

ПРОИЗВОДИТЕЛЬНОСТЬ ЗАПРОСОВ К ГЕТЕРОГЕННЫМ ИСТОЧНИКАМ В СИТУАЦИОННО-ОРИЕНТИРОВАННЫХ БАЗАХ ДАННЫХ

А. С. ГУСАРЕНКО

Аннотация. В области разработки баз данных продолжается активное движение по направлению создания нереляционных NoSQL-баз данных. Вопросы эффективности функционирования реляционных баз данных рассматривались в современных опубликованных работах в части оптимизации запросов, измерения времени выполнения SQL-выражений, здесь уже реализованы и известны программные средства для тестирования, определены виды тестов для измерения производительности СУБД. Такие программные средства и тесты с небольшими трудозатратами переносятся от одной реляционной СУБД к другой для измерения и оценки производительности запросов. В то же время для NoSQL-баз данных общими могут быть только объемы данных, для которых требуется измерить время обработки ядром нереляционной СУБД. В связи с этим трудно встретить универсальные и легко переносимые средства для тестирования СУБД направления NoSQL. На аналогичном этапе разработки находятся ситуационно-ориентированные базы данных — подвид NoSQL-баз данных, создаваемый в рамках проекта СОБД. Вместе с тем есть возможность измерить производительность за счет использования специализированных средств профилирования на платформе PHP путем встраивания их в программный код СОБД. Данная работа продолжает серию исследований по СОБД и рассматривает вопросы измерения и оценки производительности на примерах работы с гетерогенными источниками данных в микросервисе СОБД, документо-ориентированным хранилищем, офисными документами, архивами, реляционной СУБД, архивами. Специально для этого были разработаны или приспособлены варианты примеров применения СОБД. Полученные результаты измерений сопровождаются таблицами и графическими протоколами графов вызова программных функций СОБД для методов загрузки объемов данных целиком (Bulk Load) и порциями (Portion Load). Полученные результаты позволяют сделать вывод о сопоставимости результатов в части обработки данных с другими NoSQL-базами и двигаться дальше в направлении повышения эффективности работы СОБД.

Ключевые слова: ситуационно-ориентированная база данных; встроенная динамическая модель; гетерогенные источники документов; виртуальные мультидокументы; динамические объекты обработки данных; RESTful-сервисы; тесты производительности.

ВВЕДЕНИЕ

В области разработки систем баз данных на этапе тестирования и эксплуатации программного обеспечения возникает актуальная задача проведения измерений производительности обработки данных, оптимизации запросов к ядру БД с целью повышения эффективности работы СУБД. По данной задаче есть опубликованные работы, связанные с улучшением работы запросов [1, 2] и предложениями по мультимодельным языкам запросов [3, 4]. Такие работы ведутся в связи с совместным использованием SQL и NoSQL, направлений в области обработки больших данных с использованием машинного обучения, где используются структурированные, неструктурированные и слабо структурированные данные из гетерогенных источников [5, 6].

Гетерогенные источники представляют собой разнородные хранилища в виде микросервисов [7, 8], реляционных и нереляционных баз данных, документов и архивов. Микросервисы и веб-приложения могут быть источниками всех перечисленных гетерогенных данных [9], предоставляемых удаленно по запросу пользователя.

Следующим важным моментом является увеличение количества источников и веб-приложений различного назначения, в том числе и образовательного характера [10]. С одной стороны, это вызывает проблемы с проведением измерений и необходимостью в разработке средств так называемого «бенчмаркинга» (от англ. benchmarks), то есть программ, выдающих контрольные показатели для сравнения результатов измерений, а с другой – ориентация таких средств на направления SQL, NoSQL и веб-сервисы [11]. Имеющиеся в настоящее время средства не являются универсальными и ориентированы в основном на отдельные классы СУБД, разделяемые по их видам. Для веб-сервисов используются специализированные средства профилирования, универсальные для веб-приложений, запрограммированных на каком-либо языке программирования.

