

10.01.2023

## Übungsblatt 10

### Discussion Part (solve in the PS; no submission needed)

- a) ☐ ★ Diskutieren Sie, ob ein Datenbanksystem für alle möglicherweise auftretenden Anfragen optimiert werden kann und begründen Sie Ihre Antwort. Falls nein: bezüglich welcher Abfragen sollte ein Datenbanksystem optimiert werden? Wie finden Sie diese Abfragen?
- b) ☐ ★★ Man spricht oft davon, dass eine Query direkt “aus dem Index beantwortet werden kann” (“index-only”). Was bedeutet das für die Query-Abarbeitung und für die Performance der Datenbank? Welche Bedingungen müssen dafür erfüllt sein?

### Homework Part (solve at home; submission required)

#### Exercise 1 (Tooling)

[5 Points]


Ziel dieser Aufgabe ist es, Ihnen zu zeigen, welche Tools Ihnen PostgreSQL zur Verfügung stellt, um Ihnen beim Tunen einer Datenbank und beim Lösen von Problemen behilflich zu sein. Die Tools, die wir hier behandeln, gehören allesamt zum sogenannten Statistics Collector von PostgreSQL. Beginnen Sie daher, indem Sie sich die Dokumentation<sup>1</sup> dazu durchlesen. Falls nötig, konfigurieren Sie Ihr System so, dass es Statistiken sammelt.

- a) ☐ 2 Points Der Leiter der Finanzabteilung kommt zu Ihnen — dem Datenbankadministrator Ihres Arbeitgebers — und informiert Sie darüber, dass das Buchhaltungssystem “hängt”. Bei jeder Operation, die Buchungen aus der Datenbank auslesen bzw. Buchungen in diese schreiben soll, reagiert das System nicht mehr. Andere Operationen, wie zum Beispiel das Auslesen von Kundendaten, laufen einwandfrei.

Was könnte hier die Ursache sein? Sehen Sie sich die View `pg_stat_activity` an. Welche Informationen können Sie aus der Ausgabe dieser View herauslesen? Können Sie diese View verwenden, um das Problem zu lösen? Wie würden Sie vorgehen?

Abgabe



 1a.pdf

<sup>1</sup><https://www.postgresql.org/docs/current/monitoring-stats.html>

- b) **3 Points** Ihr Chef beauftragt Sie damit, die aktuelle Indizierungsstrategie für eine von Ihrem Arbeitgeber betriebene Datenbank zu evaluieren. Sie sollen also ermitteln, ob die aktuell vorhandenen Indizes ihren Zweck erfüllen, ob einige davon eventuell gelöscht werden könnten und ob Indizes fehlen.

Welche Views, die Ihnen der Statistics Collector zur Verfügung stellt, können Sie hier verwenden? Wie würden Sie vorgehen?

#### Abgabe



1b.pdf

## Exercise 2 (Query-Tuning)

**[5 Points]**

In dieser Aufgabe werden Sie verschiedene Möglichkeiten anwenden, eine gegebene Abfrage zu optimieren. Wir werden uns dabei auf zwei Möglichkeiten konzentrieren:

- Optimieren einer Abfrage, indem wir sie anders formulieren (Stichwort Rewriting).
- Optimieren einer Abfrage, indem wir die physische Datenstruktur der Datenbank so ändern bzw. erweitern, dass die Abfrage schneller beantwortet werden kann (Stichwort Indizes).

