Prof. Dr. T. Härder
Fachbereich Informatik
AG Datenbanken und Informationssysteme
Universität Kaiserslautern

Übungsblatt 3

für die freiwillige Übung

Unterlagen zur Vorlesung: „wwwdvs.informatik.uni-kl.de/courses/DBSREAL/“

Aufgabe 1: *Hot-Set* Modell

186

Es seien 3 Relationen R1, R2 und R3 mit 10, 20 bzw. 30 Seiten gegeben. Jede Seite beinhalte 10 Tupel. Zur Berechnung der beiden 1:n-*Joins* ((R1 |X| R2) |X| R3) werde ein *Nested-Loop-Join* benutzt. Die Ergebnisse der *Joins* werden in Zwischenrelationen mit 20 bzw. 30 Seiten abgelegt.

Welchen Verlauf nimmt die Fehlseitenhäufigkeit in Abhängigkeit von der Anzahl der verfügbaren Pufferrahmen, wenn für die Seitenersetzung das LRU-Verfahren gewählt wird?

Aufgabe 2: Rekursives Laden einer Tabelle beim DBCache-Ansatz

Tabelle CUST habe u.a. zwei UNIQUE-Spalten Cno und Cid sowie zwei NON-UNIQUE-Spalten CType und CLoc. Als Cache Keys seien CType und CLoc definiert. Die Belegung im BE sei

BE_CUST	Cno	CType	CLoc	Cid ...
	789	silver	SF	NULL
	891	silver	LA	d07
	333	platinum	SJ	a21
	444	unspec.	LA	a07
	123	gold	SJ	a14
	456	gold	SF	b21
	555	gold	NY	c17
	666	bronze	Chi	a49
	999	NULL	Chi	a54
	001	bronze	NULL	c18

a) Wie verläuft das erstmalige Laden bei FE-CUST, wenn Q1 das Prädikat (CType = ‘platinum’) auswertet?

b) Wo wird Q2 mit (Cid = ‘a21') ausgewertet?

c) FE-CUST sei leer. Wie sieht die Belegung nach Auswertung von Q2 mit (Cid = ‘a21’) aus?

d) Was bewirkt Anfrage Q3 mit (CLoc = ‘Chi’) bei der Belegung nach Ausführung von c)

Aufgabe 3: Laden einer CacheGroup

Tabelle CUST (C) habe u.a. zwei Spalten Cno und Cid vom Typ UNIQUE (U) sowie zwei Spalten CType und CLoc vom Typ NON UNIQUE (NU).

Tabelle ORDER (O) habe u. a. die U-Spalte Oid und die NU-Spalte Cno.

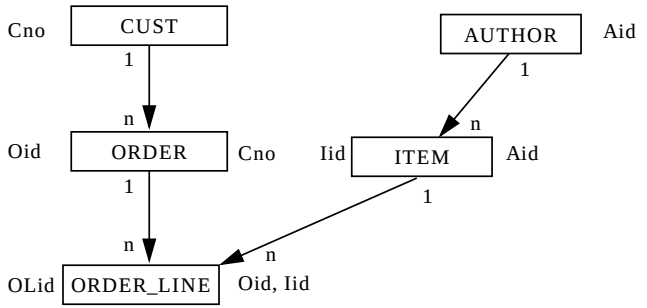
Tabelle ORDER_LINE (OL) habe u.a. die U-Spalte OLid und die NU-Spalten Oid und Iid.

Tabelle AUTHOR habe u.a. die U-Spalte Aid und eine NU-Spalte AName, während Tabelle ITEM die U-Spalte Iid und die NU-Spalte Aid besitzt.

