

МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ И КОМПЬЮТЕРНЫХ СЕТЕЙ

УДК 004.65

ЭНЕРГОЭФФЕКТИВНОСТЬ SQL-ЗАПРОСОВ В POSTGRESQL: ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ МЕТОДИК ИЗМЕРЕНИЙ И ОПТИМИЗАЦИИ

Д. Рахмани, старший преподаватель Московского технического университета связи и информатики, Москва, Россия;

orcid.org/0009-0002-8371-3358, e-mail: jahed@mtuci.ru

Д. И. Григоренко, студентка Московского технического университета связи и информатики, Москва, Россия;

orcid.org/0009-0007-2014-9215, e-mail: grigorenkodarya2004@gmail.com

В условиях роста энергопотребления центров обработки данных возникает необходимость оптимизации систем управления базами данных не только по производительности, но и по энергоэффективности. Целью работы являются разработка воспроизводимой методики прямого измерения энергопотребления SQL-запросов в PostgreSQL с использованием аппаратных счётчиков Intel RAPL и экспериментальная оценка влияния индексов, состояния кэша и параллелизма на энергозатраты и метрику Energy-Delay Product (EDP). Эксперименты проводились на таблице из 10 млн записей, каждый тест повторялся 25 раз для обеспечения статистической достоверности. Установлено, что при селективности 1 % индексное сканирование снижает энергопотребление на 71,9 % и улучшает EDP на 94,6 % по сравнению с последовательным сканированием. Прогрев кэша уменьшает долю энергии, затрачиваемой на I/O, с 54,9 % до 5,3 %, а общее энергопотребление с 192,1 Дж до 68,9 Дж. Установлено, что оптимальный уровень параллелизма по EDP достигается при четырёх рабочих процессах, указанное значение применимо для конфигурации Intel Core i7-11700 (8 ядер). Полученные результаты позволяют формулировать практические рекомендации по энергоэффективной настройке PostgreSQL в высоконагруженных средах и служат основой для разработки «зелёных» оптимизаторов запросов, учитывающих как производительность, так и энергетические затраты.

Ключевые слова: энергоэффективность, PostgreSQL, RAPL, индексы, кэширование, параллелизм, EDP, методика измерений.

DOI: 10.21667/1995-4565-2026-95-66-72

Введение

Быстрый рост больших данных и повсеместное использование машинного обучения привели к значительному увеличению энергопотребления [1], что выдвигает на первый план задачу оптимизации не только по производительности, но и по энергоэффективности. СУБД являются значительными потребителями ресурсов в ИТ-инфраструктуре. Традиционные методы оптимизации запросов ориентированы на время выполнения, но не учитывают энергопотребления. Существующие методики измерений энергопотребления СУБД часто характеризуются недостаточной точностью и не обеспечивают необходимой воспроизводимости результатов. Актуальность разработки специализированных инструментальных решений подтверждается появлением методик с использованием интерфейса RAPL для сравнительного анализа энергоэффективности различных СУБД [3]. Необходимо адаптировать и расширить

подход для углубленного анализа факторов оптимизации именно в среде PostgreSQL, что позволяет получить воспроизводимые и статистически значимые результаты по влиянию конкретных механизмов СУБД на энергопотребление.

Постановка задачи

Целью исследования являются разработка и верифицирование воспроизводимой методики измерения энергопотребления и энергоэффективности SQL-запросов в PostgreSQL по таким показателям, как энергия E (Дж), время выполнения T (с) и комплексный показатель EDP.

Исследовательские гипотезы

H1: Использование индексов снижает E и EDP при низкой селективности запросов ($\leq 1\%$).

H2: Прогрев кэша значительно снижает общее энергопотребление за счёт сокращения активности ввода-вывода.

H3: Существует оптимальный уровень параллелизма, минимизирующий EDP.

Исследовательские вопросы

1. Как индексирование влияет на энергетические характеристики при различных уровнях селективности?

2. Как прогрев кэша перераспределяет энергопотребление между подсистемами?

3. Как параллелизм влияет на компромисс «энергия – время»?

