Assignment 3: Indexing

Goal:

In this assignment you study the response times of search queries in big databases and learn to create indexes which make repeating queries faster.

Instructions:

This is an individual assignment. Try to do at least the first first five tasks. For full points, also complete Task 6.

To complete the assignment, you need to have MariaDB installed on your laptop. You can download MariaDB Community Server from https://mariadb.com/downloads/.

Tasks:

Tasks 1 to 5 can be done with **HeidiSQL**. It is a database editor that comes with MariaDB. You can also download it separately from https://www.heidisql.com/.

1. Download a database script named **big.sql** from the **Documents/Data** folder in the course's workspace.

   Start HeidiSQL. Load (**File / Load SQL file**) and execute the script. As there is quite a lot of data, the operation might take some time (tens of seconds or even a few minutes). HeidiSQL may ask if the script should be downloaded into memory. That should not be necessary.

   Check that the creation is successful and verify that the contents of **Employee** and **Phone_call** tables are OK.

   Describe the resulting database structure (tables, fields, relationships). A text description is sufficient.

2. Now, your goal is to make the following three queries as fast as possible:

   Query 1: Find the first names and salaries of all employees with family name 'Virtanen'

   Query 2: Find the sum of prices of all the calls that have been made from telephone number 041-951114.

   Query 3: Find the ID and price of all the calls that have been made by any of the employees with family name 'Virtanen'.

   The idea is that these queries represent repeating queries from an application. The underlined parameter values varies, but the structure remains.

   Write the queries in SQL and check the response time of each query. What is the reason for slowness?

Show each SQL query as well as the observed execution time. Describe the reason for slowness.

3. Design an index to speed up query 1. Create the index and re-run Query 1. Check the new response time.

   Show the index creation statement and the new response time.

4. Do a similar indexing for Query 2. How does the response time change?

   Show the index creation statement and the new response time.

5. What kind of indices do you need for Query 3? How does the response time change in comparison to the original response time?

   Show the index creation statement(s) and the new response time.

6. (A more challenging task.) Test by simulation how the use of and index affects the performance time of queries. For this, use **`mysqlslap`** program.

   First, create a text file **`queries.sql`** that contains 5-10 different queries on family name in table Employee). Each SQL query (SELECT command) should be written on its own row in the file.

   The simulation is meaningful only if the queries are structurally similar: only the family name should differ.

   Observe the performance times of the queries with **`mysqlslap`** program.

   Below is an example on a simulation command. Please note that **`mysqlslap`** is an independent program that will be run in the command prompt (in Windows, select **MariaDB/Command Prompt**). It cannot be run from within MariaDB or via HeidiSQL.

```
[root@olliv-001 /]# mysqlslap --user=root --password --concurrency=5 --iterations=10
--create-schema=firma2 --query=queries.sql
```

   The command specifies the number of concurrent database connections (**`concurrency`**) and number of iterations for the queries (**`iterations`**) within each of the connections. Additionally it specifies that the database to be used is **firma2** (replace with the correct database name), and tells that the queries are to be found in the file **`queries.sql`** (Here it is assumed that the query file is located in the same directory where you give the command).

   Write down the average execution time for at least four different combinations of parameter values of **`concurrency`** and **`iterations`** both with index and without index.

Show the experimental design and the results in your answer.

Deliverables:

Submit a pdf document that has the answers to the questions above. You can add screenshots to the document.