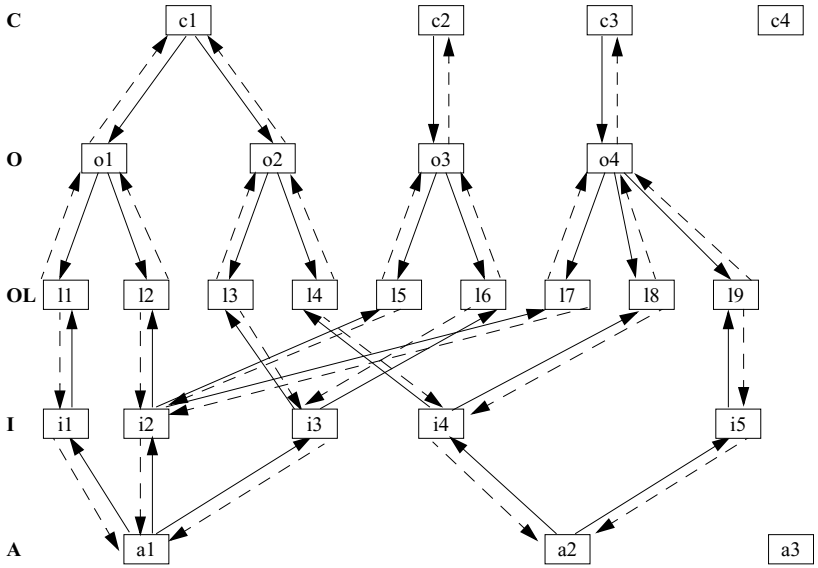
Im Backend-DB-Server (BE) seien für diesen Schemaausschnitt die referentiellen Integritätsbedingungen C.Cno → O.Cno und O.Oid → OL.Oid (als Primär-/Fremdschlüsselbeziehungen) definiert.

Weitere referentielle Integritätsbedingungen sind
A.Aid → I.Aid und I.Iid → OL.Iid.

Graphisch läßt sich dieser Schemaausschnitt wie folgt darstellen:



Ein Ausschnitt aus der BE-DB sei als folgende Graphstruktur veranschaulicht:



Der Pfeiltyp → veranschaulicht die (1:n)-Richtung und der Pfeiltyp - -> die (n:1)-Richtung der referentiellen Integritätsbedingung. Die Pfeile repräsentieren nur die Wertegleichheit von PS/FS oder FS/PS und sind nur eine verkürzende Darstellung im Vergleich zu Tabellen, die mit den entsprechenden Werten belegt sind.

Die Primärschlüsselwerte in C.Cno seien 123, 456, 789, 012 für c1, c2, c3 und c4.

c1 habe den Wert ‘gold’ und c3 den Wert ‘platinum’ in C.CType, während c2 und c4 den Wert ‘silver’ in C.CType besitzen.

Die Primärschlüsselwerte in A.Aid seien ‘a47’ in a1, ‘a53’ in a2 und ’a62’ in a3.

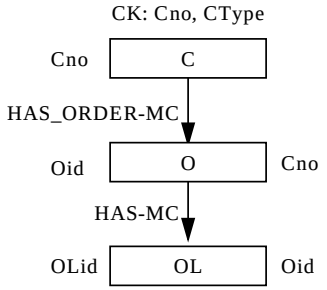
a1 habe den Wert ‘mayr’ in A.AName, während a2 und a3 ‘codd’ in A.AName haben.

Im Frontend-DB-Server (FE) sei die Wurzel-Tabelle FE-C einer CacheGroup definiert durch die Cache Keys C.Cno und C.CType. Weiterhin habe Tabelle FE-A einen Cache Key A.AName. Als RCCs (referential cache constraints) lassen sich beispielsweise C.Cno → O.Cno, O.Oid → OL.Oid, A.Aid → I.Aid und I.Iid → OL.Iid spezifizieren. Diese RCCs sind vom Type U → NU und werden auch als Member-Constraints (MC) bezeichnet. Solche MCs gewährleisten, daß für jeden Owner einer solcher (1:n)-Beziehung alle zugehörigen Member sich im Cache befinden. Die dazu umgekehrte Beziehung vom Typ NU → U heißt auch Owner-Constraint (OC), was besagt, daß, sobald sich ein Member dieser Beziehung im Cache befindet, auch der zugehörige Owner sich im Cache befinden muß.

Beispiel: Die RCCs C.Cnr → O.Cno, O.Oid → OL.Oid usw. sind Member-Constraints und können mit HAS_ORDER-MC bzw. HAS-MC bezeichnet werden. Die RCCs O.Cno → C.Cno und OL.Oid → O.Oid dagegen sind Owner-Constraints und lassen sich mit HAS_ORDER-OC bzw. HAS-OC bezeichnen.

Im folgenden gehen wir immer von einem leeren Cache in der FE-DB aus.

a) Die CacheGroup G1 sei wie folgt definiert:

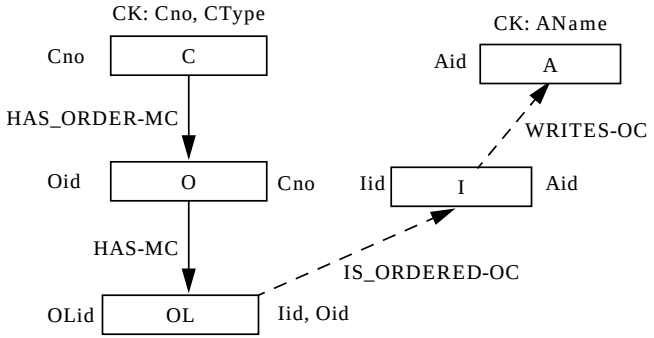


In Anfrage Q1 werde das Prädikat (Cno = 123) verwendet. In einer zweiten Anfrage Q2 wird nach (CType = ‘silver’) gesucht.