Теоретическая часть

Энергоэффективность СУБД определяется взаимодействием программных и аппаратных компонентов. Основными потребителями энергии являются центральный процессор (CPU), оперативная память (DRAM) и подсистема ввода-вывода (I/O). В то время как CPU и DRAM могут быть измерены напрямую через RAPL, I/O-энергия оценивается косвенно как разность между общей энергией и суммой PKG + DRAM с оговоркой о погрешности (примечание ниже).

Комплексная метрика Energy-Delay Product (EDP) определяется как произведение энергии выполнения запроса (E) (Дж) и времени выполнения (T) (с):

$$EDP = E_{query} \times T. \quad (1)$$

Это комплексная метрика, отражающая компромисс между энергопотреблением и задержкой. Доли энергии по подсистемам рассчитывались следующим образом: CPU и DRAM – напрямую по доменам RAPL; I/O – как остаток:

$$E_{I/O} = E_{total} - E_{CPU} - E_{DRAM}. \quad (2)$$

Инструментальные счётчики фиксируют домены PKG и DRAM; показатель «I/O» является оценкой и приводится с указанием допущений и погрешности $\pm 8\%$.

Исследование проводилось. Эксперименты проведены на платформе:

– CPU: Intel Core i7-11700 (8 ядер, 2,5 ГГц, governor = performance);

– RAM: 32 ГБ DDR4;

– Storage: NVMe SSD;

– OS: Ubuntu 22.04 LTS, ядро 5.15;

– СУБД: PostgreSQL 15.0 с параметрами по умолчанию (shared_buffers = 8 ГБ, effective_cache_size = 24 ГБ). Тестовая таблица «orders» содержит 10 млн записей.

Листинг 1. Структура тестовой таблицы

```
CREATE TABLE orders (
  order_id BIGSERIAL PRIMARY KEY,
  customer_id INTEGER,
  order_date DATE,
  amount DECIMAL(10,2),
  status VARCHAR(20)
);
```

Каждый эксперимент повторялся 25 раз. Состояние кэша контролировалось: «холодный» сброс через /proc/sys/vm/drop_caches, «тёплый» – после 3 прогонов без замера. Анализ проводился с использованием U-критерия Манна – Уитни и Cohen's d. Основная часть исследования включала три блока: сравнение последовательного и индексного сканирования, анализ влияния кэша, оценка параллелизма. Методика измерения энергии реализована в Python:

Листинг 2. Функция измерения энергии запроса

```
def measure_query_energy(query):
    start_energy = read_rapl_energy()
    start_time = time.time()
    execute_sql(query)
    end_time = time.time()
    end_energy = read_rapl_energy()
    energy = end_energy - start_energy
    time_taken = end_time - start_time
    edp = energy * time_taken
    return energy, time_taken, edp
```

На основе измерений вычисляются метрики по формулам (3, 4):

$$E_{query} = E_{end} - E_{start}, \quad (3)$$

$$E_{per_row} = \frac{E_{query}}{N_{rows}}, \quad (4)$$

где: E_{query} – энергия, потреблённая непосредственно на выполнение запроса (Дж), измеряется как разность показаний RAPL до и после выполнения (переменная `energy` в Листинге 2); N_{rows} – количество строк в результате запроса; E_{per_row} – энергия на одну возвращённую строку (Дж/строка); T – время выполнения запроса (с), соответствует переменной `time_taken`.

Экспериментальные исследования

Пример расчёта. Рассмотрим выполнение запроса с селективностью 1 % с использованием индексного сканирования. Пусть до выполнения запроса показание RAPL составляло $E_{start} = 1250,3$ Дж, а после – $E_{end} = 1302,4$ Дж. Тогда энергия запроса рассчитывается по формуле (3):

$$E_{query} = 1302,4 - 1250,3 = 52,1 \text{ Дж}.$$

Время выполнения: $T = 0,52$ с. (измерено таймером). Следовательно, согласно формуле (1):

$$E_{end} = 52,1 \times 0,52 = 27,1 \text{ Дж} \cdot \text{с},$$

что согласуется с данными в таблице 1 (небольшое расхождение объясняется усреднением по 25 запускам). Расчёт демонстрирует прямую связь между аппаратными измерениями, программной реализацией и итоговыми метриками. Запрос вернул 100 000 строк, поэтому энергия на строку вычисляется по формуле (4):

$$E_{per_row} = \frac{52,1}{100\,000} = 0,521 \text{ мДж/строка}.$$

Такой подход позволяет количественно сравнить разные стратегии выполнения запросов не только по скорости, но и по энергоэффективности.