В ситуационно-ориентированных базах данных (СОБД) [12] возникает аналогичная задача проведения измерений обработки данных с целью повышения эффективности работы баз данных. Известные SQL средства для тестирования и сами тесты с концептуального уровня также не подходят, если не брать в расчет использование нереляционных гетерогенных источников, таких как XML [13], JSON, офисных документов [14–16]. Если брать средства измерения, учитывающие только NoSQL-гетерогенные источники, то возникает проблема с выпаданием данных из реляционных СУБД. Таким образом, нет универсальных средств, подходящих для использования с СОБД [17–19], но с тем, чтобы приблизиться к ним, существуют средства для профилирования веб-приложений, и на программном уровне можно использовать такие средства, если рассматривать СОБД [20–22] как гетерогенное хранилище в виде микросервиса, RESTful-сервиса или веб-приложения.

ИЗМЕРЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ СОБД В РАБОТЕ С ГЕТЕРОГЕННЫМИ ИСТОЧНИКАМИ

Выбор программного обеспечения для фиксации результатов измерения производительности СОБД. Для тестирования веб-приложений и систем, функционирующих на базе веб-сервера, разработаны программные комплексы, решающие задачи профилирования, поиска узких мест в разработанном ПО с целью устранения недостатков в производительности, сокращения времени работы приложения и снижения потребления памяти.

Известны программные средства для системы программирования PHP, такие как:

- Pinba;
- XHprof;
- PHP-SPX;
- Yahoo! Cloud Service Benchmark.

В данной работе использовалось средство XHprof для получения результатов измерений. XHprof работает на веб-сервере Apache и известно как ранее уже использовавшееся для целей тестирования и улучшения производительности СОБД. Программа XHprof была разработана для повышения эффективности работы PHP-скриптов социальных сетей, она является знакомым и легко устанавливаемым средством, доступным на бесплатной основе любому веб-программисту.

Тест извлечения данных из СОБД. В виртуальной машине под управлением операционной системы Linux была развернута СОБД, а в ее код внедрены установочные строки для записи результатов измерений, полученных с XHprof. В качестве первого примера для тестирования производительности был выбран микросервис с данными о прохождении студентами курсового проектирования. Такой микросервис выдает данные, извлекаемые из СОБД, на экран или для сохранения в файл на внешнем диске. Запросы в микросервис посылаются строкой по протоколу HTTP с указанием в параметрах количества извлекаемых узлов. Данные выдаются в формате XML. На рис. 1 показан скриншот таблицы с результатами измерений извлечения данных из микросервиса СОБД с данными по студентам, а на рис. 2 показан граф вызова функций СОБД, привязанный к таблице результатов измерений для 10 000 узлов.

Overall Summary						
Total Incl. Wall Time (microsec): 13,071,717 microseconds						
Number of Function Calls: 443,103						
[View Full Callgraph]						
Displaying top 100 functions: Sorted by Incl. Wall Time (microsec) [display all]						
Function Name	Calls	Calls%	Incl. Wall Time (microsec)	IWall%	Excl. Wall Time (microsec)	EWall%
main()	1	0.0%	13,071,717	100.0%	42	0.0%
??_op	1	0.0%	13,071,344	100.0%	226	0.0%
Pass1	1	0.0%	13,070,712	100.0%	49	0.0%
SubModelProc1	1	0.0%	13,070,531	100.0%	36	0.0%
Pass1@1	1	0.0%	13,070,489	100.0%	49	0.0%
JumpProc1	2	0.0%	13,070,265	100.0%	28	0.0%
JumpExec	1	0.0%	13,070,111	100.0%	44	0.0%
Pass1@2	1	0.0%	13,070,051	100.0%	109	0.0%
ForProc1	1	0.0%	12,094,511	92.5%	37,684	0.3%
Pass1@3	2,557	0.6%	12,050,356	92.2%	175,213	1.3%
DOMProc	2,557	0.6%	8,330,027	63.7%	145,052	1.1%
DOMInterProc	2,557	0.6%	7,251,342	55.5%	72,343	0.6%
DOMRcvProc	2,557	0.6%	7,178,999	54.9%	705,674	5.4%
XSLTProcessor::transformToXml	2,557	0.6%	4,835,244	37.0%	4,835,244	37.0%
GlbProc1	5,118	1.2%	3,496,750	26.8%	3,150,353	24.1%
DOMDocument::load	5,115	1.2%	1,443,422	11.0%	1,443,422	11.0%
XSLTProcessor::importStylesheet	2,557	0.6%	921,208	7.0%	921,208	7.0%

Рис. 1 Таблица с результатами измерений извлечения данных из микросервиса СОБД с данными по студентам.

