

Database Tuning, ITU, Spring 2007

Rasmus Pagh

March 14, 2007

Project – deliverable 3

Deadline (strict): April 2, 23.59 Danish time.

Time plan for deliverable 3

March 15. Description, question/answer session.

March 20 and March 27. Office hours for project supervisor (Milan Ruzic).

April 2. Hand-in by e-mail (pdf file) to `milan@itu.dk`

April 11 or 12. Feedback meetings.

Purpose

The purpose of this deliverable is to gain experience in analyzing query plans, and applying methods for tuning queries.

Overview

In this deliverable you continue your work on the VXL database. Your first task is to look at and understand Oracle's query plan for each query. You should then consider the possibility of:

- Improving the available access methods (e.g., by making additional indexes, partitioning or denormalizing),
- rewriting queries,
- updating statistics,
- . . . and other ways of influencing the query plan.

For example, if a B-tree index is used only for an index nested loop join, consider replacing it with a hash index.

We provide you with a program skeleton for loading additional data into the database, and for generating queries to the database. You will be working with the 10 parameterized queries from the second deliverable, plus some new queries found below.

You can download the new skeleton program from the ITU file system: `H:/milan/DBT-files` or `/import/home/milan/DBT-files`. Milan will be presenting the changes on March 15, and is the person to ask questions regarding these programs.

New queries

11. Find the total number of hours worked per week by drivers in the region named *s*.
`public int queryHours(String region)`
12. Find the number of customers that has not sent a parcel after date *d*.
`public int queryNotRecentCustomers(Date d)`
13. Find all parcels sent from postal code *p1* to postal code *p2*
`public ResultSet queryParcelFromTo(int p1, int p2)`
14. Find all vehicles that can carry more than 1000 kilograms, *and* have driven less than 50000 kilometers.
`public ResultSet queryGoodVehicles()`
15. Find the newest (most recently inserted) schedule of the vehicle with id *vid*.
`public ResultSet queryLatestSchedule(int vid)`
16. Find all the vehicles the were bought before date *d* *or* have driven more than *k* kilometers.
`public ResultSet queryOldVehicles(Date d, int k)`
17. Find the name of every driver who transported a parcel for the customer at address (*sid,house-no*) on date *d*.
`public ResultSet queryCustomerDrivers(int sid, int no, Date d)`
18. Find, for every customer, the price and delivery address of the most expensive parcel sent. Report the result sorted by price in decreasing order.
`public ResultSet queryBiggestCustomers()`

To be handed in

A pdf file with the following:

- Your data model (indicating any changes from the second deliverable).
- All 18 SQL queries (indicating any queries from the second deliverable that you revised).
- A description of the indexes you created (attributes and type).
- For every query, a description of the query plan chosen by Oracle. You should not copy the plan from the browser window, but only state (in words) the most important choices made: What indexes are used, in what order, what join methods are used, etc.
- For every query, a brief discussion of the query plan: What are the competing alternatives (if any)? Have any of them been tried, and what were the results? Is there any change that could make this query faster, but was not chosen because of other constraints?