

УДК 004.657

ИССЛЕДОВАНИЕ АЛГОРИТМОВ ПОИСКА ОПТИМАЛЬНОГО ПЛАНА АНАЛИТИЧЕСКОГО ЗАПРОСА В СИСТЕМАХ УПРАВЛЕНИЯ БАЗ ДАННЫХ ДЛЯ ГЕТЕРОГЕННЫХ СИСТЕМ

Никольская Д.В. (Университет ИТМО, г. Санкт-Петербург)

Научный руководитель – к. т. н., доцент Косяков М.С.

(Университет ИТМО, г. Санкт-Петербург)

В последнее время растет спрос на повышение гетерогенности в вычислительных системах, который связан с необходимостью осуществления высокопроизводительных вычислений. Недавние исследования показывают преимущества в производительности параллельных реализаций под архитектуру GPU (графических процессоров) некоторых реляционных операторов СУБД над классическими реализациями под архитектуру ЦПУ. Целью работы является реализация преобразования плана выполнения аналитического запроса в СУБД к гетерогенному виду путем добавления специального оператора, распределяющего данные между ЦПУ и GPU, и исследование алгоритмов оптимизации такого плана запроса.

Введение.

Для решения широкого спектра задач исследователей данных, бизнес аналитиков и множества других профессий, связанных с данными, используются OLAP (online analytical processing) запросы. Аналитические запросы обрабатывают большие объемы данных, без их точечного изменения. Также работа с такими запросами не требует архивирования данных. Эти особенности спровоцировали появление СУБД для аналитики, задача которых эффективно выполнять OLAP запросы. Один из подходов к повышению эффективности - использование специализированных аппаратных ускорителей, таких как графический процессор.

Использование GPU в качестве основного вычислителя не обеспечивает оптимизации всех типов запросов. Графический процессор в отличие от ЦПУ предоставляет широкие возможности использования параллелизма за счет многоядерной архитектуры, однако проигрывает в выполнении сложных арифметических операций и исполнении программ с ветвлением. Поскольку эффективность устройств отличается при различных типах нагрузки, возникает необходимость выбора нужного устройства для выполнения запроса или его частей.

Под гетерогенностью подразумевается возможность совместного исполнения некоторой задачи на двух или более устройствах с различной архитектурой. В нашем случае такой задачей является аналитический запрос в СУБД.

Основной проблемой гетерогенного исполнения являются задержки, связанные с передачей данных между устройствами. Поэтому оптимизация требуется не на уровне отдельных операторов, а на уровне всего плана в совокупности. Существующие СУБД с поддержкой гетерогенности в основном используют жадную стратегию или перебор с ограничением переходов между устройствами.

Основная часть.

В NetExchange предлагается способ преобразования физического плана выполнения запроса СУБД к гетерогенному. Идея состоит в следующем: на каждом шаге плана перед стандартным оператором добавляется специальный оператор управления потоком данных (Router), задающий распределение данных между ЦПУ и графическим процессором для выполнения последующего оператора.

Чтобы минимизировать полное время исполнения запроса, необходима модель затрат, позволяющая определить «стоимость» выполнения гетерогенного плана и алгоритм, который, используя существующую модель затрат, способен подобрать параметры операторов типа

Router. При этом прирост производительности достигается тогда, когда время, затрачиваемое на оптимизацию, значительно меньше времени исполнения запроса.

Выбор распределения данных между устройствами для каждого шага плана это NP-трудная задача. Оценка стоимости плана с помощью модели затрат – это функция от параметров операторов типа Router, нам нужно найти минимум этой функции. Для решения таких задач хорошо подходят алгоритмы оптимизации, так как в нашем случае нет необходимости в поиске самого оптимального решения, достаточно приблизиться к нему за небольшое количество итераций. Многие алгоритмы оптимизации основаны на итеративности: с каждой новой итерацией они улучшают решение, найденное на предыдущем шаге. Эта особенность в нашем случае позволит найти баланс между временем, затраченным на оптимизацию и временем выполнения запроса.

Чтобы выбрать алгоритм для оптимизации гетерогенных планов, предлагается реализовать проект оптимизатора, который преобразует физический план выполнения запроса СУБД в гетерогенный с использованием различных алгоритмов расстановки параметров и позволяет измерить время работы оптимизации. Модель затрат в оптимизаторе основывается на замерах производительности операторов СУБД OmniSciDB. Алгоритмы, выбранные для оптимизатора на данный момент: генетический алгоритм, алгоритм пчелиной колонии и алгоритм светлячка. Также для сравнения в нем реализованы полный перебор и жадный алгоритм выбора параметров.

Выводы.

Оптимизатор позволяет оценить, насколько хорошо работают оптимизационные алгоритмы в сравнении с переборными и жадными. А также позволит добавить новые алгоритмы расстановки параметров, например, использующие машинное обучение.

На данный момент наилучший результат сходимости на запросах бенчмарка TPC-H показывает алгоритм пчелиной колонии.

Исследование в частности направлено на улучшение реализации гетерогенности в in-memory базе данных OmniSciDB, которая в свою очередь используется в платформе Modin для ускорения pandas – популярной библиотеки для обработки и анализа данных. Поэтому лучший вариант оптимизации имеет потенциал интеграции в OmniSciDB, в которой на данный момент есть поддержка выполнения отдельных операторов на GPU, однако процесс выбора того, как будет исполнен запрос, не автоматизирован и отсутствуют кросс-операторные оптимизации.

Никольская Д.В. (автор)

Подпись

Косяков М.С. (научный руководитель)

Подпись