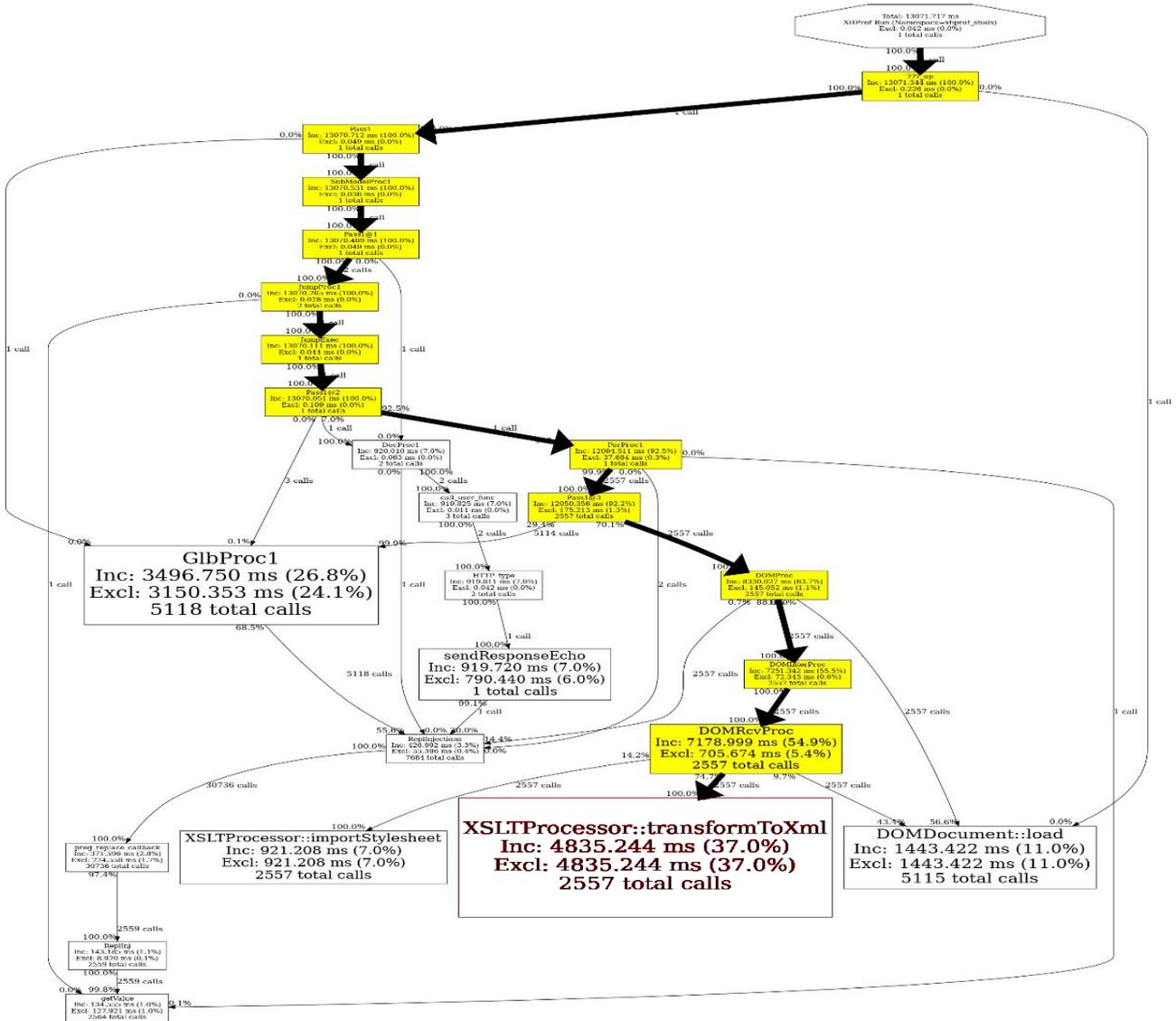


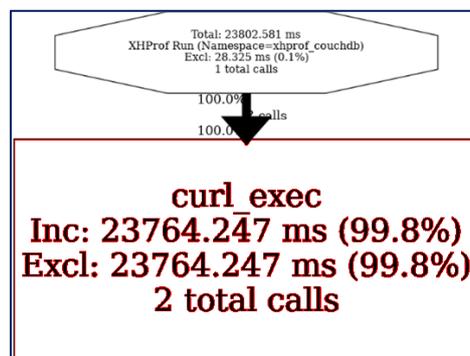
Рис. 2 Граф с результатами измерений извлечения данных из микросервиса СОБД с данными по студентам.

