

5. Übungsblatt

Ausgabe: 3. November 2014 · Besprechung: 17. und 19. November 2014

1 Praktischer Einsatz von Indexen

Die Antwortzeit von Anfragen kann oft durch den Einsatz von Indexstrukturen beschleunigt werden, da Indexe die Anzahl der zu lesenden Seiten signifikant reduzieren können. In dieser Aufgabe sollen TPC-H-Anfragen auf Postgres optimiert werden. Dazu sollten folgende Schritte durchgeführt werden.

1. Bestimmen Sie für jede Anfrage des TPC-H Benchmarks die Laufzeit auf Ihrer Datenbank. Welche Anfragen dauern am längsten?
2. Postgres bietet die Möglichkeit, über das Kommando `explain <SQL Anfrage>` bzw. `explain analyze <SQL Anfrage>` den genutzten Anfrageplan anzuzeigen. Bei `explain analyze` werden zusätzlich die realen Kosten angezeigt. Analysieren Sie für alle lang laufenden Anfragen die Anfragepläne. Für welche Anfragen würden sie eine Performanzverbesserung durch Indexe erwarten?
3. Erzeugen Sie Indexe auf ihrer Postgres Datenbank. Wie wirken sich die neuen Indexe auf die Performanz aus? (Hinweis: Sie müssen nach Anlegen der Indexe die Postgres Statistik aktualisieren, damit der Anfrageoptimierer Indexpläne betrachtet.)

2 Curse of Dimensionality

1. Welches Problem ist mit „Curse of Dimensionality“ gemeint?
2. Was bedeutet das für Datenbanken in der Praxis?
3. Schreiben Sie ein Programm, das zufällig Vektoren in einem n-dimensionalen Raum generiert. Der Wertebereich jedes Dimensionswertes soll zwischen 0 und 9 liegen. Nehmen Sie für die Werteverteilung eine Gleichverteilung an. Erzeugen Sie wiederholt Paare von zufälligen Vektoren, und berechnen Sie den Euklidischen Abstand. Wie ändern sich Mittelwert, Varianz und Standardabweichung der Distanzen zwischen den Vektoren für 3, 10, 100, und 1000 Dimensionen?