



10. Übungsblatt

(Testwoche: 22. - 24. Juni 2010)

Einführung in Datenbanksysteme
Datenbanken für die Bioinformatik

Heinz Schweppe, Katharina Hahn

Aufgabe 1 (Concurrency Control)

4+3 Punkte

a) Betrachten Sie folgende Historie:

r1(a), r1(b), r2(a), r2(b), w2(a), c2, w1(b), c1

- (1) Geben Sie die Ausgabe für das strikte 2PL an. Geben Sie dort, wo es im Protokoll vorgesehen ist, die entsprechenden Lock- und Unlock-Operationen an. Die Locks werden schrittweise vergeben, d.h. bei read und write Zugriff erst ein Read-Lock (lockS), dann ein Write-Lock (lockX).
- (2) Geben Sie die Ausgabe für das Snapshot Isolation Protokoll für obige Historie an. Geben Sie die Ausgabe für die Variante „first-committer-wins“ an.

b) Betrachten Sie folgende Historie:

r2(a), w2(a) r1(b), r3(b), w3(b), r1(a), c1, r2(c), c2, r3(c), w3(c), c3

Geben Sie die Ausgabe für das *MVCC für read-only Transaktionen* für obige Historie an. Transaktion T1 ist als read-only Transaktion gekennzeichnet. Kennzeichnen Sie die entsprechenden Versionen der Objekte, die gelesen bzw. geschrieben werden.

Aufgabe 2 (Recovery)

2+2+2 Punkte

- a) Die Commit-Regel (auch „force at commit“) besagt bekanntlich, dass der Commit-Log-Eintrag in den persistenten Speicher geschrieben werden muss. Gilt das auch für den Abort-Log-Eintrag für abgebrochene Transaktionen (mit Begründung)?
- b) Damit logisches Protokollieren (logical log – es werden die Operationen protokolliert) funktioniert, muss die jeweilige Operation ein Inverses haben. Geben Sie ein Beispiel für eine Datenbankoperation an, die kein Inverses hat. Wie könnte ein Verfahren aussehen, dass mit dieser Situation umgehen kann.
- c) Wir betrachten die Undo-Verarbeitung für eine Transaktion TA1, die abgebrochen wird. Sie hat ein Objekt in einer Seite verändert, der LSN-Eintrag 233. Danach hat TA2 ein Objekt der Seite verändert, LSN 234, und die Seite wurde zurückgeschrieben (LSN der Datenseite jetzt 234). TA2 endet mit commit, danach erfolgt das abort von TA1. Welche LSN hat die Datenseite nach der Undo-Verarbeitung?

Aufgabe 3 (Data-Warehouse)**3+3 Punkte**

Für diese Aufgabe wird Ihnen die sogenannte TPC-H Datenbank zur Verfügung gestellt. Hierbei handelt es sich um eine Datenbasis für einen Benchmark im Bereich Decision Support Systeme. Die Datenbank bildet den Beschaffungs- und Verkaufsprozess eines global agierenden Unternehmens ab. Details entnehmen Sie bitte der Dokumentation des TPC-H Benchmarks (<http://www.tpc.org/tpch/spec/tpch2.11.0.pdf>): Für Sie relevant ist Seite 12, auf der das Schemas abgebildet ist.

Die Datenbank steht Ihnen auf *esel* **nur unter Oracle** im Schema *tpch* zur Verfügung. Beantworten Sie die folgenden Anfragen in jeweils gültigem SQL mit Hilfe des CUBE bzw. ROLLUP Operators.

- a) Schreiben Sie eine Anfrage, die den durchschnittlichen Preisnachlass (*discount*) der Einzelposten von Bestellungen nach Herkunft der Kunden (*customer*) gruppiert. Dabei sollen die Hierarchisierungsstufen Nation (*nation*) und Region (*region*) der Dimension Herkunft betrachtet werden. Das Ergebnis soll sowohl den Preisnachlass pro Nation als auch pro Region auflisten.
- b) Wie hoch ist der Durchschnittspreis von Waren (*part.retailprice*) einer bestimmten Marke (*part.brand*) bei unterschiedlichen Lieferanten (*supplier*) aus den USA, deren Kontoguthaben größer als 9950 ist? Ordnen Sie das Ergebnis aufsteigend nach Marke, dann nach Durchschnittspreis. In dem Ergebnis sollen sowohl die Durchschnittspreise pro Marke und Lieferant, als auch pro Marke, als auch pro Lieferant enthalten sein.

Zusatzaufgabe (Data-Warehouse)

- a) In der ER-Modellierung eines Sternschemas modelliert man die Faktentabelle oft als Beziehung. Erläutern Sie das unter Verwendung eines Beispiels
- b) Entwerfen Sie einen Auswertungsalgorithmus für den ROLLUP-Operator, der sich dadurch auszeichnet, dass bereits berechnete Aggregationen wiederverwendet werden, statt sie neu zu berechnen.