Fünf-Schichten-Modell)
- Verteilte DBS: Einsatz von Mehrrechner-DBS
- Schichtenmodelle für Client/Server-DBS u. a.

2. Externspeicherverwaltung

- Datei- und Segmentkonzepte
- Maßnahmen zur Fehlertoleranz, Einbringverfahren
- Eigenschaften von Externspeichern

3. DB-Pufferverwaltung

- Nutzung von Lokalität
- Verwaltung des DB-Puffers (Suche, Speicherzuordnung)
- Ersetzungsverfahren und Vorausplanung

4. Web- und DB-Caching

- Client-to-Server Path and User-to-Data Path
- Typen von Web-Caches, Unterschiede DB-Puffer – DB-Cache
- Materialisierte Sichten
- Cache Groups

5. Speicherungsstrukturen

- Freispeicherverwaltung, Adressierung von Sätzen
- DeweyIDs für baumstrukturierte Organisation von Sätzen
- Satzabbildung, Cluster-Bildung

ÜBERSICHT (2)

6. Baumbasierte Zugriffspfade

- Anforderungen an Zugriffspfade
- Mehrwegbäume
- Digitalbäume

7. Hash-basierte Zugriffspfade

- Statische Hash-Verfahren
- Dynamische Hash-Verfahren
- Externes Hashing mit Separatoren
- Lineares Hashing
- Vergleich der wichtigsten eindimensionalen Verfahren

8. Sekundäre und hierarchische Zugriffspfade

- Zugriffspfade für Sekundärschlüssel
- Verfahren der Bitlistenkomprimierung
- Zugriffspfade für Set-Strukturen
- Verallgemeinerte Zugriffspfadstrukturen
- Verbund- und Pfadindexe

9. Mehrdimensionale Zugriffspfade

- Unterstützung für raumbezogene Zugriffe
- Organisation der Datensätze
- Organisation des umgebenden Datenraums
- Grid-File, R-Baum, GiST u. a.
- Vergleich verschiedener Zugriffsverfahren

ÜBERSICHT (3)

10. Satzorientierte Schnittstelle

- Data-Dictionary-Funktionen
- Scan-Konzepte
 - Tabellen-, Index-, Link-, k-d-Scans
 - Anwendung von Scans
- Sortier-Operator
- Externes Sortieren
 - Nutzung von Vorsortierung
 - Optimierung von Mischbäumen

11. Tabellenoperationen – Implementierung

- Operatoren auf einer und auf mehreren Relationen
- Implementierung der Verbundoperation
 - Nested Loop
 - Sort/Merge
 - Hash
- Implementierung von anderen binären Operatoren

12. Histogramme

- Arten von Histogrammen
 - Equi-width, equi-depth/-height
 - Mehrdimensionale Histogramme
- Probleme bei der Optimierung
- Coarse-grained optimization

ÜBERSICHT (4)

13. Mengenorientierte Schnittstelle

- Formen der Spracheinbettung
- Übersetzung von DB-Anweisungen
- Anfrageoptimierung (Standardisierung, Vereinfachung, Restrukturierung und Transformation, Kostenmodelle)
- Code-Erzeugung
- Ausführung von DB-Anweisungen

14. Betriebssystemeinbettung und DB-Tuning

- Anforderungen und Probleme
- Single-process / multi-process
- Single-tasking / multi-tasking
- Index-, Speicher-, Anfrage-Tuning
- Monitoring und Troubleshooting
- ...

15. Künftige DBMS-Architekturen

- Evolution des Schichtenmodells
- DBS-Einsatz für neue Anwendungen
- Wo liegen die Grenzen des Schichtenmodells?
- Künftige Architekturen: DBMS Ecosystems
- ...

LITERATURLISTE

Härder, T., Rahm, E.: Datenbanksysteme — Konzepte und Techniken der Implementierung, 2. Auflage, Springer-Verlag, 2001.

Gray, J., Reuter, A.: Transaction Processing—Concepts and Techniques, Morgan Kaufmann Publishers, Inc., San Mateo, CA., 1998 (5th printing).

O’Neil, P.: Database—Principles, Programming, Performance, Morgan Kaufmann Publishers, Inc., San Mateo, CA., 1994.

Mitschang, B.: Anfrageverarbeitung in Datenbanksystemen — Entwurfs- und Implementierungskonzepte, Reihe Datenbanksysteme, Vieweg, 1995.

Saake, G., Heuer, A., Sattler, K.-U.: Datenbanken: Implementierungstechniken, 2. Auflage, MITP, 2005.

Stallings, W.:

Betriebssysteme – Prinzipien und Umsetzung, 4. Auflage, Pearson Studium, 2003.

ZEITSCHRIFTEN:

ACM TODS Transactions on Database Systems, ACM-Publ. (vierteljährl.)

