

# 1. Übungsblatt

Ausgabe: 15. Oktober 2013 · Besprechung: 23. Oktober 2013

## 1 Buffer Manager — Verdrängungsstrategien

In der Vorlesung wurde die Funktionsweise des *Buffer Managers* diskutiert. Dabei ist bei der Wahl eines zu überschreibenden *Frames* eine *Verdrängungsstrategie* notwendig.

Die wohl bekannteste Verdrängungsstrategie ist LRU (“Least Recently Used”). In Datenbanksystemen werden jedoch oft auch andere Strategien verwendet, etwa

- “clock” und
- “LRU-k” (*LRU-Variante, wobei der k-letzte Zugriff berücksichtigt wird*).

Wie funktionieren diese beiden Verdrängungsstrategien? Welche Vor- und Nachteile haben sie jeweils gegenüber einer reinen LRU-Strategie?

## 2 PostgreSQL

Wir werden uns später noch das Verhalten von PostgreSQL anschauen, dessen Implementation oft ein gutes Beispiel ist für die in der Vorlesung diskutierten Techniken.

1. Laden Sie den Sourcecode der aktuellen PostgreSQL-Version auf Ihren Rechner. Anschließend kompilieren und installieren Sie ihn.
2. In eine leere PostgreSQL-Datenbank laden Sie Daten des TPC-H-Benchmarks.  
TPC-H ist der wohl bekannteste Benchmark für “decision support”-Aufgaben. Auf der Webseite des TPC-H-Konsortiums finden Sie ein Paket, das Generatoren für jeweils Daten (**dbgen**) und Anfragen (**qgen**) enthält.
3. Führen Sie die TPC-H-Anfragen aus und messen Sie jeweils die Laufzeit.

## Hinweise:

- Anleitungen dazu finden Sie im Internet.
- Der Datengenerator **dbgen** erzeugt ein Format, in dem jede Zeile mit einem 'I' endet. Dieses kann mit dem **COPY**-Kommando von PostgreSQL nicht korrekt eingelesen werden. Entfernen Sie daher vor dem Laden diese 'I' am Zeilenende (z. B. mittels **sed -i 's/I\$/g'**).
- In den Anfragen zum Benchmark (Verzeichnis **queries**) müssen einige Klauseln umformuliert werden, die sich auf Zeitintervalle beziehen. PostgreSQL versteht hier z. B. die Formulierung

```
l_shipdate <= date '1998-12-01' - interval '118 day'
```

- Verwenden Sie bitte keine zu großen Benchmark-Instanzen. Ohne explizites Anlegen von Datenbank-Indizes können sonst sehr lange Laufzeiten entstehen und/oder die Auswertung viel Platten- und Hauptspeicher benötigen.