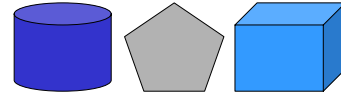


Datenintegration

Datenintegration



Kapitel 0: Organisatorisches

Andreas Thor
Sommersemester 2008

Universität Leipzig
Institut für Informatik
<http://dbs.uni-leipzig.de>

1



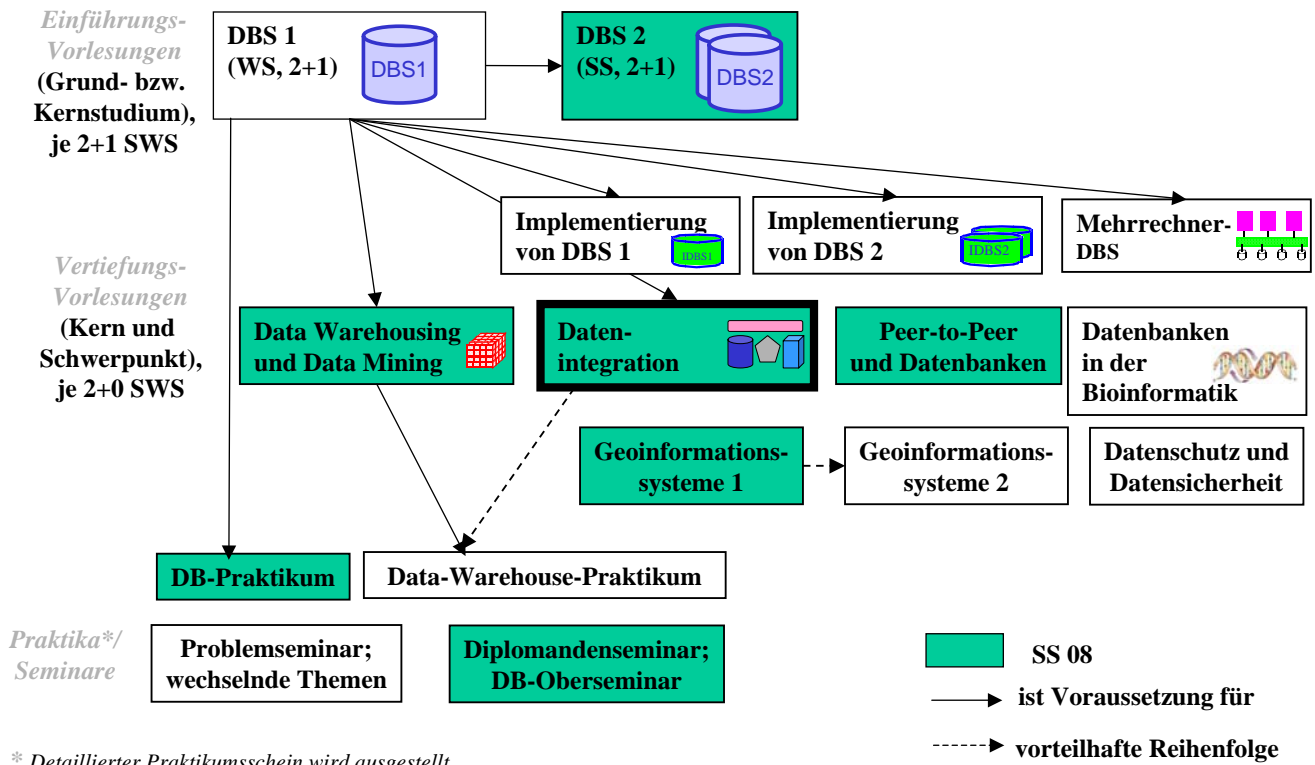
Organisatorisches

- Termin: freitags, 11.15 – 12.45 Uhr, KH 2-07
- Webseite mit allen wichtigen Infos:
<http://dbs.uni-leipzig.de/de/stud/ss2008/datenintegration>
- Diplomstudiengang Informatik (ab 5. Semester)
 - Modulklausur für Kerngebiet Praktische Informatik (2 SWS)
 - oder als Teil der mündlichen Schwerpunktprüfung
- Bachelor/Master Informatik (alte PO)
 - APL oder PL (3 credits) über Klausur
- Master Informatik (neue PO)
 - Teil des Moduls „Anwendungsbezogene Datenbankkonzepte“
- Klausur
 - Voraussichtlich Juli/August 2008
 - 60 Minuten

2



Lehrveranstaltungen zu „Datenbanken“ (SS'08)



Verwandte Vorlesungen (Prof. Rahm)

- Datenbanksysteme 1+2
 - Grundvorlesungen
 - DBS1 ist Voraussetzung für diese Vorlesung
- Implementation von Datenbanksystemen
 - Systemnahe Aspekte
 - Eine monolithische Datenbank
- Data Warehousing
 - Spezifische Form einer integrierten Datenbank
 - Fokus: Modellierung, Optimierung, OLAP ...
- Mehrrechner-Datenbanksysteme
 - Verteilte Datenbanken
 - Fokus: Datenverteilung, verteilte Abfrageverarbeitung, Performanz, ...



