

## Lösungen der Übungsaufgaben von Kapitel 9

1. Formulieren Sie die Definition eines Indexes mit eigenen Worten. Was ist der Zweck der Anlage eines Indexes für ein Attribut einer Tabelle bzw. einer Datei?

Vergleichen Sie die entsprechende Definition am Ende von Abschnitt 9.3 Ein Index, der für ein spezielles Attribut angelegt wird, hilft durch Angabe der Record-Id dabei, den Datensatz, der zu einem Wert dieses Attributs gehört, schnell zu finden.

2. Was ist ein Primärindex, was ein Sekundärindex? Wie viele eindeutige Indizes gibt es für eine Tabelle mit 3 Attributen, von denen genau ein Attribut der einzige Schlüsselkandidat und damit auch der Primärschlüssel ist?

Ein Index für einen Primärschlüssel heißt **Primärindex**, Indizes anderer Attribute heißen dementsprechend **Sekundärindizes**. Indizes zu Schlüsselkandidaten heißen **eindeutige Indizes**. Damit gibt es in diesem Falle genau einen eindeutigen Index.

3. Beschreiben Sie den Unterschied zwischen einem dichten und einem lichten Index mit eigenen Worten. Wie viele lichte Indizes kann man grundsätzlich für eine Datei definieren? Warum ist das so?

Ein dichter Index für ein Attribut oder eine Attributkombination speichert für jeden Wert dieses Attributs bzw. dieser Attributkombination die Record-Id des zugehörigen Satzes. Bei einem lichten Index erfolgt diese Speicherung nur für diejenigen Sätze der Datei, die auf ihrer jeweiligen Seite als letzte Sätze stehen.

Man kann für eine Datei grundsätzlich nur **einen** lichten Index erstellen, denn ein solcher Index macht nur dann Sinn, wenn die Sätze dieser Datei in der Sortierung gemäß den Attributwerten dieses Index physikalisch gespeichert sind.

4. Die „Standard“-Hash-Funktion, die mit dem Primärschlüssel ***Id*** einer Tabelle arbeitet, lautet:  $\text{hash}(\mathbf{Id}) := \mathbf{Id} \bmod p$  für eine Zahl  $p$ . Geben Sie einen Grund dafür an, warum man für  $p$  grundsätzlich am liebsten Primzahlen wählt.

Der Grund ist ein mathematischer: Man kann  $\bmod p$  viel unproblematischer rechnen, wenn  $p$  eine Primzahl ist. Genauer finden Sie in jedem Buch über elementare Zahlentheorie. In meinem Buch „Mathematik für Informatiker“ diskutiere ich diese Frage genau.

5. Betrachten Sie noch einmal Abschnitt 5.2 aus dem 5. Kapitel. Ich behaupte: Das Hash-Verfahren ist ein weiterer Grund dafür, Primärschlüssel möglichst frei von Anwenderinformationen zu halten. Geben Sie für diese Behauptung eine sorgfältige Begründung.

- (i) Man wendet beim Hash-Verfahren die Hash-Funktion im allgemeinen auf den Primärschlüssel an.
- (ii) Die Hash-Funktion berechnet die Speicheradresse.
- (iii) Also bedeutet eine Änderung des Primärschlüssels eine Änderung der Speicheradresse und damit ein physikalisches Delete und ein anschließendes physikalisches Insert auf der Festplatte. Das kostet Zeit.
- (iv) Darum will man den Primärschlüssel so wenig wie möglich ändern.
- (v) Da sich aber Anwenderinformationen wie Namen, Bezeichnungen, Artikelnummern usw. völlig unabhängig von den Anforderungen der internen Datenbankverwaltung ändern können, möchte man sich beim Primärschlüssel davon unabhängig machen.

Datenbanken

Theorie, Entwurf und Programmierung relationaler

Datenbanken

Schubert, M.

2007, XII, 344 S. Mit Online-Extras., Softcover

ISBN: 978-3-8351-0163-0