Методы загрузки и извлечения данных в СОБД. Для сравнения результатов измерения производительности тестировали два известных метода извлечения данных в СОБД: 1) метод записи данных целиком, более известный в других СУБД как "bulk-load" [23]; 2) метод, называемый «порционным», позволяющий разбивать большой объем данных на порции изменяемого размера в соответствии с указаниями пользователя. Здесь надо отметить, что заранее известно, что метод загрузки порциями может быть более затратным, поскольку для него в начале и конце требуется инициирование соединения, а затем его разрыв по окончании операции записи в гетерогенное хранилище данных. Примером, где продемонстрирована разница в двух методах для тех же данных, являются результаты, приведенные на рис. 3, где обработка проводится целиком, и на рис. 4, где обработка проводится порционно [24].

Displaying top 100 functions: Sorted by Incl. Wall Time (microsec) [[display all](#)]

Function Name	Calls	Calls%	Incl. Wall Time (microsec)	IWall%	Excl. Wall Time (microsec)	EWall%
main()	1	6.2%	23,802,581	100.0%	28,325	0.1%
curl_exec	2	12.5%	23,764,247	99.8%	23,764,247	99.8%
curl_setopt	8	50.0%	9,290	0.0%	9,290	0.0%
curl_close	2	12.5%	701	0.0%	701	0.0%
curl_init	2	12.5%	18	0.0%	18	0.0%
xhprof_disable	1	6.2%	0	0.0%	0	0.0%

а



б

Рис. 3 Результаты измерений производительности чтения из микросервиса СОБД с записью в документо-ориентированное хранилище CouchDB:
 а – таблица с результатами измерений обработки bulk-load целиком;
 б – граф вызова с результатом измерений обработки bulk-load целиком.

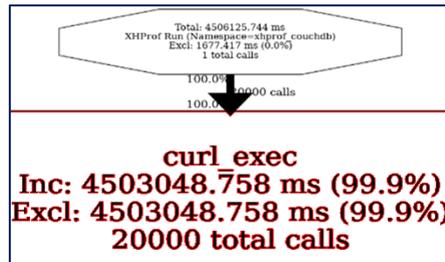
Измерение производительности работы СОБД с документо-ориентированным хранилищем CouchDB. На рис. 3 и 4 приводятся результаты измерений работы СОБД с хранилищем CouchDB – одного из популярных хранилищ обработки больших данных с использованием функций Map-Reduce [25, 26]. Данный пример является реальным практическим примером, рассмотренным в работе по документо-ориентированным хранилищам и СОБД [20].

Измерение производительности работы СОБД с реляционной СУБД. На рис. 3 и 4 приводятся результаты измерений работы СОБД реляционной базой данных в среде MySQL. Методы используются те же: bulk-load (рис. 5) и порционный (рис. 6).

Displaying top 100 functions: Sorted by Incl. Wall Time (microsec) [[display all](#)]

Function Name	Calls	Calls%	Incl. Wall Time (microsec)	IWall%	Excl. Wall Time (microsec)	EWall%
main()	1	0.0%	4,506,125,744	100.0%	1,677,417	0.0%
curl_exec	20,000	18.2%	4,503,048,758	99.9%	4,503,048,758	99.9%
curl_setopt	80,000	72.7%	1,144,877	0.0%	1,144,877	0.0%
json_decode	10,000	9.1%	254,560	0.0%	254,560	0.0%
curl_close	2	0.0%	122	0.0%	122	0.0%
curl_init	2	0.0%	10	0.0%	10	0.0%
xhprof_disable	1	0.0%	0	0.0%	0	0.0%

а



б

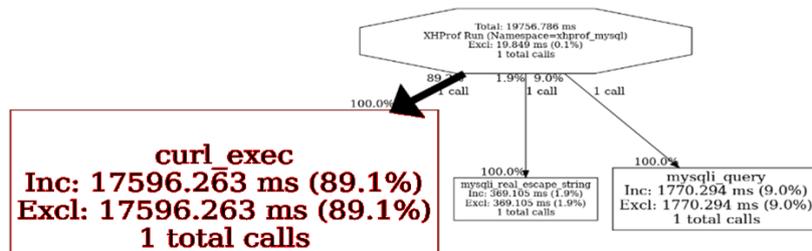
Рис. 4 Результаты измерений производительности чтения из микросервиса СОБД с записью в документ-ориентированное хранилище CouchDB:

а – таблица с результатами измерений порционной обработки;
б – граф вызова с результатом измерений порционной обработки.

