



FAKULTÄT FÜR  
INFORMATIK

Seminar: Datenqualität und Optimierungs- und Modellierungstechniken von Datenbanken

# Maßgeschneiderte Datenhaltung

Lan Nguyen Si

17.12.2009

# Agenda

- Einleitung
  - Begriffe
  - Motivation
- Beispiel: LCARS
  - Integration
  - Implementierung
- Zusammenfassung
- Quellen

# Einleitung – Begriffe

- Datenhaltung:
  - Datensicherung und –archivierung
  - Speichermedien
  - Wiederherstellung
  - Datenkonvertierung
  - Datenschutz
- Maßgeschneiderte Datenhaltung:
  - Verbesserung des Zugriffs auf Dateien
  - Integration und Indexierung hochdimensionaler Datenobjekte
  - Referenz-Library

# Einleitung – Motivation

- Veränderung der Anforderungen an Computersysteme
- Keine Weiterentwicklung der Dateisysteme → Verzeichnishierarchie als Ordnungsschema veraltet
- Alternativer Zugang zu Dateien dringend notwendig → Nachteile der Ordnerstruktur
- Praxis: zum Teil fatal

# Einleitung – Motivation

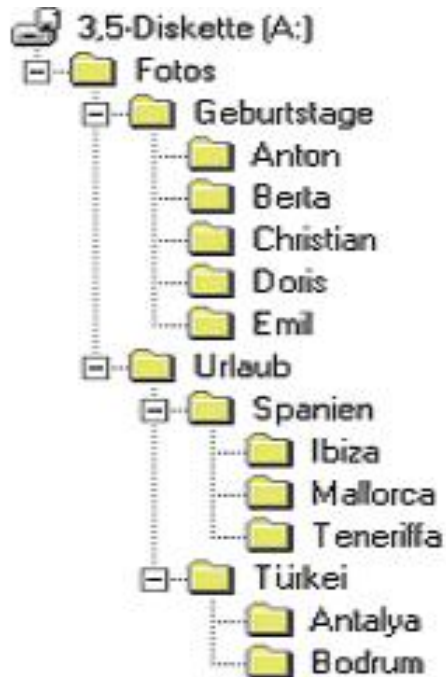


Abb. 1: Verzeichnishierarchie eines fiktiven Nutzers, Quelle: <http://www.deskwork.de/INFOS/LCARS.HTM>

- Externe und interne Suchprogramme → Ansatz
- Ziel: Entwicklung zur maßgeschneiderten Datenhaltung

# Beispiel LCARS: Integration

- Abbildung von BLOB–relationalen DB und semantischen Dateisystemen auf das „Library“–Modell
- Verwendung von Metadaten und Integration von relationalen Datenbanken
- Ziel: alle Attribute einer Datei (einschließlich Metadaten) durch Benutzeranfrage zugänglich machen
- Problem:
  - Traditionelle Dateisysteme können Metadaten nicht auslesen
  - Extraktion der Attribute

