

УДК 004.65

doi:10.18720/SPBPU/2/id24-200

*Бандель Елизавета Станиславовна*¹,
студент магистратуры, бакалавр;
*Нестеров Сергей Александрович*²,
доцент, канд. техн. наук, доцент

МЕТОДИКА ПОВЫШЕНИЯ СКОРОСТИ ВЫПОЛНЕНИЯ ЗАПРОСОВ В СУБД POSTGRESQL

^{1,2} Россия, Санкт-Петербург,
Санкт-Петербургский политехнический университет Петра Великого;
¹ bandel.es@edu.spbstu.ru, ² nesterov@spbstu.ru

Аннотация. В работе описывается методика повышения скорости выполнения запросов на основе анализа эффективности различных средств, таких как создание индексов, переписывание запроса, использование материализованного представления, сокращение количества сканирований. На основе предложенной методики разработано программное средство, формирующее рекомендации по повышению скорости обработки конкретного запроса к заданной базе данных, работающей под управлением СУБД PostgreSQL.

Ключевые слова: PostgreSQL, реляционные базы данных, язык SQL, стоимость выполнения запроса.

*Elizaveta S. Bandel*¹,
BSc, Master Student;
*Sergey A. Nesterov*²,
Candidate of Technical Sciences (PhD), Associate Professor

THE METHODOLOGY FOR INCREASING THE SPEED OF QUERY EXECUTION IN THE POSTGRES SQL DBMS

^{1,2} Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia; ¹ bandel.es@edu.spbstu.ru, ² nesterov@spbstu.ru

Abstract. The work develops a methodology for increasing the speed of query execution based on an analysis of the effectiveness of various tools, such as creating indexes, rewriting a query, using a materialized view, and reducing the number of scans. Based on the proposed methodology, a software tool has been developed that generates recommendations for increasing the speed of processing a specific request to a given database running under the PostgreSQL DBMS.

Keywords: PostgreSQL, relational databases, SQL language, request execution cost.

Введение

Сейчас в базах данных хранятся огромные объемы данных. Системы управления базами данных (СУБД) играют ключевую роль в обработке, хранении и анализе этих данных. Быстрота выполнения запросов является ключевым фактором для эффективной работы системы. Это важно для компаний, работающих с большими объемами данных. Некорректно или неэффективно написанные запросы могут привести к значительному снижению производительности баз данных и к увеличению времени отклика приложений, а каждая задержка или простой в работе системы могут привести к серьезным проблемам. Поэтому скорость выполнения запросов является критически важным элементом для обеспечения эффективной работы приложений [1].

1. Постановка задачи

База данных (БД) — именованная совокупность данных, отражающая состояние объектов и их отношений в заданной предметной области [4]. Каждый запрос к БД имеет свою скорость выполнения, которая напрямую зависит от стоимости C выполнения запроса. Она определяется многими параметрами, например, структурой запроса. Выразим это с помощью формулы (1):

$$C = f(D, S, I, M, H), \quad (1)$$

где D — множество данных, хранящихся в системе, S — структура запроса, I — набор использованных при выполнении запроса индексов, M — используемые материализованные представления, H — технические характеристики системы.

Так как скорость выполнения запросов является критически важным элементом, в работе будет рассматриваться взаимодействие с некоторыми параметрами запроса для минимизации стоимости C выполнения запроса, то есть $C \rightarrow C_{min}$. На основе полученных результатов разработана методика и написано приложение для повышения скорости выполнения запросов в СУБД PostgreSQL.

2. Разработка методики повышения скорости выполнения запросов

Для экспериментов была выбрана база данных, содержащая информацию об авиаперевозках, в которой хранятся данные за один год.

Для анализа эффективности применяемых средств необходимо оценить планы выполнения запросов, сравнивая выбранные показатели. Время выполнения не является надежным показателем, так как зависит от разных параметров (например, от окружения, в котором СУБД находится в данный момент). Фактическое время выполнения (execution time) будет рассматриваться как дополнительный показатель. В качестве основного показателя выбрана стоимость (cost), при снижении которой повышается скорость выполнения.

На основании экспериментов разработана методика для повышения скорости выполнения запросов в СУБД PostgreSQL. Схема алгоритма представлена на рисунке 1.

В методике предлагается разделить запросы на короткие и длинные. *Короткий запрос* — это запрос, который требует небольшого количества строк для получения результата, независимо от того, какие таблицы были задействованы [2]. Запрос считается *длинным*, если селективность запроса высока по крайней мере для одной из больших таблиц; то есть результат, даже если он невелик, определяется почти всеми строками [3].

