

Architektur von Datenbanksystemen Skript

In der Hoffnung, dass es was nützt ...

Christian Kroh

29. Juni 2014, Dresden

Inhaltsverzeichnis

1. Allgemeine Einführung	3
1.1. Qualifikationsziele	3
1.2. Basisliteratur	3
1.3. Spezielle Literatur	3
 I. Allgemeines	 4
1. Datenbanksystem	5
2. Datenbank Anfragen	6
2.1. Anfragesprachen	6
2.1.1. Typen von Anfragesprachen	6
2.1.2. Optimierungsarten	6
2.2. Anfrageverarbeitung	6
3. Komponenten eines Datenbanksystems	7
4. Schichtenmodell	8
 II. ADBS-1: Betriebssystem, Puffer- und Speicherverwaltung	 9
1. Einführung	10
2. Segment- und Seitenverwaltung	11
3. Einbringstrategien	12
4. Systempuffer	13
5. Row-based Record Management (klassische Satzverwaltung)	14
6. Column-based Record Management	15
7. Distributed Key-Value Stores	16
8. Replication	17
 III. ADBS-2: Zugriffssystem, Datensystem	 18
 IV. Appendix	 19

1. Allgemeine Einführung

1.1. Qualifikationsziele

Die Lehrveranstaltung widmet sich im Wesentlichen der Architektur von Datenbanken, wobei die Thematik von unterschiedlichen Perspektiven betrachtet wird. Zum einen wird der Aufbau traditioneller Datenbanksysteme schrittweise hinsichtlich der verschiedenen funktionalen Schichten diskutiert, wobei datenbankspezifische Konzepte der Externspeicherverwaltung, des Systempuffer-Managements und der internen Satzschnittstelle ausführlich untersucht werden. Des Weiteren werden orthogonale Konzepte wie aus dem Bereich der parallelen Datenbankarchitektur (Parallelisierung von Datenbankoperatoren) und aus dem Umfeld verteilter Sperr- und Cache-Synchronisationsverfahren betrachtet. Ebenfalls werden im Bereich der verteilten Datenbanken notwendige Technologien für einen zuverlässigen transaktional ausgerichteten Betrieb aufgearbeitet (2-Phasen-Commit-Protokoll, Replikation, etc.). Neben den traditionellen Datenbanksystemen wird auch auf alternative Konzepte, wie z.B. „Column-Store“ Datenbanken und Datenstromsysteme eingegangen. Außerdem werden Prinzipien und Techniken wie Fragmentierung, Replikation, Virtualisierung, Multi-Tenancy oder CAP-Theorem sowohl im Bezug auf operationale wie auch auf analytische Probleme diskutiert.

Datenbank- und Informationssysteme können in zwei große Bestandteile zerlegt werden. Auf der einen Seite steht das Speichermanagement, d.h. die Komponenten, die für die physische Organisation der Datenbestände zur Verfügung stehen. Die unterschiedlichen Methoden und Komponenten werden dabei im ersten Teil der Vorlesungsreihe besprochen. Auf der anderen Seite steht die Verarbeitungseinheit, also die Komponenten eines Systems, die transaktional geschützt, nebenläufig korrekt und effizient Datenbank Anfragen bearbeiten. Die Methoden und die Aufgaben der unterschiedlichen Komponenten werden in diesem zweiten Teil der Vorlesung besprochen. Als Schnittstelle zwischen den Schichten innerhalb der Datenbankarchitektur dient die Auffassung eines Key-Value-Paares. Dabei kann Key-Value bedeuten, dass dies direkt von der Applikation genutzt bzw. programmiert werden kann oder höhere Systemschichten darauf aufsetzen. Ziel der Vorlesung ist es ein tiefes Verständnis von Anfrageverarbeitungs- und optimierungstechniken unter den Randbedingungen transaktionaler Korrektheit zu vermitteln.