# Beispiel LCARS: Integration

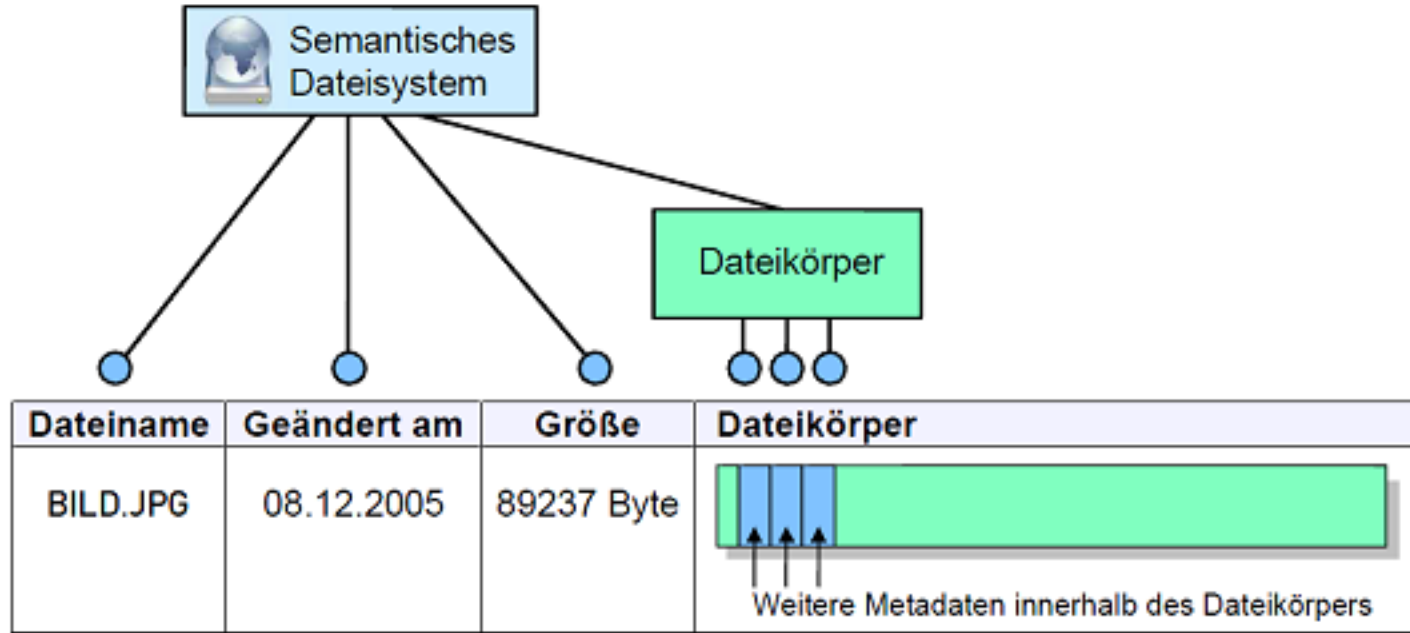


Abb. 2: JPG-Bild mit Metadaten, Quelle: Koll, K.: *Integration, Indexierung und Interaktion hochdimensionaler Datenobjekte*

- Metadaten: Informationen zur Beschreibung der Nutzdaten einer Datei
- Attribute im Dateikörper für herkömmliche Dateisysteme nicht ersichtlich