Displaying top 100 functions: Sorted by Incl. Wall Time (microsec) [[display all](#)]

Function Name	Calls	Calls%	Incl. Wall Time (microsec)	IWall%	Excl. Wall Time (microsec)	EWall%
main()	1	7.1%	19,756,786	100.0%	19,849	0.1%
curl_exec	1	7.1%	17,596,263	89.1%	17,596,263	89.1%
mysql_query	1	7.1%	1,770,294	9.0%	1,770,294	9.0%
mysql_real_escape_string	1	7.1%	369,105	1.9%	369,105	1.9%
curl_close	1	7.1%	931	0.0%	931	0.0%
mysql_connect	1	7.1%	223	0.0%	223	0.0%
mysql_set_charset	1	7.1%	64	0.0%	64	0.0%
curl_init	1	7.1%	29	0.0%	29	0.0%
mysql_close	1	7.1%	23	0.0%	23	0.0%
curl_setopt	2	14.3%	3	0.0%	3	0.0%
printf	1	7.1%	2	0.0%	2	0.0%
mysql_character_set_name	1	7.1%	0	0.0%	0	0.0%
xhprof_disable	1	7.1%	0	0.0%	0	0.0%

а



б

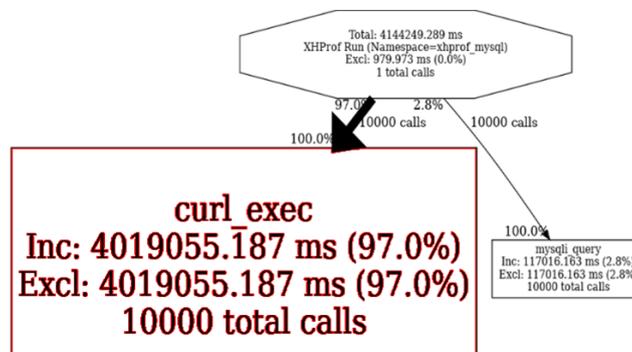
Рис. 5 Результаты измерений производительности чтения из микросервиса СОБД с записью в реляционную СУБД MySQL:

а – таблица с результатами измерений обработки bulk-load целиком;
б – граф вызова с результатом измерений обработки bulk-load целиком.

Displaying top 100 functions: Sorted by Incl. Wall Time (microsec) [[display all](#)]

Function Name	Calls	Calls%	Incl. Wall Time (microsec)	IWall%	Excl. Wall Time (microsec)	EWall%
main()	1	0.0%	4,144,249,289	100.0%	979,973	0.0%
curl_exec	10,000	20.0%	4,019,055,187	97.0%	4,019,055,187	97.0%
mysql_query	10,000	20.0%	117,016,163	2.8%	117,016,163	2.8%
mysql_real_escape_string	10,000	20.0%	7,059,983	0.2%	7,059,983	0.2%
curl_setopt	20,000	40.0%	137,495	0.0%	137,495	0.0%
mysql_connect	1	0.0%	309	0.0%	309	0.0%
curl_close	1	0.0%	76	0.0%	76	0.0%
mysql::set_charset	1	0.0%	67	0.0%	67	0.0%
mysql_close	1	0.0%	26	0.0%	26	0.0%
curl_init	1	0.0%	7	0.0%	7	0.0%
printf	1	0.0%	2	0.0%	2	0.0%
mysql::character_set_name	1	0.0%	1	0.0%	1	0.0%
xhprof_disable	1	0.0%	0	0.0%	0	0.0%

а



б

Рис. 6 Результаты измерений производительности чтения из микросервиса СОБД с записью в реляционную СУБД MySQL:
 а – таблица с результатами измерений порционной обработки;
 б – граф вызова с результатом измерений порционной обработки.