Welche Belegung von G1 stellt sich nach Ausführung beider Anfragen ein?

Geben Sie Beispiele für Verbundprädikate, die in der FE-DB ausgewertet werden können.

b) CacheGroup G2 sei wie folgt definiert:

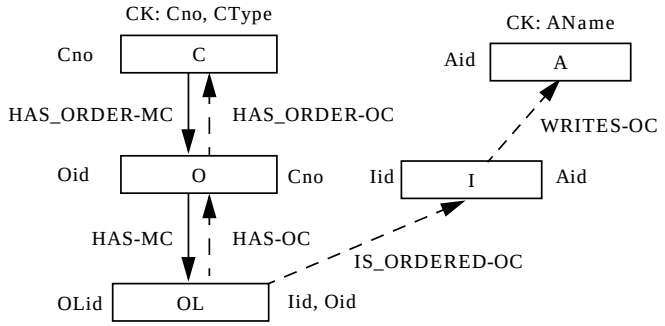


Q3 referenziert (C.CType = ‘gold’). Wie sieht die Belegung von G2 nach Ausführung von Q3 aus?

In dieser Situation referenziert Q4 (A.AName = 'codd').

Was ändert sich an der Belegung von G2?

c) CacheGroup G3 sei wie folgt definiert:

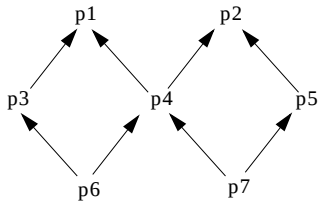


Wie sieht die Belegung von G3 nach Ausführung von Q3 und Q4 aus?

Aufgabe 4: Rekursives Laden von CacheGroups

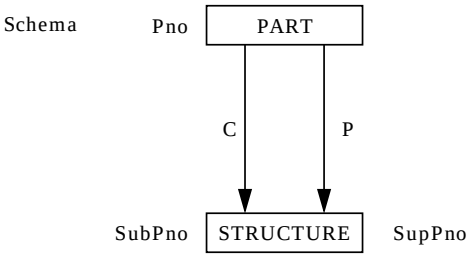
Das Lösen dieser Aufgabe ist **preiswürdig!**

Der Gozinto-Graph GG1 einer Stückliste sei



Die Pfeilrichtung entspricht der Beziehung PART_OF

Im Relationenmodell läßt sich eine Stückliste wie folgt modellieren, wobei ein Pfeil eine PS/FS-Beziehung (referentielle Integritätsbedingung) auf Typebene ausdrückt:



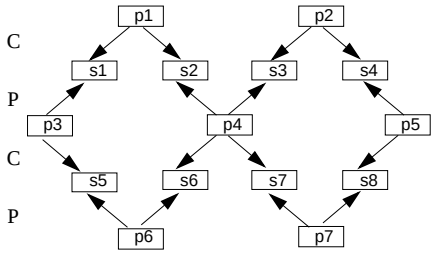
C = COMPOSED_OF

P = PART_OF

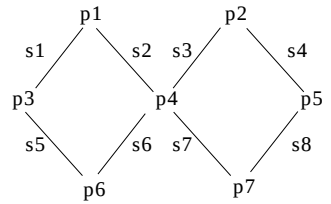
Gemäß dieses Schemas sei die obige Stückliste GG1 in der Backend-DB (BE-DB) in den Tabellen BE-PART und BE-STRUCTURE gespeichert.

Für GG1 wählen wir eine Graphdarstellung, die sich als kompakte Alternativform der Tabellendarstellung für PART und STRUCTURE auffassen läßt. GG1 läßt sich nach obigem Schema folgendermaßen veranschaulichen, wobei jeweils die PS/FS-Beziehungen zwischen den Sätzen (Tupel) als Pfeile dargestellt sind. Die Kantenbezeichnungen wurden nach folgender Darstellung gewählt:

GG1 als Graph
mit PS/FS-Beziehungen für C und P



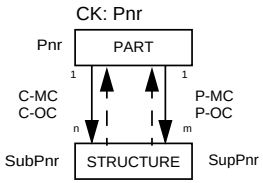
Kantenbezeichnungen



Im Cache sei der Cache Key Pno für FE-PART definiert.