## Beispiel LCARS: Integration

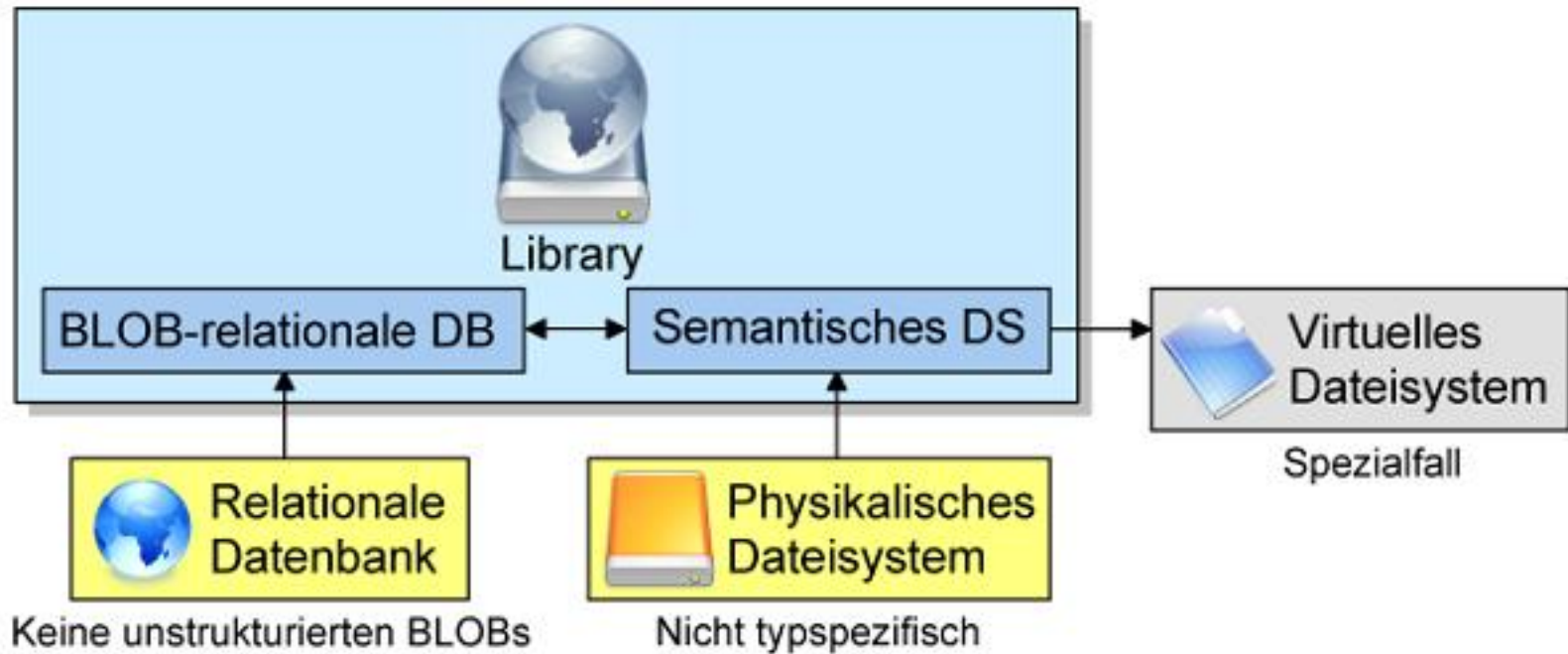


Abb. 3: Beziehung zwischen Datenbanken und Dateisystemen, Quelle: <http://www.deskwork.de/INFOS/LCARS.HTM>

- Datenbanken und Dateisysteme auf logischer Ebene nicht mehr unterschiedlich → Library
- Semantisches Dateisystem: typspezifisch → Verwaltung eigener Strukturen für Dateiformate möglich

# Beispiel LCARS: Implementierung

- Problem:
  - Behandlung der Attribute
  - Ausführung von Applikationen auf alle Dateien für jede Anfrage unmöglich
- Isolierung des Betriebssystems vom unterliegenden physikalischen Dateisystem → Zusammensetzung aus mehreren Domains
- Domains speichern Dateien, Attribute und führen Anfragen aus
- Steuerung der Domains über Library und Registry

# Beispiel LCARS: Implementierung

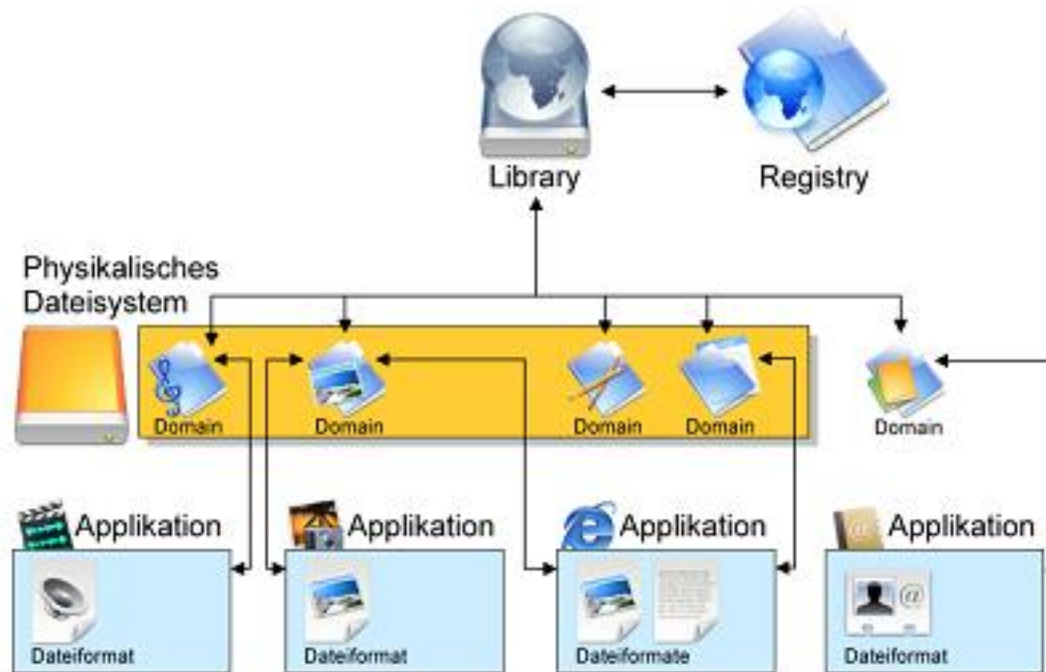


Abb. 4: Architektur des LCARS-Dateisystems, Quelle: <http://www.deskwork.de/INFOS/LCARS.HTM>

- Applikationen greifen nicht direkt auf physikalisches Dateisystem zu → Domains
- Praxis: Ersetzung der API-Funktionen innerhalb des Betriebssystems