Применение индексных структур данных формирует рациональный компромисс между скоростью доступа к информации и энергетическими издержками, возникающими при поддержании их внутренней организации.

В процессе выборки данных использование индексов обеспечивает значительное сокращение объема просматриваемых записей, что приводит к заметному уменьшению нагрузки на подсистемы ввода-вывода и вычислительные ресурсы центрального процессора [2].

При низкой селективности выборки (1 %) переход от последовательного сканирования (Seq Scan) к индексному (Index Scan) позволяет сократить общее энергопотребление с 185,6

до 52,1 Дж, что эквивалентно снижению на 71,9 % ($p < 0,001$). Аналогичная тенденция наблюдается для интегрального показателя энергозатратки (EDP), который уменьшается с 434,3 до 27,1 Дж·с, демонстрируя улучшение на 94,6 %.

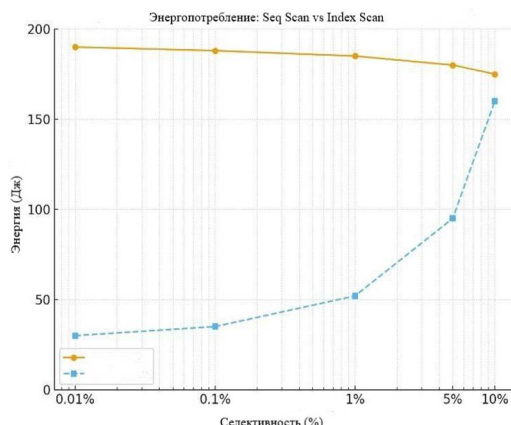


Рисунок 1 – Зависимость энергопотребления от селективности запроса

Figure 1 – Dependence of energy consumption on query selectivity

Итоговые значения представлены в таблице 1 и на рисунке 1.

Таблица 1 – Сравнение Seq Scan и Index Scan (селективность 1 %)

Table 1 – Comparison of Seq Scan and Index Scan (1 % selectivity)

| Метрика | SEQ SCAN | INDEX SCAN | Δ% |
|--------------|--------------|------------|--------|
| Энергия (Дж) | 185,6 ± 12,3 | 52,1 ± 3,8 | -71,9% |
| EDP (Дж·с) | 434,3 ± 35,1 | 27,1 ± 2,8 | -94,6% |

Интерпретация представленных данных подтверждает статистически достоверное преимущество индексного сканирования по энергетической эффективности при низком уровне селективности запросов. Энергопотребление снижается более чем на две трети, а показатель EDP демонстрирует почти двадцатикратное улучшение. Таким образом, гипотеза H1 получает эмпирическое подтверждение: использование индексных структур данных позволяет существенно повысить энергоэффективность операций выборки за счет сокращения нагрузки на вычислительные и I/O-компоненты системы.

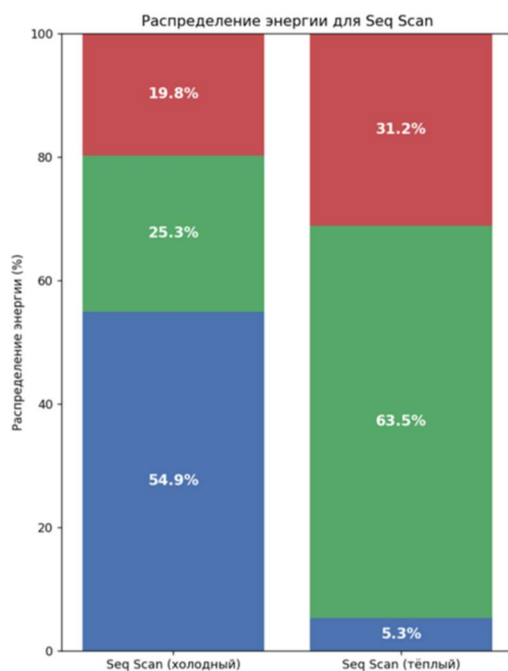


Рисунок 2 – Распределение энергии по компонентам системы

Figure 2 – Energy distribution across system components

Анализ распределения энергетических затрат (рисунок 2) показывает, что процесс прогрева кэша оказывает выраженное влияние на общую структуру энергопотребления. В частности, доля операций ввода-вывода в совокупных энергозатратах снижается с 54,9 % до 5,3 %, а общая потребляемая энергия уменьшается с 192,1 до 68,9 Дж. Это подтверждает, что эффективность индексных операций во многом обусловлена оптимизацией кэш-поведения и сокращением обращений к внешней памяти.

