

Data Warehousing

Sommersemester 2005

Ulf Leser

Wissensmanagement in der
Bioinformatik



-
- „... Der typische Walmart Kaufagent verwendet täglich mächtige Data Mining Werkzeuge, um die Daten der **300 Terabyte Datenbank** zu erforschen“

[Jim Gray, Computer Zeitung 17/2003]

- Was verkaufen wir wo am häufigsten ?
- Gibt es einen Zusammenhang zwischen der Regalposition und den Umsatzzahlen von Produkten ?
- Wie lange müssen wir Produkt X durchschnittlich lagern und wo ?

Data Warehousing

- Vorlesung: Mo, 15.00 – 17.00 Uhr
- Keine Übung
- Welche Termine fallen aus ?
 - 16.5. Pfingsten
- Folien im Web jeweils vorab verfügbar
- Sprechstunde
 - Nach Vereinbarung / Offene Türen
 - IV.105
 - (030) 2093 – 3901
 - [leser\(...\)informatik.hu-berlin.de](mailto:leser(...)informatik.hu-berlin.de)

Einbettung

- Voraussetzungen
 - DBS-I
 - Relationalenmodell, ER-Modell
 - Joins, Anfrageübersetzung
 - Grundzüge de Anfrageoptimierung
 - SQL
 - Logik
- Keine Voraussetzung
 - DBS-II

Ziel der Vorlesung

- Techniken und Konzepte im Umgang mit großen Datenbanken (=Data Warehouse)
 - Architektur und Prozesse eines DWH
 - Performance-optimierte Schemata
 - Multidimensionales „statt“ relationales Datenmodell
 - Advanced Features: Bitmap Index, Materialisierte Views, STAR-Joins, multidimensionale Indexstrukturen
- Unterschiede in **Nutzungscharakteristika** von Datenbanken kennen lernen
 - Andere Anforderungen – andere Methoden
 - Missachtung lässt Projekte scheitern

Inhaltsübersicht –1-

- Einleitung & Motivation
 - Was ist ein DWH ?
- Architektur & Prozesse
 - Konzepte, Komponenten und Begriffe
- Modellierung von DWH
 - Multidimensionale Modellierung
 - Spezielle Operatoren auf MDDM
- Umsetzung des multidimensionalen Datenmodells
 - Relationale Abbildung
 - SQL-basierte DWH-Operatoren
 - MDX

Inhaltsübersicht –2-

- Indexstrukturen für DWH
 - Bitmap Indexe
 - Multidimensionale Indexstrukturen
- Logische Optimierung
 - Star-Join und Partitionierung
- Materialisierte Sichten
 - Precomputed statt „On demand“
 - „Answering Queries using Views“
- Extraction, Transformation & Load (ETL)
 - Integrierte und aktuelle Daten für das DWH
 - Updatestrategien
 - Datenqualität und Data Cleansing

Gastvortrag

- 20.6.: Microsoft SQL Server – OLAP and BI Features
 - Vorlesung eines Microsoft-Mitarbeiters
- Außerdem (mit DBS-2)
 - Teradata Parallel Database Server
 - Microdiscovery: BI Tools
 - Oracle Vortrag und Hands-On Workshop

Anrechenbarkeit

- Halbkurs zusammen mit
 - Der Vorlesung „XML und Datenbanken“
 - Prof. Naumann, dieses Semester
 - Der Vorlesung „Molekularbiologische Datenbanken“
 - Dazu ist ein Übungsschein vom SoSe2004/2003 notwendig
- Weitere Kombinationen auf Anfrage

Literatur

- Primär
 - Lehner: „Datenbanktechnologie für Data Warehouse Systeme“, dpunkt.Verlag, 2003, ca. 40.- Euro
- Weitere
 - Bauer/Günzel: „Data Warehouse Systeme“, dpunkt.Verlag, 2001, ca. 50.- Euro
 - Kimball et al. „The Data Warehouse Life Cycle Toolkit“, John Wiley & Sons, 1998
 - Jürgens: „Index Structures for Data Warehouses“, LNCS 1859, Springer-Verlag, 2000
 - Oehler: „OLAP: Grundlagen, Modellierung und betriebswirtschaftliche Grundlagen“, Hanser Verlag, 2000
- Übersichtsartikel
 - Chaudhuri, Dayal: „An Overview of Data Warehousing and OLAP Technology“, SIGMOD Record, 1997
 - Widom: „Research Problems in Data Warehousing“, CIKM, 1994

Fragen ?

Fragen

- Diplominformatiker?
- Semester?
- DBS-I?
- Prüfung?

Wissensmanagement in der Bioinformatik

- Schwerpunkte
 - Algorithmen der Bioinformatik
 - Management molekularbiologischer Daten
 - Datenintegration
 - Text Mining
- Laufend möglich
 - Studienarbeiten
 - Diplomarbeiten
 - ... oft in Kooperation mit Bioinformatikfirmen
 - ... oft in Kooperation mit Bioinformatikinstituten

Werbung

