

«УТВЕРЖДАЮ»

Заместитель директора по научной работе
Института проблем управления
им. В. А. Трапезникова
кандидат физико-математических наук

И.Н. Барабанов
21 ноября 2022 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного бюджетного учреждения науки «Институт проблем управления им. В.А. Трапезникова Российской академии наук»
(ИПУ РАН)

на диссертационную работу
Бучацкого Рубена Артуровича

«Метод динамической компиляции SQL-запросов для реляционных СУБД»,

представленную к защите на соискание учёной степени кандидата технических наук по специальности 2.3.5 – «математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Диссертационная работа Бучацкого Р.А. посвящена исследованию и разработке метода динамической компиляции запросов и её реализации в СУБД с целью ускорения выполнения SQL-запросов. Разработанный метод генерации кода под конкретный запрос с трансформацией алгоритмов операторов плана запроса из модели итераторов в модель явных циклов может быть применён к СУБД с Volcano моделью для реализации динамического компилятора запросов, что отражает практическую значимость работы.

Актуальность темы диссертационной работы

Наиболее распространённые в индустрии реляционные СУБД выполняют SQL-запросы посредством интерпретации его плана, основанного на модели итераторов. Данная модель проста для восприятия и удобна в реализации, но неоптимальна с точки зрения производительности на современных процессорах. Издержки

интерпретации особенно заметны в запросах, оперирующими большим количеством данных. Одним из средств сокращения накладных расходов интерпретации является динамическая компиляция запросов. В связи с этим диссертационная работы Р.А. Бучацкого посвящена *актуальной задаче* – разработке и реализации метода динамической компиляции запросов для ускорения их выполнения в СУБД.

Структура и основное содержание диссертационной работы

Диссертация имеет общий объем 168 страниц и состоит из введения, четырёх глав, заключения, списка литературы из 80 наименований, рисунков, таблиц и двух приложений.

Во введении обосновывается актуальность темы исследования, формулируется ее цель и задачи, излагается научная новизна, теоретическая и практическая значимость полученных результатов, формулируются основные положения, выносимые на защиту.

В первой главе содержится обзор предметной области по применению методов динамической компиляции в современных СУБД. Рассматриваются достоинства и недостатки имеющихся методов. Описываются модели выполнения запроса в СУБД и указывается применимость каждой из моделей. Делается вывод о перспективности применения метода динамической компиляции для ускорения выполнения запросов.

Во второй главе приводится описание предлагаемого метод динамической компиляции запросов с трансформацией операторов плана запроса из Volcano-модели в модель явных циклов. Описывается разработанный интерфейс оператора в модели явных циклов необходимого для трансформации плана запроса. Также описывается процесс декомпозиции алгоритмов основных операторов итераторной модели на функции разработанного интерфейса для выполнения запросов с использованием динамического компилятора. Описывается метод динамической компиляции выражений, используемых в запросе. Описываются разработанные эвристики стратегии выполнения и метод кэширования скомпилированного кода запроса, который избавляет от накладных расходов на компиляцию.

В третьей главе приводится описание программной реализации динамического компилятора запросов для открытой СУБД PostgreSQL на основе предложенных в диссертации методов. Описывается архитектура динамического компилятора, которая реализована на базе популярной компиляторной инфраструктуры LLVM.

В четвёртой главе приводятся результаты измерений производительности реализованного динамического компилятора запросов для СУБД PostgreSQL на

промышленных тестовых наборах TPC-H и TPC-DS. Тестирование проводилось на платформах x86-64 и ARM. Получено ускорение до 5 раз от применения динамического компилятора по сравнению с интерпретатором запросов PostgreSQL.

В заключении перечислены основные результаты диссертационной работы.