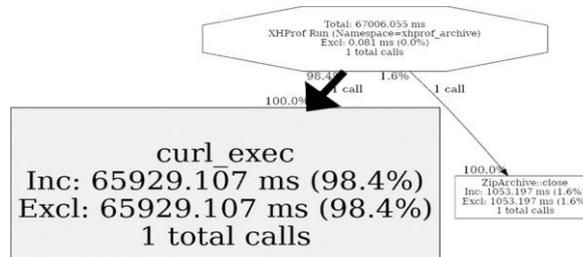
Измерение производительности работы СОБД с офисными документами docx и архивами. На рис. 7 и 8 приводятся примеры таблиц [27] с результатами измерений производительности чтения из микросервиса СОБД с записью в архив docx, а также соответствующий граф вызова. Этот пример демонстрирует не лучшие результаты в связи с тем, что в документах содержатся метки, на поиск и запись в которых может затрачиваться какое-то время. Чтобы избежать таких потерь, было использовано ограниченное количество меток для записи.

Сравнительная диаграмма по времени обработки данных в СОБД. В данной работе на рис. 1 – 8 приведены примеры таблиц и графов вызова функций СОБД с результатами обработки данных [28] по гетерогенным источникам. Поскольку таких рисунков получено значительное количество, нет возможности привести их здесь целиком. Для каждого теста и метода получают несколько таблиц и рисунков, отличающихся друг от друга объемом и цифровыми значениями временной метки измерения. Чтобы сократить данный объем, на рис. 9 объединены результаты для каждого теста в диаграммы, где для каждого метода и теста показаны усредненные данные по объемам от 10 до 10 000 узлов. По диаграммам видно, что характер зависимостей является линейным. А в порционном методе близкий к линейному в связи с тем, что на больших объемах данных образуется очередь на запись в хранилище [29, 30] гетерогенного источника.

Displaying top 100 functions: Sorted by Incl. Wall Time (microsec) [[display all](#)]

Function Name	Calls	Calls%	Incl. Wall Time (microsec)	IWall%	Excl. Wall Time (microsec)	EWall%
main()	1	8.3%	67,006,055	100.0%	81	0.0%
curl_exec	1	8.3%	65,929,107	98.4%	65,929,107	98.4%
ZipArchive::close	1	8.3%	1,053,197	1.6%	1,053,197	1.6%
ZipArchive::addFromString	1	8.3%	11,290	0.0%	11,290	0.0%
substr_replace	1	8.3%	11,194	0.0%	11,194	0.0%
curl_close	1	8.3%	1,044	0.0%	1,044	0.0%
ZipArchive::open	1	8.3%	94	0.0%	94	0.0%
ZipArchive::getFromName	1	8.3%	40	0.0%	40	0.0%
curl_init	1	8.3%	7	0.0%	7	0.0%
curl_setopt	2	16.7%	1	0.0%	1	0.0%
xhprof_disable	1	8.3%	0	0.0%	0	0.0%

а



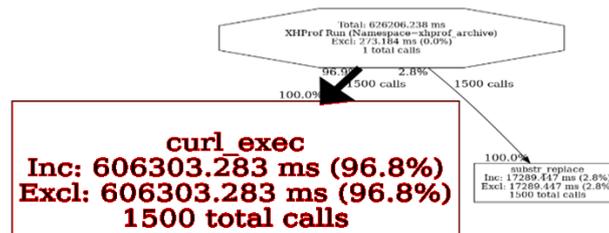
б

Рис. 7 Результаты измерений производительности чтения из микросервиса СОБД с записью в архив docx:
а – таблица с результатами измерений обработки bulk-load целиком;
б – граф вызова с результатом измерений обработки bulk-load целиком.

Displaying top 100 functions: Sorted by Incl. Wall Time (microsec) [[display all](#)]

Function Name	Calls	Calls%	Incl. Wall Time (microsec)	IWall%	Excl. Wall Time (microsec)	EWall%
main()	1	0.0%	626,206,238	100.0%	273,184	0.0%
curl_exec	1,500	20.0%	606,303,283	96.8%	606,303,283	96.8%
substr_replace	1,500	20.0%	17,289,447	2.8%	17,289,447	2.8%
ZipArchive::addFromString	1,500	20.0%	2,095,156	0.3%	2,095,156	0.3%
curl_init	1	0.0%	124,005	0.0%	124,005	0.0%
ZipArchive::close	1	0.0%	96,653	0.0%	96,653	0.0%
curl_setopt	3,000	40.0%	24,286	0.0%	24,286	0.0%
ZipArchive::open	1	0.0%	121	0.0%	121	0.0%
ZipArchive::getFromName	1	0.0%	61	0.0%	61	0.0%
curl_close	1	0.0%	42	0.0%	42	0.0%
xhprof_disable	1	0.0%	0	0.0%	0	0.0%

а



б

Рис. 8 Результаты измерений производительности чтения из микросервиса СОБД с записью в архив docx:
а – таблица с результатами измерений порционной обработки;
б – граф вызова с результатом измерений порционной обработки.

