

Query Optimization Exercise

Please carry out the following instructions in the order specified and enter the original query exactly as written.

- Get a copy of dict.txt from Moodle.
- Create a new database named as your userid, for example, mine is brendahmcfarland.
- Create a new table named 'dict' in the database you just created with a varchar(30) field named 'word' (No indexes should be created including no primary key index).

Load the data from dict.txt into the dict table. You can FTP dict.txt to the server and then use this command as the MariaDB prompt: **load data local infile 'dict.txt' into table dict;** Be sure to make your database the active database.

You will have built a table that has 76,596 words in it.

<pre>+-----+ word +-----+ (CABLIS) (ESA-IRS (EUCLID) (ISBD abaci aback abacus abacuses abaft abalone abandon abandoned abandonee abandoner abandoning +-----+</pre>	<pre>MariaDB [brendahmcfarland]> select count(*) from dict; +-----+ count(*) +-----+ 76596 +-----+ 1 row in set (0.05 sec)</pre>
--	--

Query Optimization Exercise

1. Execute the following SQL statement:

```
SELECT d1.word, d2.word, d3.word
FROM dict AS d1 JOIN dict AS d2
      ON d1.word=reverse(d2.word)
JOIN dict AS d3
      ON d1.word=right(d3.word,6)
WHERE length(d1.word)=6 AND length(d3.word)>6;
```

You should get output as follows. Please make sure that you understand the query and what it is trying to accomplish.

word	word	word
dialer	relaid	autodialer
warder	redraw	awarder
warder	redraw	forwarder
remark	Kramer	premark
repaid	diaper	prepaid
redder	redder	shredder
spoons	snoops	tablespoons
spoons	snoops	teaspoons
drawer	reward	top-drawer
reined	denier	unreined
repaid	diaper	unrepaid
11 rows in set		

How long does it take that query to run for you? I ran it several times and it took anywhere from 1 minute 7 seconds to 1 minute 10 seconds depending on the time of day and resources being used. Suffices to say I think the average time is over one minute.

You can optimize queries in several different ways. Since we do not have administrative privileges and therefore can't adjust a lot of things on the database itself, I would like use to optimize it using SQL only. Here are the "rules" for this exercise:

Do one or more of the following to improve the query execution time, keeping in mind tht the query must execute successfully in the end ... i.e. generate the same output rows):

Change a clause in the query

Query Optimization Exercise

Add or drop a function

Add or drop a table field (NEVER drop the original field "word")

Change a field property (NEVER change a field property for "word")

Add or drop an index

Modify an index

NEVER delete or add rows to or from the table

NEVER create any additional tables

Open up a notepad++ file (or editor of your choosing). Each time you are ready to make a change, please log the following items:

- what the change is
- any SQL used to make the change
- show the query you ran, the result set, and the time it ran.

Continue this process until you believe that you have the lowest time possible for your query.... Which would be the optimal results. Always making sure that the output exactly matches the output from the initial query with those same 11 rows.

Once you have your optimal time, please write an explanation of what you did and highlight the changes that are associated with the improved execution time and why you think that change caused the execution time to improve. If you make some changes that cause the time to increase, figure out why that is the case and write that up too.

There is no exact answer. There are many things that you can do to optimize this query your query in this fashion.

I have done this exercise before and everyone always wants to know what time I got. I will be happy to tell you later in the week when you have tried it out. I hope someone follows the “rules” and optimizes better than I did!!!

Verify that the output of each query is the same.