

T-202-GAG1: Assignment 11

Readings

Ramakrishnan & Gehrke: Chapter 8, 9, 12.1-12.3

Preparation

For this assignment, it is very important to first listen to lecture V11A (this is an old recording from 2014, but it is nevertheless the correct one).

The Assignment

For this assignment, you can work groups of two or three. Do as much of the following as you can in the span of 90 minutes.

In particular, if your computer has known issues with the FILL scripts, please arrange to work with someone with a computer that does not have such issues.

This assignment is created with psql (command prompt) usage in mind.

If you use some other tool you need to try and avoid outputting results when filling the database, as otherwise your results will be skewed. Also, do not commit after each insert, as that will skew your results too. The FILL.sql script avoids both those pitfalls.

Part I: No Keys

Start by creating a database called SCALEDDB.

Then run the commands in 1_CREATE_NOKEY.sql, which create the database schema from P2 but without any PRIMARY KEY and FOREIGN KEY constraints.

Then populate the schema with the commands in 2_FILL.sql. You need to modify the path to where you store the 2_FILL.sql file. Note (= write down) how long it takes, so that you can compare the insertions with and without the overhead of index maintenance.

Run 3_QUERIES.sql. Note how long it takes to execute each of the two queries, along with how many buffers were read.

- What could explain the difference in execution time and buffers read between Query I and Query II?

Store the explain plan output for both queries and compare it to the explain plans generated in Part II.

You can now drop the tables using 4_DROP.sql. Make sure you have noted all the query processing times, buffers read and have made a copy of the output of the execution plan (explain output). If you are missing something, you can repeat the whole process from 1_CREATE_NOKEY to 4_DROP.

Part II: Keys

Now, run the commands in 1_CREATE_KEY.sql, which create the database schema from P2 with all the PRIMARY KEY and FOREIGN KEY constraints.

Then populate the schema again with the commands in 2_FILL.sql. Note how long it takes, so that you can compare the insertions with and without the overhead of index maintenance.

Rerun 3_QUERIES.sql and compare the execution time of Query I and Query II with that of the database without constraints.

- Do you see indices used in the query plans (explain)?
- Why is the difference, in both execution time and buffers read, between Part I and Part II considerably larger for Query II than Query I?

Look at all your numbers, and see if you can see both the benefits and drawbacks of indices.

If you can get results from both a hard disk and an SSD, even better, and it is fine to share results for this purpose. In all cases, you should note hardware used to obtain your results.

Part III: Query Tuning

Consider Query 7 from Project 2 (you can use the official solution, if you did not have the correct query). In this part, use SCALEDDB with relations created using 1_CREATE_KEY.sql.

Run Query 7 and note how long the execution takes (it may take a long time; if it runs more than a minute or five you should just cancel the execution, as it is already unacceptably long). While it is running, continue reading and thinking ahead!

Then try creating an index to speed up the execution.

- Which attribute(s) should be included in the key and in which order, for best results. You can try some combinations
- How long does the execution take using the index?
- Do you see the index used in the query plans?

Since most likely you cannot run Query 7 to conclusion in the SCALEDDB database, repeat the same experiment in the original P2 database and answer the same questions above!

Deliverables

Submit a PDF file containing the outcomes of your experiments (including hardware information) and your thoughts about the results of the experiments (in particular the answers to the questions in bullets).

In the (short) report, section headings for each part are useful, as are tables.

Again, if you are well prepared, you should not spend more than 90 minutes on the assignment.

Additional Exercises

Exercises: 8.3, 8.7, 8.9, 8.11.