

УДК 004.655
ББК 32.973.26-018.2
Д66

Домбровская Г., Новиков Б., Бейликова А.

Д66 Оптимизация запросов в PostgreSQL / пер. с англ. Д. А. Беликова. – М.: ДМК Пресс, 2022. – 278 с.: ил.

ISBN 978-5-97060-963-7

Книга поможет вам писать запросы, которые выполняются быстро и вовремя доставляют результаты. Вы научитесь смотреть на процесс написания запроса с точки зрения механизма базы данных и начнете думать, как оптимизатор базы данных. Объясняется, как читать и понимать планы выполнения запросов, какие существуют методы воздействия на них с точки зрения оптимизации производительности, и показано, как эти методы используются вместе для создания эффективных приложений.

Издание предназначено разработчикам и администраторам баз данных, а также системным архитекторам, использующим PostgreSQL.

УДК 004.655
ББК 32.973.26-018.2

First published in English under the title PostgreSQL Query Optimization; The Ultimate Guide to Building Efficient Queries by Henrietta Dombrovskaya, Boris Novikov and Anna Bailliekova, edition: 1. This edition has been translated and published under licence from APress Media, LLC, part of Springer Nature.

APress Media, LLC, part of Springer Nature takes no responsibility and shall not be made liable for the accuracy of the translation. Russian language edition copyright © 2022 by ДМК Пресс. All rights reserved.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN 978-1-4842-6884-1 (англ.)

© Henrietta Dombrovskaya,
Boris Novikov,
Anna Bailliekova, 2021

ISBN 978-5-97060-963-7 (рус.)

© Перевод, оформление, издание,
ДМК Пресс, 2022

Содержание

От издательства	11
Об авторах	12
О техническом редакторе	13
Благодарности	14
Вступление	15
Глава 1. Зачем нужна оптимизация?	21
Что подразумевается под оптимизацией?.....	21
Императивный и декларативный подходы: почему это сложно	22
Цели оптимизации	25
Оптимизация процессов.....	26
Оптимизация OLTP и OLAP.....	27
Проектирование базы данных и производительность	27
Разработка приложений и производительность	28
Другие этапы жизненного цикла	29
Особенности PostgreSQL.....	29
Выводы	30
Глава 2. Теория: да, она нужна нам!	31
Обзор обработки запросов	31
Компиляция.....	31
Оптимизация и выполнение.....	32
Реляционные, логические и физические операции	32
Реляционные операции	33
Логические операции.....	36
Запросы как выражения: мыслить множествами.....	36
Операции и алгоритмы.....	37
Выводы	38
Глава 3. Еще больше теории: алгоритмы	39
Стоимостные модели алгоритмов.....	39
Алгоритмы доступа к данным	40
Представление данных	41
Полное (последовательное) сканирование.....	42

Доступ к таблицам на основе индексов	42
Сканирование только индекса.....	43
Сравнение алгоритмов доступа к данным	44
Индексные структуры	46
Что такое индекс?	46
В-деревья.....	48
Почему так часто используются В-деревья?.....	49
Битовые карты	50
Другие виды индексов	51
Сочетание отношений.....	51
Вложенные циклы	52
Алгоритмы на основе хеширования	54
Сортировка слиянием	55
Сравнение алгоритмов	56
Выводы	56
Глава 4. Планы выполнения.....	57
Собираем все вместе: как оптимизатор создает план выполнения	57
Чтение планов выполнения	58
Планы выполнения.....	61
Что происходит во время оптимизации?.....	62
Почему планов выполнения так много?	62
Как рассчитываются стоимости выполнения?.....	63
Почему оптимизатор может ошибаться?.....	65
Выводы	66
Глава 5. Короткие запросы и индексы.....	67
Какие запросы считаются короткими?.....	67
Выбор критериев фильтрации	69
Селективность индексов.....	69
Уникальные индексы и ограничения	70
Индексы и неравенства.....	74
Индексы и преобразования столбцов.....	74
Индексы и оператор like	78
Использование нескольких индексов	80
Составные индексы	81
Как работают составные индексы?	81
Меньшая селективность	83
Использование индексов для получения данных.....	83
Покрывающие индексы	84
Избыточные критерии отбора	85
Частичные индексы	88
Индексы и порядок соединений.....	90
Когда индексы не используются.....	93
Избегаем использования индекса.....	93
Почему PostgreSQL игнорирует мой индекс?.....	94

Не мешайте PostgreSQL делать свое дело.....	96
Как создать правильные индексы?	98
Создавать или не создавать	98
Какие индексы нужны?.....	100
Какие индексы не нужны?.....	101
Индексы и масштабируемость коротких запросов.....	101
Выводы	102

Глава 6. Длинные запросы и полное сканирование..... 103

Какие запросы считаются длинными?	103
Длинные запросы и полное сканирование.....	104
Длинные запросы и соединения хешированием	105
Длинные запросы и порядок соединений	106
Что такое полусоединение?.....	106
Полусоединения и порядок соединений.....	108
Подробнее о порядке соединений	109
Что такое антисоединение?	111
Полу- и антисоединения с использованием оператора JOIN	113
Когда необходимо указывать порядок соединения?.....	115
Группировка: сначала фильтруем, затем группируем	117
Группировка: сначала группируем, затем выбираем.....	123
Использование операций над множествами	124
Избегаем многократного сканирования	128
Выводы	133