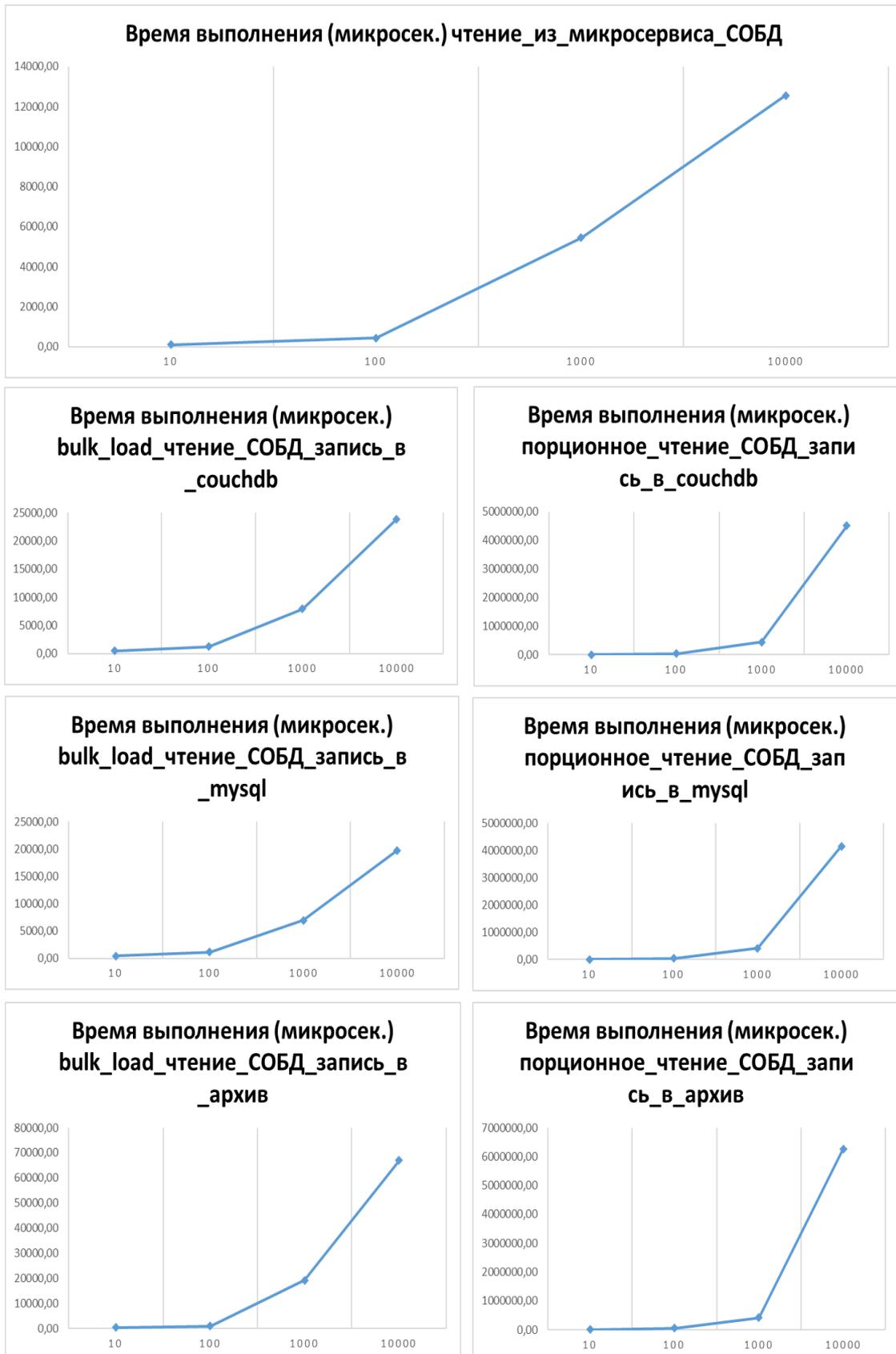


Рис. 9 Диаграммы с результатами измерений производительности СОБД по методам порционной обработки и bulk-load на объемах данных от 10 до 10 000 узлов.