3. Анализ эффективности разработанной методики

Для того чтобы оценить эффективность разработанной методики, рассмотрим следующий запрос, который возвращает информацию о билете пассажира из таблицы tickets, при этом есть два условия: на столбцы passenger_name и passenger_id.

```
EXPLAIN ANALYZE SELECT upper(passenger_name), passenger_id, contact_data
FROM tickets
WHERE upper(passenger_name) like 'SER%'
AND passenger_id='0031 522293';
```

Это короткий запрос и, следуя методике, необходимо сделать следующие действия:

1) создать индексы для полей passenger_name, passenger_id, contact_data;

- 2) создать покрывающий индекс с использованием преобразований внутри него;
- 3) посчитать итоговую стоимость;
- 4) использовать двойное условие вместо оператора like;
- 5) проверить статистику планировщика.

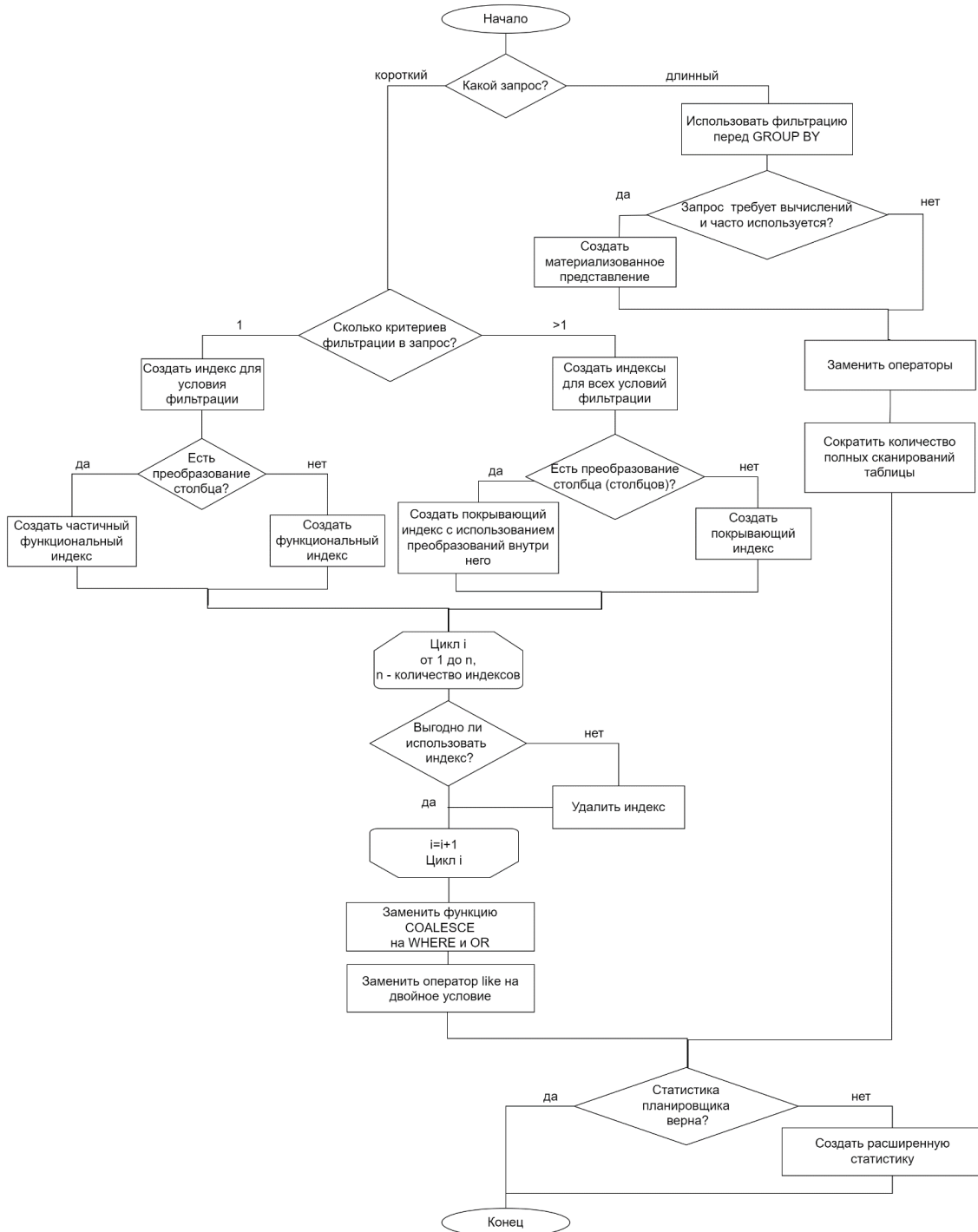


Рис. 1. Схема алгоритма повышения скорости выполнения запросов

Сравнение стоимости и времени выполнения запроса

Запрос	Стоимость (cost)	Фактическое время (execution time), мс
До применения методики	71 924.48	361.608
После применения методики	490.93	0.115