Глава 7. Длинные запросы: дополнительные приемы..... 134

Структурирование запросов.....	134
Временные таблицы и общие табличные выражения.....	135
Временные таблицы.....	135
Общие табличные выражения (СТЕ).....	137
Представления: использовать или не использовать	140
Зачем использовать представления?.....	145
Материализованные представления	146
Создание и использование материализованных представлений.....	147
Обновление материализованных представлений.....	148
Создавать материализованное представление или нет?	148
Нужно ли оптимизировать материализованные представления?	150
Зависимости	151
Секционирование	151
Параллелизм	155
Выводы	156

Глава 8. Оптимизация модификации данных..... 157

Что такое DML?.....	157
Два способа оптимизации модификации данных.....	157
Как работает DML?	158

Низкоуровневый ввод-вывод	158
Влияние одновременного доступа	159
Модификация данных и индексы	161
Массовые обновления и частые обновления	162
Ссылочная целостность и триггеры	163
Выводы	164

Глава 9. Проектирование имеет значение..... 165

Проектирование имеет значение	165
Зачем использовать реляционную модель?	168
Типы баз данных.....	168
Модель «сущность–атрибут–значение».....	169
Модель «ключ–значение».....	169
Иерархическая модель	170
Лучшее из разных миров.....	171
Гибкость против эффективности и корректности	172
Нужна ли нормализация?	173
Правильное и неправильное использование суррогатных ключей.....	175
Выводы	180

Глава 10. Разработка приложений и производительность..... 181

Время отклика имеет значение	181
Всемирное ожидание.....	182
Показатели производительности	183
Потеря соответствия.....	183
Дорога, вымощенная благими намерениями.....	184
Шаблоны разработки приложений	184
Проблема списка покупок	186
Интерфейсы.....	188
Добро пожаловать в мир ORM	188
В поисках более подходящего решения	189
Выводы	191

Глава 11. Функции..... 193

Создание функций	193
Встроенные функции	193
Пользовательские функции	194
Знакомство с процедурным языком.....	194
Долларовые кавычки.....	195
Параметры и возвращаемое значение	196
Перегрузка функций	197
Выполнение функций.....	198
Как происходит выполнение функций	200
Функции и производительность.....	203
Как использование функций может ухудшить производительность	203
Могут ли функции улучшить производительность?	205

Функции и пользовательские типы	205
Пользовательские типы данных.....	205
Функции, возвращающие составные типы.....	206
Использование составных типов с вложенной структурой.....	209
Функции и зависимости типов	213
Управление данными с помощью функций	213
Функции и безопасность.....	215
Как насчет бизнес-логики?.....	216
Функции в системах OLAP	217
Параметризация	217
Отсутствие явной зависимости от таблиц и представлений	217
Возможность выполнять динамический SQL.....	217
Хранимые процедуры	218
Функции, не возвращающие результат.....	218
Функции и хранимые процедуры	218
Управление транзакциями.....	219
Обработка исключений.....	219
Выводы	220

Глава 12. Динамический SQL..... 221

Что такое динамический SQL.....	221
Почему в Postgres это работает лучше.....	221
Что с внедрением SQL-кода?.....	222
Как использовать динамический SQL в OLTP-системах.....	222
Как использовать динамический SQL в системах OLAP	227
Использование динамического SQL для гибкости.....	230
Использование динамического SQL в помощь оптимизатору	236
Обертки сторонних данных и динамический SQL	239
Выводы	239

Глава 13. Как избежать подводных камней

объектно-реляционного отображения 240

Почему разработчикам приложений нравится NORM.....	240
Сравнение ORM и NORM.....	241
Как работает NORM.....	242
Детали реализации	248
Сложный поиск.....	251
Обновления.....	254
Вставка.....	254
Обновление.....	254
Удаление.....	258
Почему бы не хранить JSON?	258
Прирост производительности.....	258
Совместная работа с разработчиками приложений.....	259
Выводы	259

Глава 14. Более сложная фильтрация и поиск	260
Полнотекстовый поиск.....	260
Многомерный и пространственный поиск.....	261
Обобщенные типы индексов PostgreSQL	262
Индексы GiST.....	262
Индексы для полнотекстового поиска	263
Индексирование очень больших таблиц.....	264
Индексирование JSON и JSONB.....	265
Выводы	268
Глава 15. Полный и окончательный алгоритм оптимизации	269
Основные шаги.....	269
Пошаговое руководство	270
Шаг 1. Короткий запрос или длинный?	270
Шаг 2. Короткий запрос.....	270
Шаг 3. Длинный запрос	271
Шаг 4. Инкрементальные обновления.....	272
Шаг 5. Неинкрементальный длинный запрос	272
Но подождите, это еще не все!	272
Выводы	273
Заключение	274
Предметный указатель	276