Таблица 2 – Распределение энергии по компонентам
Table 2 – Energy distribution across system components

| Состояние | CPU (%) | DRAM (%) | I/O (%) | Общая энергия (Дж) |
|---------------------|---------|----------|---------|--------------------|
| Seq Scan (холодный) | 25,3 | 19,8 | 54,9 | 192,1 |
| Seq Scan (тёплый) | 63,5 | 31,2 | 5,3 | 68,9 |

H2 подтверждена. Прогрев кэша снижает общую энергию на 64,1 % и радикально перераспределяет нагрузку в пользу CPU/DRAM.

Для комплексной оценки энергоэффективности систем обработки данных широко применяется метрика EDP (1), которая успешно интегрирует показатели энергопотребления и производительности в единый критерий оптимизации. Использование EDP позволяет количественно оценить компромисс между скоростью выполнения операций и энергетическими затратами, что соответствует современным подходам к разработке энергоэффективных решений в области систем управления базами данных. Данная метрика особенно ценна при сравнительном анализе различных стратегий выполнения запросов, поскольку обеспечивает:

- единую количественную основу для сравнения энергоэффективности;
- учет как временных, так и энергетических характеристик системы;
- возможность выявления оптимальных параметров конфигурации СУБД.

Такой подход позволяет перейти от традиционной оптимизации исключительно по времени выполнения к сбалансированной оценке, учитывающей как производительность, так и энергозатраты.

Таблица 3 – Влияние параллельных workers
Table 3 – Impact of parallel workers

| Workers | Время (с) | Энергия (Дж) | EDP (Дж·с) | Δ EDP |
|---------|-----------|--------------|------------|--------------|
| 0 | 3,21 | 195,6 | 627,9 | – |
| 2 | 1,85 | 210,3 | 389,1 | –38,0% |
| 4 | 1,32 | 245,8 | 324,5 | –48,3% |
| 8 | 1,15 | 298,3 | 343,0 | –45,4% |

Параллелизм улучшает EDP до определённого предела (таблица 3).

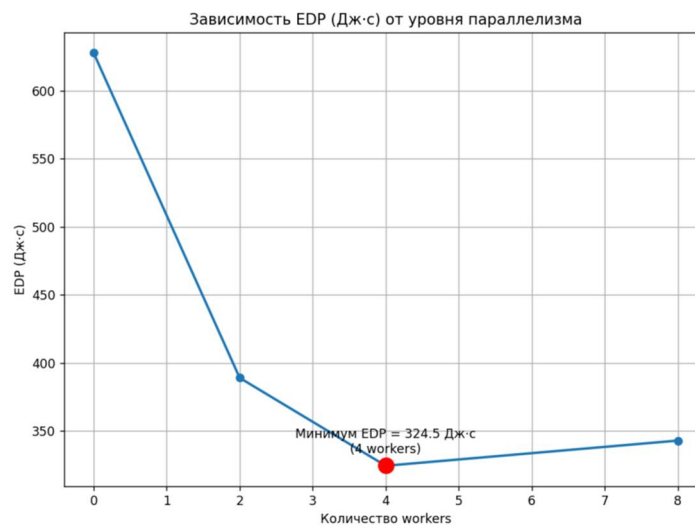


Рисунок 3 – Зависимость метрики EDP от уровня параллелизма
Figure 3 – Dependence of EDP Metric on the Level of Parallelism

НЗ подтверждена. Оптимальный уровень параллелизма по метрике EDP достигается при 4 рабочих процессах (рисунок 3); данное значение справедливо для конфигурации Intel Core i7-11700 (8 ядер) и требует калибровки для других аппаратных платформ.