THE VLDB Journal VLDB Foundation (vierteljährlich)

Information Systems Pergamon Press (8-mal jährlich)

Informatik – Forschung und Entwicklung Springer-Verlag (vierteljährlich)

ACM Computing Surveys ACM-Publikation (vierteljährlich)

TAGUNGSBÄNDE:

ICDE Proceedings, „International Conference on Data Engineering“, jährliche Konferenz der IEEE

SIGMOD Proceedings, jährliche Konferenz der ACM Special Interest Group on Management of Data

VLDB Proceedings, „International Conference on Very Large Data Bases“, jährliche Konferenz der VLDB Foundation

BTW Tagungsbände der alle 2 Jahre stattfindenden Tagungen „Datenbanksysteme für Business, Technologie und Web“ der GI, und weitere Tagungen innerhalb des GI-FB „DBIS“



Datenbanktechnologie - Was ist das?

Konzepte, Methoden, Werkzeuge und Systeme für die

- dauerhafte Lebensdauer Daten > Dauer Erzeugungsprozeß
- zuverlässige Integrität, Konsistenz, Verlustsicherheit
- unabhängige wechselseitige Änderungssimmunität AWP-DB

Verwaltung und

- komfortable „höhere“ abstrakte Schnittstelle (Objekte des Benutzers)
- flexible Ad-hoc-Zugriffsmöglichkeit (generische DB-Sprache)

Benutzung von

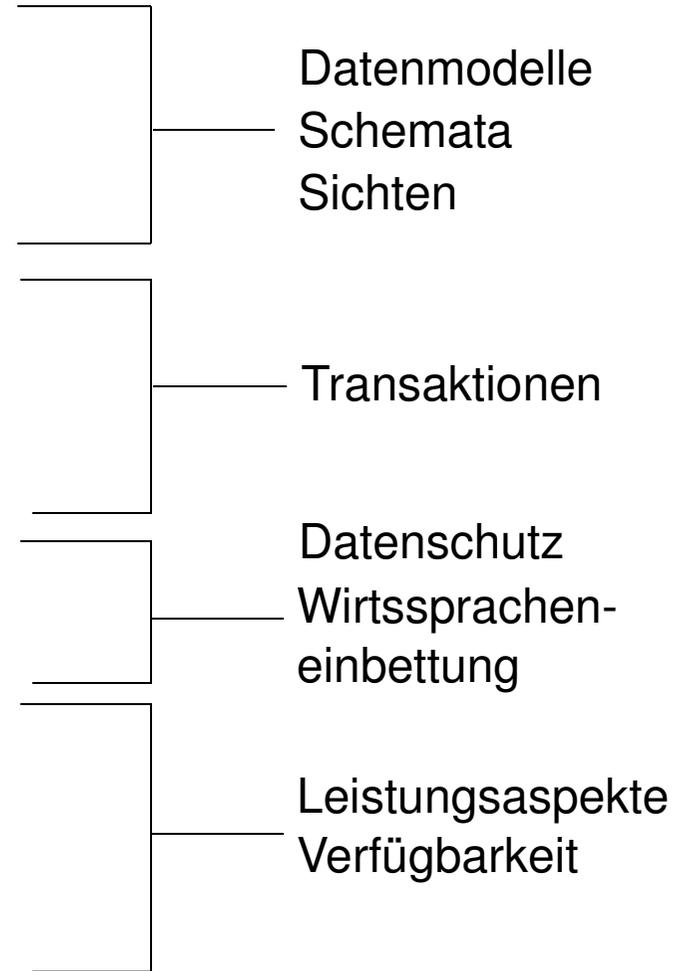
- großen Größe Daten >> Größe Hauptspeicher
- integrierten kontrollierte Redundanz von/für mehrere Applikationen,
- mehrfachbenutzbaren paralleler Zugriff

Datenbasen



Wichtige DBS-Konzepte und ihre Einordnung

- Datenintegration
- anwendungsorientierte Datenbeschreibung
- Datenunabhängigkeit
- Konsistenzkontrolle
- Mehrbenutzerbetrieb
- Datenintegrität
- Wiederanlauf
- Zugriffskontrolle
- Programmkopplung
- Ad-hoc-Anfragen
- Zugriffspfade
- Speicherungsstrukturen
- Verteilung



Notwendigkeit effizienter DBMS

- „Informationen sind in unserer vom Wettbewerb geprägten Welt ähnlich wie die Luft, die wir atmen – überall vorhanden und absolut lebenswichtig.“
- „Datenbanktechnik ist eine nützliche Infrastruktur wie fließendes Wasser, das wir erst bemerken, wenn es fehlt“
- **Nicht nur relationale Tabellen, sondern auch VITA-Daten**
 - Speicherung und Verwaltung, aber auch
 - inhaltsorientierte Suche, Verknüpfung und Modifikation