Data-Warehouse-Praktikum

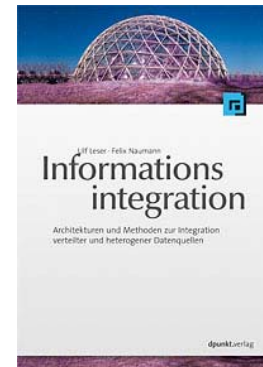
- Data Warehouse ist eine Architektur zur Datenintegration
 - große wirtschaftliche Bedeutung
 - Unterstützung durch kommerzielle Datenbanksysteme
- Praktische Realisierung typischer Elemente eines DWH-Projektes
 - Datenimport, Datenextraktion
 - Data Cleaning (Objektkonsolidierung, Datennormalisierung, Ableitung neuer Attribute)
 - Konzeption eines Warehouse-Schemas mit entsprechender Datenbefüllung
 - Analysen mittels MDX-Anfragen
- Technisches
 - Ca. 200 MB Daten (CSV, XML)
 - Software: SQL Server 2005 und Visual Studio
 - Gruppenarbeit (2 Studenten) mit 3 Testdaten
- Organisatorisches
 - Wintersemester 2008/09 (evtl. Beginn bereits ab August 2008 möglich)

5



Literatur

- Ulf Leser, Felix Naumann:
„Informationsintegration“,
dpunkt.verlag, Oktober 2006
- Hinweise auf wissenschaftliche Arbeiten
in den einzelnen Kapiteln
- Vorlesungsskript basiert auf Skripten zur Vorlesung
„Informationsintegration“ von
 - Prof. Dr. Ulf Leser
 - Professur „Knowledge Management in Bioinformatics“
 - Humboldt-Universität zu Berlin
 - Prof. Dr. Felix Naumann
 - Fachgebiet „Informationssysteme“
 - Hasso-Plattner-Institut, Potsdam



6



Inhaltsverzeichnis (vorläufig)

1. Einführung (1 VL)
 - Begriffsdefinition
 - Beispiele und Anwendungsgebiete
 - Integration am Beispiel
2. Verteilung, Autonomie und Heterogenität (1)
 - Verteilung
 - Autonomie
 - Heterogenität
3. Eigenschaften von Integrationssystemen (1)
 - Bzgl. Verteilung, Autonomie und Heterogenität
 - Weitere Kriterien
 - Materialisiert vs. virtuell
4. Integrationsarchitekturen (2)
 - Data Warehouse
 - Föderierte Datenbanken
 - Mediator-Wrapper-Ansatz
 - Peer-Data-Management-Systeme
 - Architekturen vs. Integrationsaufgaben
5. Anfrageverarbeitung (2)
 - SchemaSQL
 - Global-as-View (GaV), Local-as-View (LaV)
 - Quellen mit beschränkten Anfragemöglichkeiten
6. Schemamanagement (2)
 - Schemaintegration
 - Schema Matching
 - Schema Mapping
7. Datenfusion (2)
 - Datenqualität
 - Duplikaterkennung
 - Konfliktbehandlung
8. Webdatenintegration (2)
 - Hidden Web
 - Semantic Web
 - Mashup-Frameworks



Freie SHK-Stellen

- Voraussetzungen
 - Diplom, Bachelor- oder Master-Studium der Informatik (ab dem 3. Semester)
 - Gute Datenbankkenntnisse
 - Gute Kenntnisse bei der Erstellung von Web-Applikationen
 - Verantwortungsbewusstsein, Zuverlässigkeit
- 1. DFG-Projekt Dynamische Fusion verteilter Webdaten
 - Implementationen zur Anbindung neuer Web-Quellen
 - Evaluation von Query- und Match-Verfahren zur Datenintegration
- 2. Lehre
 - Betreuung der eingesetzten E-Learning-Plattform LOTS
 - Unterstützung praktischer Übungen zu den angebotenen Vorlesungen
- Kontakt: Andreas Thor (thor@informatik.uni-leipzig.de)

