



Übung zur Vorlesung *Grundlagen: Datenbanken* im WS14/15

Harald Lang (harald.lang@in.tum.de)

<http://www-db.in.tum.de/teaching/ws1415/grundlagen/>

Blatt Nr. 14

Am 28.01.2015 findet anstelle der Vorlesung eine Zentralübung/Fragestunde statt.

**Hausaufgabe 1 - Mehrbenutzersynchronisation: Eigenschaften von Historien**

Kreuzen Sie alle Eigenschaften an, die von den Historien erfüllt werden.

a)  $H_1 =$

Schritt	$T_1$	$T_2$	$T_3$
1	$w(x)$		
2		$r(x)$	
3		$w(y)$	
4		$c$	
5			$r(y)$
6			$w(z)$
7			$c$
8	$c$		

richtig	falsch	Aussage
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Die Historie ist serialisierbar (SR)
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Die Historie ist rücksetzbar (RC)
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Die Historie ist vermeidet kaskadierendes Rücksetzen (ACA)
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Die Historie ist strikt (ST)

b)  $H_2 =$

Schritt	$T_1$	$T_2$
1	$w(x)$	
2	$w(z)$	
3		$w(z)$
4	$c$	
5		$w(x)$
6		$c$

richtig	falsch	Aussage
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Die Historie ist serialisierbar (SR)
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Die Historie ist rücksetzbar (RC)
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Die Historie ist vermeidet kaskadierendes Rücksetzen (ACA)
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Die Historie ist strikt (ST)

## Hausaufgabe 2 - Mehrbenutzersynchronisation: Deadlocks

Nennen Sie die Vorteile und Nachteile von Deadlockerkennung und -vermeidung durch:

- Timeouts
- Wartegraphen
- Preclaiming
- Zeitstempel

Sind Kombinationen denkbar/sinnvoll?

### Lösung:

- Timeouts
  - + Verfahren ist einfach zu implementieren und erfordert sehr geringen Verwaltungsaufwand.
  - Deadlocks werden erst mit Verzögerung erkannt.
  - Es gibt false positives.
- Wartegraphen
  - + Es werden echte Deadlocks schnell erkannt.
  - Aufwendig: Pflegen des Graphen, Zyklenerkennung
- Preclaiming
  - + Es treten keine Deadlocks auf,
  - verringert jedoch die Parallelität.
- Deadlockvermeidung durch Zeitstempel
  - + Deadlocks treten nicht auf.
  - Viele Transaktionen müssen zurückgesetzt werden, obwohl nie ein Deadlock auftreten würde – false positives.

!!! **Anmerkung:** Nicht zu verwechseln mit *zeitstempelbasierter Synchronisation*. Dort können auch keine Verklemmungen auftreten, aber existieren auch keine Sperren! (vgl. Blatt 13/Aufgabe 5)

Zur Erinnerung; Zeitstempel treten an verschiedenen Stellen auf:

- \* bei der Deadlockvermeidung (wound-wait, wait-die),
- \* bei zeitstempelbasierter Synchronisation, und auch
- \* bei optimistischer Synchronisation, wo Transaktionen einen Zeitstempel beim Eintritt in die Validierungsphase erhalten. Über die Zeitstempel wird ermittelt, welche Transaktionen parallel gelaufen sind, und mit welchen WriteSets das aktuelle ReadSet verglichen werden muss.

Beispielsweise können Timeout und Wartegraph kombiniert werden. Hierbei wird der Wartegraph erst dann erzeugt, wenn ein Timeout auftritt. Es treten so im Vergleich zum Timeout-Verfahren keine false positives mehr auf und das Verfahren bleibt im Regelfall "billig", da der Wartegraph erst bei Bedarf erzeugt wird. Das Problem der verzögerten Erkennung von Deadlocks bleibt allerdings bestehen.

### Hausaufgabe 3 - Mehrbenutzersynchronisation in SQL

SQL-92 spezifiziert mehrere Konsistenzstufen (*isolation level*) durch welche der Benutzer (bzw. die Anwendung) festlegen kann, wie "stark" eine Transaktion von anderen parallel laufenden Transaktionen isoliert werden soll.

- Nennen und erläutern Sie kurz die Isolation Level. Geben Sie an, welche Nebenläufigkeitsprobleme mit dem jeweiligen Level vermieden werden.
- Warum kann zwischen den Konsistenzstufen gewählt werden?

#### Lösung:

a) Nicht abgesicherte Nebenläufigkeit kann zu folgenden Problemen führen:

- lost update
- dirty read
- non-repeatable read
- phantom problem

Die folgende Tabelle zeigt, welche Probleme durch die jeweiligen Isolation Level verhindert (✓) werden:

		lost update	dirty read	non-repeatable read	phantom problem
isolation level	read un-committed	✓			
	read committed	✓	✓		
	repeatable read	✓	✓	✓	
	serializable	✓	✓	✓	✓

- Die Isolation Level erlauben einen Kompromiss zwischen Performanz und Genauigkeit zu schließen. Je mehr Genauigkeit/Sicherheit, desto langsamer wird die Ausführung. Zusätzlich kann dem Datenbanksystem mitgeteilt werden, ob die Transaktion *read only* ist. Sind im System ausschließlich read-only-Transaktionen aktiv, dann können diese völlig uneingeschränkt parallel laufen, da Konfliktoperationen ausgeschlossen sind.