Заключение

Проведённое исследование демонстрирует, что оптимизация SQL-запросов исключительно по времени выполнения может приводить к неоправданно высокому энергопотреблению. Предложенная методика прямого измерения энергии через Intel RAPL обеспечивает воспроизводимость и позволяет количественно оценивать компромисс между производительностью и энергоэффективностью с использованием метрики EDP. Индексирование энергоэффективно при селективности $\leq 1\%$: снижает энергопотребление на 71,9 % и EDP на 94,6 % ($p < 0,001$, $d = 3,2$). Прогрев кэша кардинально перераспределяет энергозатраты между подсистемами, снижая долю I/O с 54,9 % до 5,3 %, что имеет принципиальное значение для проектирования «зелёных» СУБД. Оптимальный уровень параллелизма по метрике EDP достигается при 4 рабочих процессах (улучшение на 48,3 % по сравнению с последовательным выполнением); данное значение справедливо для конфигурации Intel Core i7-11700 (8 ядер) и требует калибровки для других аппаратных платформ.

Полученные результаты применимы при настройке высоконагруженных PostgreSQL-кластеров в центрах обработки данных, где даже небольшое снижение энергопотребления на запрос масштабируется на миллионы операций. В перспективе методика может лечь в основу адаптивного оптимизатора, выбирающего план выполнения запроса не только по скорости, но и по энергетической стоимости.

Библиографический список

1. Barua H.B., Mondal K.C., Khatua S. Green Computing for Big Data and Machine Learning // Proceedings of the 5th Joint International Conference on Data Science & Management of Data (CODS-COMAD 2022). ACM, 2022. Pp. 348-351.
2. Kumar M. Energy Efficiency in Database Systems: A Survey // Journal of Systems and Software. 2018. Vol. 145. Pp. 1-22.
3. Database architectures: Current trends and their relationships to environmental data management / J.Pokorný. Environmental Modelling & Software, 2006. Vol. 21. № 12. Pp. 1579-1591.

UDC 004.65

ENERGY EFFICIENCY OF SQL QUERIES IN POSTGRESQL: EXPERIMENTAL STUDY OF MEASUREMENT AND OPTIMIZATION METHODS

D. Rahmani, lecturer, Moscow Technical University of Communications and Informatics, Moscow, Russia; orcid.org/0009-0002-8371-3358, e-mail: j.rahmani@mtuci.ru

D. I. Grigorenko, student, Moscow Technical University of Communications and Informatics, Moscow, Russia; orcid.org/0009-0007-2014-9215, e-mail: grigorenkodarya2004@gmail.com

With the growing energy consumption of data centers, there is an increasing need to optimize database management systems not only for performance but also for energy efficiency. This study aims to develop a reproducible methodology for directly measuring the energy consumption of SQL queries in PostgreSQL using Intel RAPL hardware counters and to experimentally evaluate the impact of indexing, cache state, and parallelism on energy consumption and Energy-Delay Product (EDP) metric. Experiments were conducted on a table containing 10 million records, each test repeated 25 times to ensure statistical reliability. The results show that, at a selectivity of 1 %, index scans reduce energy consumption by 71,9 % and improve EDP by 94,6 % compared to sequential scans. Cache warming reduces the share of energy spent on I/O from 54,9 % to 5,3 %, decreasing total energy consumption from 192,1 J to 68,9 J. The results indicate that the

optimal level of parallelism based on EDP metric is achieved with four worker processes; this finding is valid for Intel Core i7-11700 (8-core) configuration. These findings enable practical recommendations for energy-efficient PostgreSQL tuning in high-load environments and lay the groundwork for developing «green» query optimizers that consider both performance and energy costs.

Keywords: energy efficiency, PostgreSQL, RAPL, indexes, caching, parallelism, EDP, measurement methodology.

DOI: 10.21667/1995-4565-2026-95-66-72

References

1. **Barua H.B., Mondal K.C., Khatua S.** Green Computing for Big Data and Machine Learning. In: Proc 5th Joint Int. Conf. on Data Science & Management of Data (CODS-COMAD 2022). ACM, 2022, pp. 348-351.
2. **Kumar M.** Energy Efficiency in Database Systems: A Survey // *Journal of Systems and Software*. 2018, vol. 145, pp. 1-22.
3. Database architectures: Current trends and their relationships to environmental data management / **J. Pokorný.** *Environmental Modelling & Software*, 2006, vol. 21, no 12, pp. 1579-1591.