Основные результаты диссертационной работы

В диссертационной работе Бучацкого Р.А. получены следующие результаты:

1. Разработан метод динамической компиляции SQL-запросов с трансформацией на леву операторов плана запроса из модели Volcano в модель явных циклов.
2. Разработан метод динамической компиляции выражений в SQL-запросах с применением открытой вставки предварительно скомпилированных встроенных функций СУБД.
3. Разработаны эвристики стратегии выполнения запроса на основе оценок затрат на интерпретацию и динамическую компиляцию. Разработан метод кэширования кода, сгенерированного динамическим компилятором для одинаковых SQL-запросов.
4. Разработано программное расширение к СУБД PostgreSQL, реализующее метод динамической компиляции SQL-запросов.

Достоверность полученных результатов

Достоверность полученных результатов подтверждается экспериментальной и теоретической проверкой работоспособности предложенного подхода, а также апробацией на научных конференциях и публикациями в рецензируемых журналах, рекомендованных ВАК, две из которых индексируются в международных системах Scopus и WoS. Кроме того, получено свидетельство о государственной регистрации программы для ЭВМ.

Практическая значимость

Практическая значимость полученных результатов состоит в том, что предложенный соискателем метод динамической компиляции запросов может быть применён к СУБД с итераторной моделью для реализации динамического компилятора. Разработанный динамический компилятор запросов для СУБД PostgreSQL позволяет получить ускорение в несколько раз при выполнении аналитических запросов.

Замечания

В работе имеются следующие недостатки:

1. Недостаточно четко сформулирована проблема, решаемая изложенными в работе методами. В частности, не определены области возможного практического использования результатов.
2. Работа не предполагает решения оптимизационных задач. Тем не менее в тексте неоднократно используются термины «оптимизация» и «эффективность» без указания соответствующих критериев и целевых функций.
3. В работе не рассматриваются альтернативные способы генерации кода в промежуточном представлении помимо используемой инфраструктуры LLVM.
4. В разделе 3.5 для оценки времени компиляции кода используется линейная регрессия для нахождения зависимости количества инструкций в модуле от времени компиляции. В диссертации отсутствует разъяснение выбора такой зависимости, а не более сложной, например, учитывая тип инструкций, количество функций и т. д.
5. Разработанный автором динамический компилятор запросов реализован для СУБД PostgreSQL версии 9.6.3, поддержка которого закончилась в 2022 году. Можно ли оценить необходимые ресурсы для перевода сделанного компилятора на последнюю версию PostgreSQL?
6. Было бы полезно измерить вклад выполняемых оптимизаций в получаемое ускорение для выбора наиболее «полезных» оптимизаций.

Заключение

Отмеченные недостатки не снижают положительной оценки диссертационной работы. Диссертация является законченным научным исследованием, написанном на высоком уровне. Результаты диссертации представлены в статьях автора, а также докладывались на российских и международных научных конференциях. Автореферат диссертации правильно и полно отражает содержание работы и оформлен надлежащим образом.

Диссертационная работа Бучацкого Р.А. полностью соответствует требованиям ВАК, предъявляемым к диссертациям на соискание учёной степени кандидата технических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей». Бучацкий Рубен Артурович заслуживает присуждения учёной степени кандидата

технических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Диссертационная работа и отзыв рассмотрены и утверждены на расширенном семинаре лабораторий №№ 11, 17, 31, 46, 49 ИПУ РАН. На заседании семинара присутствовало 10 человек, включая двух докторов и шесть кандидатов наук.

Главный научный сотрудник ИПУ РАН,
д.т.н., профессор Калянов Георгий Николаевич

/Г.Н. Калянов/

21 ноября 2022 г.

Калянов Георгий Николаевич, главный научный сотрудник лаборатории №49 «Проектирования автоматизированных систем управления многоцелевыми объектами», доктор технических наук (специальность 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей), профессор

Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В. А. Трапезникова Российской академии наук; адрес: 117997 Москва, Профсоюзная 65; телефон: +7 495 334-89-10; факс: +7 495 334-93-40; email: dan@ipu.ru; сайт: <https://www.ipu.ru/>.