



Universität Freiburg
Institut für Informatik
Prof. Dr. Georg Lausen
Io Taxidou

Georges-Köhler Allee, Geb. 51
D-79110 Freiburg
lausen@informatik.uni-freiburg.de
taxidou@informatik.uni-freiburg.de

Übungen zur Vorlesung
Datenbanken und Informationssysteme
Wintersemester 2017/2018

Ausgabe: 16.01.2018
Abgabe: 22.01.2018, 12:00 Uhr

13. Aufgabenblatt: Formaler und Physischer Datenbankentwurf

Übung 1 (1+1=2 Punkte)

Welche der folgenden Relationsschemata sind in 3NF bzw. BCNF? Begründen Sie Ihre Antwort.

- a) $R_1(A, B, C, D, E)$, $\mathcal{F}_1 = \{C \rightarrow DE, AC \rightarrow B\}$
- b) $R_2(A, B, C, D, E, F)$, $\mathcal{F}_2 = \{A \rightarrow D, D \rightarrow F, F \rightarrow ABCE, B \rightarrow D\}$

Übung 2 (2+3+2=7 Punkte)

Gegeben sei ein Relationschema R über $V = \{A, B, C, D\}$ mit den folgenden funktionalen Abhängigkeiten:

$$\mathcal{F} = \{A \rightarrow B, A \rightarrow C, A \rightarrow D, C \rightarrow D, BC \rightarrow A, AD \rightarrow B\}.$$

- a) Überprüfen Sie mittels des X^+ -Algorithmus, ob A und BC Schlüssel sind.
- b) Bestimmen Sie eine minimale Überdeckung \mathcal{F}^{min} zu \mathcal{F} .
Hinweis: Es genügt, wenn Sie hierzu $BC \rightarrow A$ und $AD \rightarrow B$ auf mögliche Linksreduktionen sowie $A \rightarrow D$ auf mögliche Rechtsreduktionen überprüfen.
- c) Ist R in 3NF? Begründen Sie.
Falls R nicht in 3NF ist, geben Sie eine Zerlegung ρ von R mit Hilfe des 3NF-Synthese-Algorithmus an.

Übung 3 (2+2=4 Punkte)

Verwenden Sie den 3NF-Synthese-Algorithmus aus der Vorlesung, um zu den folgenden Relationsschemata R_1, R_2 mit den funktionalen Abhängigkeiten $\mathcal{F}_1, \mathcal{F}_2$ jeweils eine Zerlegung in 3NF zu bestimmen.

- a) $R_1 = (A, B, C, D, E, F, G)$, $\mathcal{F}_1 = \{A \rightarrow BC, B \rightarrow A, F \rightarrow BDEG, BDE \rightarrow FG\}$
- b) $R_2 = (A, B, C, D, E)$, $\mathcal{F}_2 = \{AB \rightarrow C, CD \rightarrow A, A \rightarrow E\}$

Übung 4 (1+1=2 Punkte)

Betrachten Sie die folgenden Relationen:

$\text{Emp}(eid: \text{integer}, ename: \text{varchar}, sal: \text{integer}, age: \text{integer}, did: \text{integer})$

$\text{Dept}(did: \text{integer}, budget: \text{integer}, floor: \text{integer}, mgr_eid: \text{integer})$

Gehälter (sal) variieren zwischen \$10.000 und \$100.000, Budgets ($budget$) zwischen \$10.000 und \$1.000.000, Alter (age) zwischen 20 und 80. Ein Department hat im Durchschnitt 5 Angestellte und es gibt 10 Etagen ($floor$). Sie können von einer gleichförmigen Verteilung der Werte ausgehen.

Entscheiden Sie für die beiden folgenden Anfragen, welche der vorgegebenen Indexte sie wählen würden, um die Anfrage zu beschleunigen. Falls das Datenbank-System keine Index-Only Anfrage-Auswertung unterstützt (d.h. Dateneinträge werden immer abgefragt, auch wenn genug Informationen zur Beantwortung der Anfrage bereits im Index-Eintrag enthalten sind), würde sich Ihre Wahl verändern? Erläutern Sie jeweils Ihre Wahl.

- a) `SELECT ename, age, sal FROM Emp`
- (a) Unclustered Hash Index on $\langle ename, age, sal \rangle$ of Emp
 - (b) Clustered B⁺-Tree Index on $\langle ename, age, sal \rangle$ of Emp
 - (c) Clustered Hash Index on $\langle eid, did \rangle$ of Emp
 - (d) No Index
- b) `SELECT did FROM Dept WHERE floor = 10 AND budget < 15.000`
- (a) Clustered Hash Index on *floor* of Dept
 - (b) Clustered B⁺-Tree Index on $\langle floor, budget \rangle$ of Dept
 - (c) Clustered B⁺-Tree Index on *budget* of Dept
 - (d) No Index

Übung 5 (5 Punkte)

Fügen Sie die Suchschlüsselwerte (8, 5, 1, 7, 3, 12, 9, 6) in dieser Reihenfolge in einen zu Beginn leeren B⁺-Baum (3, 2) ein. Zeigen Sie den Baum nach jeder Einfüge-Operation.

Halten Sie sich dabei an folgende Vorgaben:

- Bei einem Überlauf und Aufteilen des entsprechenden Blattknotens enthält das linke Teilblatt anschließend einen Suchschlüsselwert mehr als das rechte Teilblatt.
- Bei einem Überlauf und Aufteilen des entsprechenden inneren Knoten wird der mittlere Suchschlüsselwert in den übergeordneten inneren Knoten verschoben. Falls kein übergeordneter innerer Knoten existiert (aufzuteilender Knoten ist der Wurzelknoten), so wird ein neuer Wurzelknoten erzeugt.

Übung 6 (4 Punkte: Bonuspunkte)

Sei ein B⁺-Baum charakterisiert durch (m, l) , $m > 2$, $l > 1$ und m gerade. m definiert die maximale Anzahl direkter Nachfolger eines Knotens und l die maximale Anzahl von Tupeln in einem Blatt des Baumes.

Angenommen in einen anfangs leeren B⁺-Baum (200, 31) werden nacheinander $n = 32.000.000$ Tupel in aufsteigender Suchschlüssel-Reihenfolge eingefügt. Bei einem Überlauf und Aufteilen des entsprechenden Blattknotens werden die Suchschlüsselwerte gleichmäßig über beide Knoten verteilt. Schätzen sie die benötigte Höhe des Baumes ab. Begründen Sie Ihre Abschätzung.