### Hausaufgabe 4 - Indexstrukturen

Gegeben seien ein Rot-Schwarz-Baum sowie ein Trie (oder Präfixbaum) mit einem fan-out von 256. Vergleichen Sie die beiden Datenstrukturen hinsichtlich der Baumhöhe in Abhängigkeit von der Anzahl der indizierten Werte  $n$ . Geben Sie an, ab wann der Rot-Schwarz-Baum frühestens und spätestens die Höhe des Präfixbaums überschreitet; jeweils für Elemente vom Typ 32-Bit Integer und 64-Bit Integer.

**Lösung:**

Die Länge des Schlüssels sei  $k$  Bits. Jeder Knoten hat bis zu 256 Kindknoten (fan-out). Somit lassen sich  $\log_2(256) = 8$  Bits = 1 Byte des Schlüssel pro Knoten speichern. Die Höhe des Präfixbaums ist unabhängig von der Anzahl der indizierten Elemente und beträgt somit  $\lceil k/8 \rceil$ .

Die Höhe eines Rot-Schwarz-Baums hingegen ist abhängig von der Anzahl der indizierten Elemente  $n$ . Ist der Baum perfekt ausbalanciert beträgt die Höhe  $h_b(n) = \lceil \log_2(n+1) \rceil$  und im ungünstigsten Fall  $h_w(n) = 2 \log_2(n+2) - 2$ .

Desweiteren gilt, dass die minimale Anzahl Knoten in einem Rot-Schwarz-Baum mit ungerader Höhe  $m_h = 3 \cdot 2^{\frac{h-1}{2}} - 2$  beträgt und bei gerader Höhe  $m_h = 2 \cdot 2^{\frac{h}{2}} - 2$ .

Für Schlüssel vom Typ 32-Bit Integer ergibt sich:

- Die Höhe des Tries beträgt  $\lceil 32/8 \rceil = 4$ .
- Ein Rot-Schwarz-Baum der Höhe 4 enthält höchstens  $2^4 - 1 = 15$  Elemente.
- Ein Rot-Schwarz-Baum der Höhe 5 enthält mindestens  $3 \cdot 2^{\frac{5-1}{2}} - 2 = 10$  Elemente.
- Die Höhe des Tries wird frühestens mit dem Einfügen des 10-ten Elements überschritten und spätestens mit dem Einfügen des 16-ten Element.

Für Schlüssel vom Typ 64-Bit Integer ergibt sich:

- Die Höhe des Tries beträgt  $\lceil 64/8 \rceil = 8$ .
- Ein Rot-Schwarz-Baum der Höhe 8 enthält höchstens  $2^8 - 1 = 255$  Elemente.
- Ein Rot-Schwarz-Baum der Höhe 9 enthält mindestens  $3 \cdot 2^{\frac{9-1}{2}} - 2 = 46$  Elemente.
- Die Höhe des Tries wird frühestens mit dem Einfügen des 46-ten Elements überschritten und spätestens mit dem Einfügen des 256-ten Element.

**Hausaufgabe 5 - Relationale Entwurfstheorie**

Überführen Sie verlustlos in die 4. NF:

Relation

$$\mathcal{R} = \{A, B, C, D\}$$

mit den FDs/MVDs

$$\begin{aligned} AB &\twoheadrightarrow C \\ BC &\twoheadrightarrow D \\ BA &\rightarrow CD \\ DA &\rightarrow B \end{aligned}$$

**Lösung:**

Kandidatenschlüssel:  $\kappa_1 = \{A, B\}$  und  $\kappa_2 = \{A, D\}$

- Betrachte  $AB \twoheadrightarrow C$  in  $\mathcal{R}$ :
  - MVD ist nicht trivial ( $AB \cap C = \emptyset$  und  $AB \cup C \subset \mathcal{R}$ ),
  - aber die linke Seite ist ein Superschlüssel
  - und stellt somit keine Verletzung der 4. NF dar.
- Betrachte  $BC \twoheadrightarrow D$  in  $\mathcal{R}$ :
  - MVD ist nicht trivial ( $BC \cap D = \emptyset$  und  $BC \cup D \subset \mathcal{R}$ ),
  - und die linke Seite ist **kein** Superschlüssel.
  - Die 4. NF wird dadurch verletzt.
- Zerlege  $\mathcal{R}$  anhand  $BC \twoheadrightarrow D$  in
  - $\mathcal{R}_1 = \{B, C\} \cup \{D\} = \{B, C, D\}$ , und
  - $\mathcal{R}_2 = \mathcal{R} \setminus \{D\} = \{A, B, C\}$ , mit den FDs/MVDs
  - $F_{\mathcal{R}_1} = \{\}$ ,  $D_{\mathcal{R}_1} = \{BC \twoheadrightarrow D\}$ , und
  - $F_{\mathcal{R}_2} = \{BA \rightarrow C\}$ ,  $D_{\mathcal{R}_2} = \{AB \twoheadrightarrow C\}$ .Die FDs  $BA \rightarrow D$  und  $DA \rightarrow B$  gehen verloren.
- Betrachte  $BC \twoheadrightarrow D$  in  $\mathcal{R}_1$ :
  - MVD ist trivial, da  $BC \cup D = \mathcal{R}_1$ ,
  - und stellt somit keine Verletzung der 4. NF dar.
  - $\mathcal{R}_1$  kann nicht weiter zerlegt werden.
- Betrachte  $AB \twoheadrightarrow C$  in  $\mathcal{R}_2$ :
  - MVD ist trivial, da  $AB \cup C = \mathcal{R}_2$ ,
  - und stellt somit keine Verletzung der 4. NF dar.
  - $\mathcal{R}_2$  kann nicht weiter zerlegt werden.

Ergebnis:  $\mathcal{R}_1 = \{\underline{B}, \underline{C}, D\}$ ,  $\mathcal{R}_2 = \{\underline{A}, \underline{B}, C\}$