

Aufgabe 1 Lineares Hashing

a) Nennen Sie je drei Vorteile und Nachteile des linearen Hashing.

Vorteile:

- Speicherallokation nur nach Bedarf (dynamische Verwaltung)
- im Regelfall schnelles Auffinden
- keine großen "Lücken" => gute Datenverteilung
- schneller Zugriff
- einfache Implementierung

Nachteile:

- "Datenballungen", wenn viele Daten den gleichen Hashwert ergeben, dadurch im schlimmsten Fall hoher Suchaufwand => evtl. schlechte Datenverteilung
- Split ist aufwendig
- Split findet unabhängig von Belegung statt

b) Entwerfen Sie je ein Szenario, für das sich der Einsatz des linearen Hashing eignen würde bzw. nicht eignen würde.

Geeignet für lineares Hashing:

- Datenzugriff über eindeutige Nummern (z.B. "sprechende Schlüssel")
- gleichmäßige Verteilung der Nummern
- keine häufigen Löschvorgänge

Ungeeignet für lineares Hashing:

- wenn die Auslastung der Buckets stark ungleichmäßig ist
- wenn auf Daten über ein nicht eindeutiges Attribut zugegriffen wird
- wenn häufig gelöscht wird
- bei Bereichsanfragen

Aufgabe 2 : Pflichtaufgabe - Lineares Hashing - Beispiel

Gegeben seien die Suchschlüssel 2, 3, 5, 7, 11, 13, ..., 97 (Primzahlen unter 100), die der Reihe nach mittels Hashing in eine Datei eingefügt werden sollen.

- a) Was ergibt sich bei (statischem) Hashing mit der Hash Funktion $h(x) = x \bmod 8$ bei einer bucket Größe von einer Seite, auf die jeweils 3 Datensätze passen?
- b) was ergibt sich bei linearem Hashing mit der Hash Funktionenfolge $h_i(x) = x \bmod (2i \cdot 4)$ wenn bei jedem Seitenüberlauf gesplittet wird? (beachten Sie, dass beim linearen Hashing nicht unbedingt die übergelaufene Seite geteilt wird!)

Aufgabe 3 : Pflichtaufgabe - physischer Entwurf

In unserer Bibliothek werden die Verwaltung der Bücher und der Ausleihvorgänge über eine Datenbank durchgeführt. Um eine größtmögliche Effizienz zu gewährleisten, sollen Zugriffspfade eingesetzt werden.

Tabellen:

Buch(Buchnr, Standort, Autor, Titel, Verlag)

AusleiherIn(Name, Vorname, Adresse, Mahngebühren)

Ausleihvorgang(Name, Vorname, Buchnr, Rückgabedatum)

Anforderungen:

Auf alle Bücher muß direkt über den Primärschlüssel zugegriffen werden (z.B. um

festzustellen, welche BenutzerIn ein Buch ausgeliehen hat).

Eingetragene Bücher werden fast nie gelöscht, Neuanschaffungen sind leider selten geworden.

Zu Beginn jedes Semesters kommen in der Regel neue AusleiherInnen hinzu bzw. werden alte aus der Tabelle entfernt.

Jeden Monat wird ein Bericht über die offenen Mahngebühren angefertigt, wobei als Anhang eine Liste der AusleiherInnen mit offenen Mahngebühren zu erstellen ist.

Für das Einfügen und Löschen von Mahngebühren muss natürlich auf jedes einzelne Tupel zugegriffen werden.

Ausleihvorgänge werden häufig erstellt und gelöscht. Insbesondere sind Einzelzugriffe über den Namen erforderlich(Welche Bücher hat xyz entliehen?) oder auch alphabetische Listen auszuwerfen, die alle aktuellen Ausleihvorgänge enthalten.

Es werden täglich Mahnungen für diejenigen Ausleihvorgänge geschrieben, die bereits 10 Tage überfällig sind.

Außerdem müssen häufig Auskünfte erteilt werden, wann ein bestimmtes Buch zurückgegeben werden wird, also wieder verfügbar ist.

Zur Steigerung der Effizienz stehen die Ihnen in der Vorlesung vorgestellten Zugriffsverfahren zur Verfügung.

a) Stellen Sie einen Kriterien - Katalog auf, mit dessen Hilfe Sie die verschiedenen Zugriffsverfahren bewerten können.

1. Effektivität des Primärschlüssel-Zugriffs
2. Effektivität des Sekundärschlüssel-Zugriffs
3. Effektivität bei Ausgabe sortierter Listen
4. Auswirkungen bei `insert`
5. Auswirkungen bei `delete`
6. Speicherplatzbedarf

b) Bewerten Sie drei ausgewählte Zugriffsstrukturen gemäß diesem Kriterienkatalog.

	Hashverfahren	B ⁺ -Bäume
1	optimal	sehr gut (konstanter Aufwand, Baumhöhe)
2	(lineare Suche)	(einfache Sache)
3	ungeeignet	optimal
4	sehr gut, einfach (wenig Aufwand, außer wenn Bereich zu klein)	hoher Aufwand (Änderungen bis zur Wurzel)
5	sehr gut, einfach (wenig Aufwand, außer wenn Bereich zu klein)	sehr gut (nur Blätter)
6	Hashtabelle	viel mehr als Hash

c) Machen Sie einen Vorschlag für die Primärorganisation sowie evtl. einzurichtende Sekundärzugriffspfade für die angegebenen Datenmengen

Tabelle	Merkmale	Empfehlung
---------	----------	------------

Buch	häufiger PS-Zugriff, wenig Bestandsveränderung	Hash
------	---	------

Ausleiher	viel Bewegung, häufiges Insert, Delete monatliche sortierte Liste monatliche Aktualisierung der Mahngebühren	B ⁺ -Baum
-----------	--	----------------------

Ausleihvorgang

viel Insert/Delete alphabetische Liste nach Namen häufiger Zugriff über Namen Selects auf Rückgabedatum+10 < heute Selects auf Join Buch, Ausleihvorgang	B ⁺ -Baum für Name
--	-------------------------------