Wir werden bei den folgenden Aufgaben zum Teil mit der Pagila Datenbank arbeiten. Da in dieser Datenbank allerdings schon recht viele Indizes definiert sind, müssen Sie diese zuerst löschen, damit Sie die folgenden Aufgaben richtig bearbeiten können. Führen Sie dazu in Ihrer Pagila Datenbank folgende Query aus:

```
1  DROP INDEX
2      film_fulltext_idx,
3      idx_actor_last_name,
4      idx_fk_address_id,
5      idx_fk_city_id,
6      idx_fk_country_id,
7      idx_fk_customer_id,
8      idx_fk_film_id,
9      idx_fk_inventory_id,
10     idx_fk_language_id,
11     idx_fk_original_language_id,
12     idx_fk_payment_p2017_01_customer_id,
13     idx_fk_payment_p2017_01_staff_id,
14     idx_fk_payment_p2017_02_customer_id,
15     idx_fk_payment_p2017_02_staff_id,
16     idx_fk_payment_p2017_03_customer_id,
17     idx_fk_payment_p2017_03_staff_id,
18     idx_fk_payment_p2017_04_customer_id,
19     idx_fk_payment_p2017_04_staff_id,
20     idx_fk_payment_p2017_05_customer_id,
21     idx_fk_payment_p2017_05_staff_id,
```

```

22     idx_fk_payment_p2017_06_customer_id,
23     idx_fk_payment_p2017_06_staff_id,
24     idx_fk_staff_id,
25     idx_fk_store_id,
26     idx_last_name,
27     idx_store_id_film_id,
28     idx_title,
29     idx_unq_manager_staff_id,
30     idx_unq_rental_rental_date_inventory_id_customer_id

```

Bearbeiten Sie dann die folgenden Aufgaben.

a) 2 Points Gegeben sei die folgende Abfrage (auf der bekannten Pagila-Datenbank):

```

1  SELECT first_name, last_name, count(film_id)
2  FROM film_actor
3  INNER JOIN actor
4  ON actor.actor_id = film_actor.actor_id
5  WHERE film_id != ALL(SELECT film_id FROM film WHERE length < 120)
6  GROUP BY first_name, last_name
7  ORDER BY first_name, last_name

```

Diese Abfrage wird von PostgreSQL relativ ineffizient abgearbeitet. Finden Sie eine äquivalente Abfrage, welche das gleiche Ergebnis liefert und effizienter ausgeführt wird. Messen Sie mithilfe Ihres Client-Tools (etwa pgAdmin), wie lange beide Varianten für die Ausführung benötigen. Sehen Sie sich weiters die Ausführungspläne der beiden Varianten an und versuchen Sie, zu erklären, warum eine Variante schneller ist als die andere. Schreiben Sie Ihre Überlegungen und alle zugehörigen Materialien (z.B. die Ausführungspläne) in einem PDF-Dokument zusammen.

#### Abgabe



2a\_faster.sql

2a\_explanations.pdf

b) 3 Points Gegeben sei die folgende Abfrage. Diese müssen Sie vorerst nicht ausführen — wir machen erst eine theoretische Betrachtung.

```

1  SELECT      customer_id,
2              last_name,
3              first_name
4  FROM        customer
5  ORDER BY    last_name ASC

```


Nehmen Sie an, die customer-Tabelle enthält 1.000.000 Zeilen. Nehmen Sie weiters an, dass auf der Tabelle keine Indizes definiert sind. Welche(n) Index/Indizes können Sie definieren, damit die Ausführung dieser Query eventuell beschleunigt wird?


Führen Sie nun die Abfrage auf der Pagila-Datenbank aus, messen Sie, wie lang die Ausführung dauert, und sehen Sie sich den Ausführungsplan an (Speichern nicht vergessen —

der Ausführungsplan muss in der Abgabe enthalten sein!). Erstellen Sie anschließend in der Pagila-Datenbank den Index bzw. die Indizes, welche(n) Sie vorhin entworfen haben. Starten Sie PostgreSQL neu (um eventuelle Einflüsse durch Caching zu vermeiden) und führen Sie die Abfrage erneut aus. Messen Sie die benötigte Zeit und speichern Sie den Ausführungsplan. Gibt es einen Unterschied? Wurde(n) der Index/die Indizes verwendet? Wenn ja, wie? Wenn nein, welchen Grund könnte das Ihrer Meinung nach haben? Recherchieren Sie und schreiben Sie Ihre Überlegungen und alle zugehörigen Materialien in einem PDF-Dokument zusammen.

#### Abgabe



 2b\_index.sql

 2b\_explanations.pdf

**Important:** Submit your solution to OLAT and mark your solved exercises with the provided checkboxes. The deadline ends at 23:59 on the day before the discussion.