

Assignment 2

Query Tuning

Database Tuning

Due date: April 09, 2018, 23:55

Grading: 5 points

Notes

- It is suggested that you also have a look at the report template before you start working on the assignment.

Access Parameters for PostgreSQL

- Host: `biber.cosy.sbg.ac.at`
- Port: 5432
- Database: `dbtuning_ss2018`
- User/Password: you should have received them via email

The database server (`biber.cosy.sbg.ac.at`) is accessible only from inside the university network. If you would like to work from home, please connect to `fanny.cosy.sbg.ac.at` via `ssh`. Java and the PostgreSQL client as well as Python are installed on this machine.

Support

If there are any ambiguities or problems of understanding regarding the assignment, you have the following possibilities to clarify them:

- Slack channel `#dbt`¹ (preferred way of communication)
- Office hours: Wednesday, 10am - 11am, Office 0.26 (ground floor)

In this assignment you will gain hands-on experience in rewriting slow queries and in experimentally evaluating the rewritten queries.

Creating Tables and Indexes

Create a database with the following database schema:

- `Employee(ssnum, name, manager, dept, salary, numfriends)`
 - unique index on `ssnum`
 - unique index on `name`
 - index on `dept`

¹<https://dbteaching.slack.com>

- Student (ssnum, name, course, grade)
 - unique index on `ssnum`
 - unique index on `name`
- Techdept (dept, manager, location)
 - unique index on `dept`
 - a manager may manage multiple departments
 - a location may contain multiple departments

Populating the Tables

Fill the database with 100k employees, 100k students, and 10 technical departments. Only about 10% of the employees are in a technical department. The types of the attributes should make sense (e.g., `ssnum` should be an integer), but the values need not be meaningful (e.g., names can be random strings).

Queries

Original Queries Choose two types of queries that might be hard for your database to optimize. Taking queries from the lecture notes is OK, but do not choose a query where the original query would use a temporary table.

Note: For at least one of your queries rewriting should make a difference.

Rewritten Queries and Execution Plans Rewrite the queries and consult the execution plans of the original and the rewritten query. The rewritten query must lead to the same output across all possible instances.

Experiments Run the original and the rewritten query and measure the runtimes.

Report

1. Describe your database instance (data types, how did you fill the tables?).
2. Give the original and the rewritten queries.
3. Show and explain the execution plans, i.e., the output of `EXPLAIN ANALYZE`. Graphical versions (e.g., `pgAdmin3` output) is not enough.
4. Report and briefly discuss the runtime results from your experiments.

Please indicate the average time per group member that was spent solving this assignment. The time that you indicate will have *no* impact on your grade.

Grading scheme:

Category	Max. Points
Description of your setup	1
Original query 1: description of query plan	0.5
Rewritten query 1: description of query plan	0.5
Difference between original and rewritten query 1	0.5
Original query 2: description of query plan	0.5
Rewritten query 2: description of query plan	0.5
Difference between original and rewritten query 2	0.5
Interpretation of the runtime results	1

Important: If the grading scheme is unclear, ask the lecturer!