После применения полученной методики стоимость запроса уменьшилась примерно в 150 раз, а время – в 3 000 раз.

4. Описание реализованного программного обеспечения

На основе полученной методики было разработано приложение для поддержки принятия решений в области проектирования данных. Выбранный язык программирования — Java.

Приложение работает следующим образом. Сначала пользователь вводит запрос (рис. 2), затем видит план выполнения запроса и его характеристики (приложение подключается к базе данных для выполнения запроса и получения информации о нем). Далее происходит анкетирование пользователя о его запросе.

Рис. 2. Ввод запроса

В результате обработки запроса и ответов пользователя программа выдает пошаговый план для снижения стоимости введенного запроса (рис. 3).

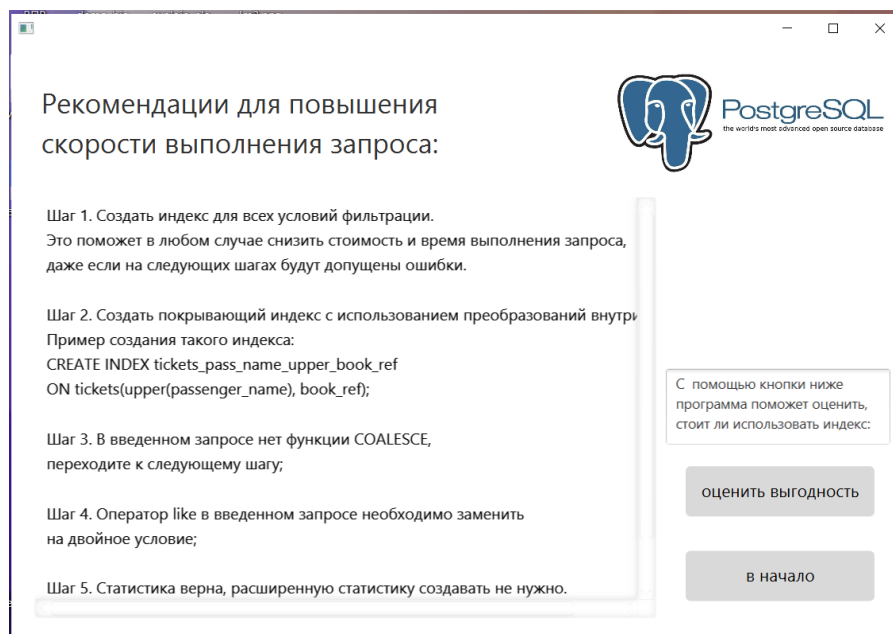


Рис. 3. Пошаговая рекомендация

Заключение

В ходе выполнения работы была создана методика повышения скорости выполнения запросов. В ней используются следующие средства: индексы различных типов, переписывание запроса, создание расширенной статистики, материализованные представления, оценка выгоды использования индексов. Также было разработано приложение, реализующее данную методику.

Созданное приложение может быть использовано начинающими разработчиками и администраторами баз данных. Дальнейшее развитие программного обеспечения заключается в возможности полной обработки запроса без анкетирования пользователя, переписывании запроса с целью его улучшения и написании запроса для создания необходимых индексов.

Список литературы

1. Джуба С., Волков А. Изучаем PostgreSQL 10 / Пер. с англ. А.А. Слинкина. – М.: ДМК Пресс, 2019. – 400 с.
2. Домбровская Г., Новиков Б., Бейликова А. Оптимизация запросов в PostgreSQL / Пер. с англ. Д. А. Беликова. – М.: ДМК Пресс, 2022. – 278 с.
3. Моргунов Е. П. PostgreSQL. Основы языка SQL: учеб. пособие. – СПб.: БХВ-Петербург, 2018. – 336 с.
4. Нестеров С. А. Базы данных. – СПб: Изд-во Политехн. ун-та, 2013. – 150 с.