# Beispiel LCARS: Implementierung

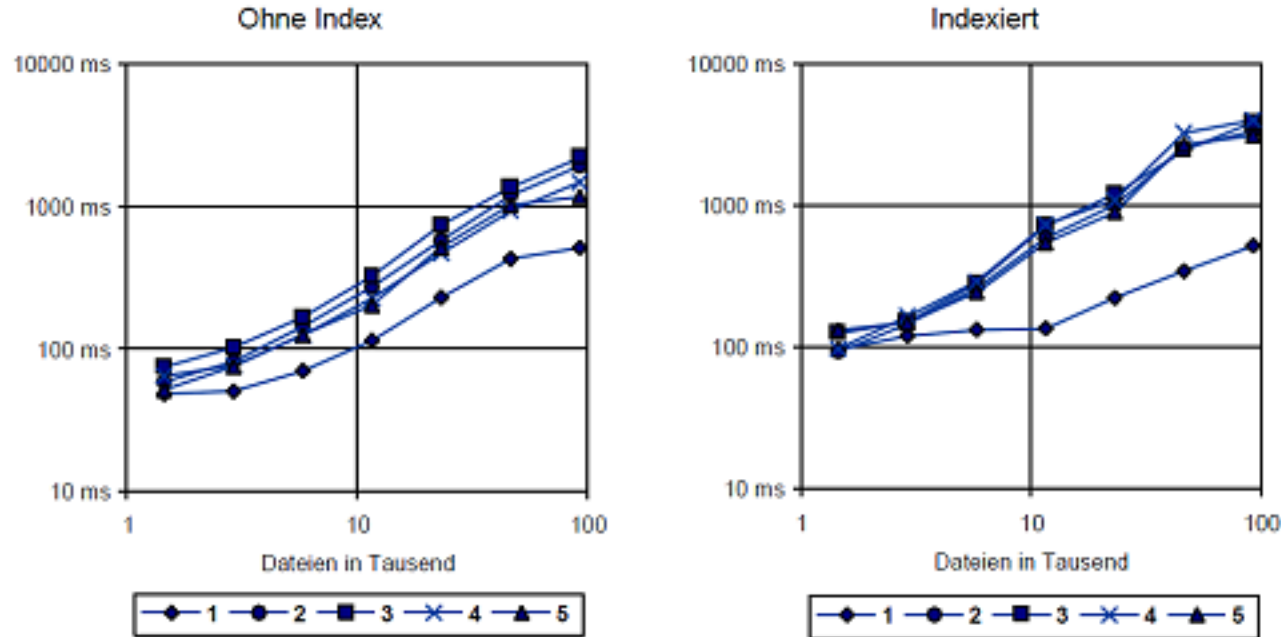


Abb. 5: Leistungsfähigkeit Microsoft SQL Server 2005, Quelle: Koll, K.: *Integration, Indexierung und Interaktion hochdimensionaler Datenobjekte*

- Indexierung von Attributen sehr schwierig
- „Fluch der Dimensionen“ → bereits bei 10 Dimensionen Überlappung von 100%

# Beispiel LCARS: Implementierung



Master index

Filename	Type	...
A.JPG	Image	...
B.MP3	MP3	...
C.JPG	Image	...
D.MP3	MP3	...
E.JPG	Image	...



Slave index for images

Filename	Width	Height
A.JPG	1024	768
C.JPG	640	480
E.JPG	2048	1536



Slave index for MP3s

Filename	Genre	Artist
B.MP3	Hip Hop	...
D.MP3	Soul	...

Abb. 6: Beispiel Master/Slave-Indexierung, Quelle: Koll, K.: *A relational file system as an example for tailor-made DMS*

- Grundidee: Hintereinanderspeicherung von Attributen aller Dateien in Heap-Datei
- Sequentielles Lesen der Indexdatei → kein Leistungsverlust durch eingesetztes Dateisystem

# Zusammenfassung

- LCARS:
  - Basierend auf „Scratch“
  - Metadaten über Applikationen
  - Domains zuständig für Attribute und Anfragen
- Allgemein:
  - Viele unterschiedliche Ergänzungen traditioneller Dateisysteme entwickelt und untersucht
  - Grundlegende Verbesserung der Organisationsstrukturen durch „Library“-Modelle
  - Master/Slave-Indexierung

# Quellen

- Koll, K.: *A relational file system as an example for tailor-made DMS*, Universität Dortmund, Paper, 2008.
- Koll, K.: *Integration, Indexierung und Interaktion hochdimensionaler Datenobjekte*, Universität Dortmund, Diss., 2009.
- DESKWORK Operation System. URL: <http://www.deskwork.de> [Stand: Dezember 2009]
- E-Teaching. URL: <http://www.e-teaching.org> [Stand: Dezember 2009]

**Vielen Dank für Ihre Aufmerksamkeit!**

**[www.ovgu.de](http://www.ovgu.de)**