Das Laden des Cache in der FE-DB beginnt immer bei leeren Tabellen FE-PART und FE-STRUCTURE.

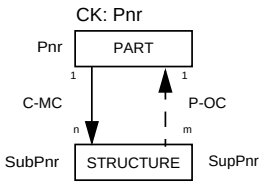
a) Gegeben sei folgende Definition der CacheGroup G1:



Wie sieht die Belegung von G1 aus, wenn in Anfrage Q1 'Pno = 1' (p1) referenziert wurde?

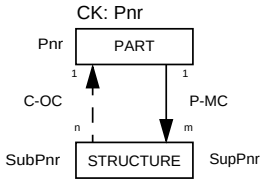
Welche Belegung wäre aus einer Referenz von p4 ('Pno = 4') oder p6 ('Pno = 6') resultiert?

b) Gegeben sei die nachfolgende Definition von G2:



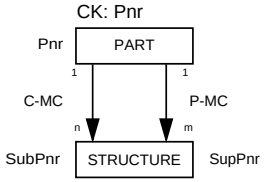
Wie sieht die Cache-Belegung aus, wenn ‘Pno = 1’ (p1) referenziert wurde?

c) Gegeben sei die nachfolgende Definition von G3:



Wie sieht die Cache-Belegung aus, wenn ‘Pno = 6’ (p6) referenziert wurde?

d) Gegeben sei die nachfolgende Definition von G4:



Wie sieht die Cache-Belegung aus, wenn ‘Pno = 4’ (p4) referenziert wurde?

e) Was ist eine notwendige Bedingung, daß in einer CacheGroup G rekursives Laden vermieden wird?

Aufgabe 5: Direkte und Indirekte Satzadressierung 12

a) Vergleichen Sie den Speicherplatzbedarf für folgende Adressierungsprinzipien:

1. logische Byte-Adresse
2. TID-Konzept
3. DBK-Konzept
4. DBK/PPP-Konzept.

Die Ausprägungen eines zu adressierenden Satztyps seien jeweils auf ein Segment der Größe 2^{32} Byte beschränkt. Es können mehrere Satztypen in einem Segment gespeichert sein. Welche Faktoren werden durch die einzelnen Adressierungskonzepte bei gegebener Pointerlänge begrenzt:

N_{REC} : Anzahl der Sätze,
S_l : logische Satzlänge,
L_s : Seitengröße,
m_k : Anzahl der Seiten (im Segment)?

Wie hoch ist der gesamte Speicherplatzbedarf für die Adressierung eines Satzes, wenn n verschiedene Pointer auf den Satz zeigen? Die Anzahl der Überläufer bei TID-Konzept liege bei 10%.

b) Schätzen sie den Zugriffsaufwand bei den verschiedenen Adressierungskonzepten ab. Wieviele physische E/A-Zugriffe sind bei einer Folge wahlfreier Zugriffe über einen Indexbaum mit 3 Stufen im Mittel zu erwarten, bis der gesuchte Datensatz gefunden ist,

1. beim TID-Konzept
2. beim DBK-Konzept
3. beim DBK/PPP-Konzept?

Der DB-Puffer besitze 40 Seitenrahmen zu 4 K Byte; die Ersetzungsstrategie sei LRU. Es seien $N_{REC}=10^5$ Datensätze über 10^4 Seiten verteilt. Der Indexbaum habe neben der Wurzel 9 Seiten auf der 2. Indexstufe und 1000 Seiten auf der 3. Indexstufe. Die Zuordnungstabelle bestehe aus Seiten der Größe 4 KB und besitze PP-Einträge der Länge 4 Byte. Die Anzahl der Überläufer sowie die Anzahl der falschen PPPs sei 10%.

Vergleichen sie die Anzahl der Zugriffe, wenn über die interne logische Adresse (TID, DB-Key) direkt zugegriffen wird.

c) Ermitteln Sie die Anzahl der physischen Zugriffe bei der fortlaufenden Verarbeitung aller Datensätze über eine Indexstruktur, wenn mit

1. dem TID-Konzept
2. dem DBK-Konzept
3. dem DBK/PPP-Konzept

adressiert wird. Welche Ergebnisse erhalten Sie unter den Annahmen von b) im *best case* (Clusterbildung) und im *worst case*?