1.2. Basisliteratur

- A. Kemper, A. Eickler: „Datenbanksysteme“, Oldenbourg-Verlag
- R. Elmasri and S. Navathe: „Grundlagen von Datenbanksystemen“, 3. überarbeitete Auflage, Addison-Wesley, 2002

1.3. Spezielle Literatur

- Gunter Saake, Andreas Heuer, Kai-Uwe Sattler: Datenbanken: Implementierungstechniken, mitp Verlag, 2. Auflage
- Härder T.; Rahm, E.: Datenbanksysteme. Konzepte und Techniken der Implementierung, Springer-Verlag
- Ramakrishnan, R.; Gehrke, J.: „Database Management Systems“. McGraw-Hill, 2000

Teil I.

Allgemeines

1. Datenbanksystem

Definition 1: Datenbanksystem:

ein System zur Beschreibung, Speicherung und Wiedergewinnung von umfangreichen Datenmengen, die von mehreren Anwendungsprogrammen genutzt werden

Komponenten

- Datenbank, in der die Daten abgelegt werden
- Datenbanksoftware, die die Daten entsprechend den vorgegebenen Beschreibungen abspeichern, auffinden oder weitere Operationen mit den Daten durchführen

Leistung in den folgenden Bereichen

- Datenmodell und Datendefinition
- Datenzugriff und -manipulation
- Steuerung und Überwachung

2. Datenbankanfragen

2.1. Anfragesprachen

2.1.1. Typen von Anfragesprachen

- Prozedurale Datenbanksprachen
 - Tupel- oder Satzorientiert
 - Programmierer denkt in Satzfolgen
 - Navigation über Zugriffspfade durch die vorhandenen Daten
 - findNext()
 - findFirst()
- Deskriptive Datenbanksprachen
 - Mengenorientiert (typisch für Relationenmodell)
 - Programmierer denkt in Mengen von Sätzen mit bestimmten Eigenschaften
 - Zugriff erfolgt durch inhaltliche Kriterien

Beispiel SQL

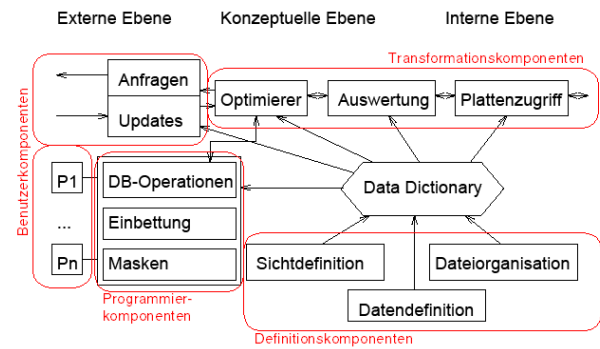
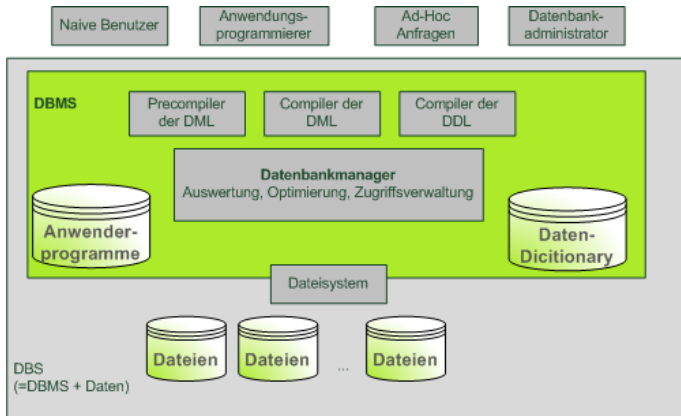
DDL (Data Definition Language)

DML (Data Manipulation Language)

2.1.2. Optimierungsarten

2.2. Anfrageverarbeitung

3. Komponenten eines Datenbanksystems



4. Schichtenmodell

Teil II.

ADBS-1: Betriebssystem, Puffer- und Speicherverwaltung

1. Einführung

2. Segment- und Seitenverwaltung

3. Einbringstrategien

4. Systempuffer

5. Row-based Record Management (klassische Satzverwaltung)

6. Column-based Record Management

7. Distributed Key-Value Stores

8. Replication

Teil III.

ADBS-2: Zugriffssystem, Datensystem

Teil IV.

Appendix