Database Tuning, ITU, Spring 2008

Rasmus Pagh

March 11, 2008

Project – deliverable 3

Deadline (strict): April 10, 23.59 Danish time.

Time plan for deliverable 3

March 25. Description, question/answer session.

April 1 and April 8. Office hours for project supervisor (Milan Ružić).

April 10. Hand-in by e-mail (pdf file) to `milan@itu.dk`

April 15. 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,

- aggregate maintenance,

- . . . 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. For more info see Milan's e-mail to the course mailing list on March 6. You will be working with the 10 parameterized queries from the second deliverable, plus some new queries found below.

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 $\hat{d}$.
    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) entry of the schedule of the bookable vehicle with id `vid`. This is the one with the highest value of schedule_id.
    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 picked up a parcel ordered by some some customer with ID `cid` on date `d`.
    public ResultSet queryCustomerDrivers(int cid, 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?