

Prof. Dr.-Ing. Dr. h. c. T. Härder
 Fachbereich Informatik
 AG Datenbanken und Informationssysteme
 Universität Kaiserslautern
 www.haerder.de

Übungsblatt 10

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

Aufgabe 1: Normalisierung, Vereinfachung, Restrukturierung

Gegeben seien folgende Tabellen:

Personal (Pnr, Pname, Beruf, Gehalt)
 Projekt (Pnr, Pname, Probudget)
 PMitarbeit (Pnr, Pnr, Dauer)

- Bestimmen Sie für die Qualifikationsbedingung der folgenden Anfragen die konjunktive und disjunktive Normalform:

```
Select * From Personal
Where (Pname Like 'M%' And Beruf = 'Techniker') Or
((Pnr > 550 Or Beruf = 'Programmierer') And Gehalt < 80000)
```

- Vereinfachen Sie die Qualifikationsbedingungen der folgenden Anfrage durch Anwendung der Idempotenzregeln:

```
Select * From Personal
Where Pnr > 456 And
Not (Beruf = 'Techniker' Or Gehalt < 50000) And
Beruf ≠ 'Techniker' And Gehalt < 50000
```

- Führen Sie die Anfragetransformation für folgende Query durch.

```
Select Pname, Pname
From Personal P, Projekt Pt, PMitarbeit Pm
Where Dauer > 10 and P.Pnr = Pm.Pnr and
Beruf = 'Programmierer' and Pt.Pnr = Pm.Pnr
```

Bestimmen Sie den Operatorbaum und führen darauf Vereinfachungen und Restrukturierungen zur algebraischen Optimierung durch.

Aufgabe 2: Einfache Join-Strategien

Sei Card (R) = 10000, Card (S) = 1000, JSF (R ⋈ S) = 0.001. Jede Tabelle soll 5 Attribute umfassen. Welche Kommunikationskosten ergeben sich für 'Ship Whole' bzw. 'Fetch as needed' bei Join-Ausführung an K_R bzw. an K_S ?

Aufgabe 3: Ship-Whole vs. Semi-Join vs. Bitvektor- Join

Auf den Tabellen Personal und PMitarbeit seit folgende Join-Query zu bearbeiten:

```
Select P.Pnr, Pname, Beruf, Pnr, Dauer
From Personal P, PMitarbeit Pm
Where P.Pnr = Pm.Pnr And P.Gehalt > 60000
```

Es gelte Card (Personal) = 1000, Card (PMitarbeit) = 1500; beide Tabellen seien an verschiedenen Knoten gespeichert. Die Anfrage soll an einem dritten Knoten K initiiert werden; das Ergebnis ist dort auch auszugeben. Die Gehaltsbedingung soll von 20 % der Angestellten erfüllt werden (SF = 0.2); 25% der Angestellten sollen in keinem Projekt mitarbeiten.

Bestimmen Sie die Kommunikationskosten (#Nachrichten, #AW (Anzahl zu übertragender Attributwerte)) für folgende Join-Strategien:

- Ship-Whole; Join-Berechnung an Knoten $K_{PMitarbeit}$
- Ship-Join; Join-Berechnung an Knoten K
- Semi-Join; Join-Bestimmung an Knoten $K_{Personal}$
- Semi-Join; Join-Berechnung an Knoten K
- Bitvektor-Join; Join-Berechnung an Knoten K

Vor der Übertragung sollen alle anwendbaren Selektionen und Projektionen durchgeführt werden. Die Länge des Bitvektors soll 5 Attributwerten entsprechen; durch Anwendung des Bitvektors soll sich die zurückzuliefernde Tupelanzahl um 5% erhöhen.

Aufgabe 4: Optimierung von mengenorientierten Anfragen mit Hilfe von Cluster-Bildung

Gegeben sei eine Tabelle Personal mit Attributen für die Personalnummer, den Namen des Angestellten, das Gehalt sowie den Manager bei dem der Mitarbeiter beschäftigt ist. Daneben ist noch weitere, hier nicht relevante Information enthalten. Insgesamt ergibt sich Personal (Pnr, Name, Gehalt, Mnr, ...), wobei die Datensätze zwischen 512 und 1024 Bytes groß sind. Die genannten Attribute haben die Größen: Pnr = 2 Byte, Name = 30 Byte, Gehalt = 4 Byte. Im Durchschnitt hat jeder Manager ca. 10 Mitarbeiter. Die Seitengröße des DBVS sowie die Slotgröße der Dateien auf dem Externspeicher betragen 4kB. Weiterhin existiere eine Indexstruktur $I_{Personal}(Pnr)$.

Folgende Anfrage soll unterstützt werden:

- Select P1.Name, P1.Gehalt P2.Gehalt
From Personal P1, Personal P2
Where P1.Mnr = P2.Pnr AND P1.Gehalt > P2.Gehalt

Untersuchen Sie die Frage, ob Cluster-Bildung der Tabelle Personal bezüglich der Mnr sinnvoll ist. Wenden Sie dabei für den Verbund-Algorithmus Nested Loops an und diskutieren Sie den Einfluß des DB-Puffers auf das E/A-Verhalten, wenn minimal nur 5 und maximal 200 Pufferrahmen zur Verfügung gestellt werden. Berücksichtigen Sie dabei auch die Größe N der Tabelle (z. B. $N=10^3$ und $N=10^5$ Tupel in Personal).

Welchen Einfluß auf die Aktualisierungskosten hat eine Cluster-Bildung über Mnr (Anpassungen von $I_{\text{Personal}}(\text{Pnr})$ werden nicht betrachtet), wenn

- die Abteilung einen neuen Manager bekommt? Hier muß die Mnr in allen Sätzen eines Clusters auf die Pnr des neuen Chefs geändert werden.
- ein Mitarbeiter in eine neue Abteilung wechselt?