ЗАКЛЮЧЕНИЕ

Данная работа сосредоточена на проведении измерений времени обработки данных в зависимости от их объема в СОБД с привлечением физических хранилищ данных различного назначения. Если раньше вопросы эффективности касались повышения уровня абстракции в создании веб-приложений и сокращения программного кода за счет использования усовершенствованной модели СОБД, то в данной работе приводятся количественные результаты, выражающиеся во временных мерах и касающиеся виртуального хранилища СОБД. Рассмотрены тесты на базе микросервиса, приведены примеры работы со следующими видами гетерогенных источников:

- документо-ориентированное хранилище данных;
- реляционная СУБД;
- офисные документы и архивы.

В качестве тестируемых были выбраны два метода чтения и записи из/в физические хранилища данных: метод bulk-load (то есть «целиком») и порционный метод. По результатам, сведенным в единый комплекс диаграмм, можно прийти к выводу, что порционный метод занимает больше времени по сравнению с методом записи bulk-load, то есть целиком. Это объясняется тем, что в процессе записи многократно повторяются операции установления соединений с хранилищем данных: первый раз во время чтения, второй – во время записи. В некоторых физических хранилищах, например, таких как CouchDB, образуется очередь из порций на запись, которые со временем записываются физически в документы. Тем не менее характер зависимости количества обрабатываемых узлов от времени является линейным. Данный характер зависимости объясняется тем, что с увеличением количества задач растет и время обработки в СОБД. Полученные диаграммы и протоколы с таблицами измерений позволяют в дальнейшем устранять недостатки в части программной реализации с целью повышения эффективности СОБД.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Zongheng Yang et al. "Balsa: Learning a query optimizer without expert demonstrations" // Proceedings of the 2022 International Conference on Management of Data. SIGMOD/PODS '22. Philadelphia, PA, USA: Association for Computing Machinery, 2022. Pp. 931-944.
2. Zongheng Yang et al. "Deep Unsupervised Cardinality Estimation" // Proceedings of the VLDB Endowment. 2019. Vol. 13. 3. VLDB Endowment. Pp. 279-292.
3. Guo Q. et al. "Multi-model query languages: taming the variety of big data" // Distributed and Parallel Databases, 2023.
4. Györfödi C. A. et al. "Performance analysis of NoSQL and relational databases with CouchDB and MySQL for application's data storage" // Applied Sciences. 2020. Vol. 10. No. 23. Pp. 8524.
5. Zongheng Yang et al. "NeuroCard: One cardinality estimator for all tables" // Proceedings of the VLDB Endowment. 2021. Vol. 14. 1. VLDB Endowment. Pp. 61-73.
6. Zongheng Yang et al. "Qd-tree: learning data layouts for big data analytics" // Proceedings of the 2020 International Conference on Management of Data. SIGMOD '20. 2020.
7. Magableh B., Almiani M. "A self-healing microservices architecture: a case study in docker swarm cluster" // Advances in Intelligent Systems and Computing. 2020. Vol. 926. Pp. 846-858.
8. Hamzehloui M. S., Sahibuddin S., Salah K. "A systematic mapping study on microservices" // Advances in Intelligent Systems and Computing. 2019. Vol. 843. Pp. 1079-1090.
9. Huf A., Siqueira F. "Composition of heterogeneous web services: A systematic review" // Journal of Network and Computer Applications. 2019. Vol. 143. Pp. 89-110.
10. Миронов В. В., Гусаренко А. С., Тугузбаев Г. А. Ситуационно-ориентированные базы данных: формирование персонализированных графических документов для поддержки учебного проектирования // Моделирование, оптимизация и информационные технологии. 2020. Т. 8. № 2 (29). С. 1–18. [[Mironov V. V., Gusarenko A. S., Tuguzbaev G. A. "Situation-oriented databases: the formation of personalized graphic documents for educational design support" // Modelirovanie, Optimizatsiya i Informatsionnye Tekhnologii = Modeling, Optimization and Information Technology, 2020, vol. 20, no. 4, pp. 1-18. (In Russian).]]
11. Abdullah M., Iqbal W., Erradi A. "Unsupervised learning approach for web application auto-decomposition into microservices" // Journal of Systems and Software. 2019. Vol. 151. Pp. 243–257.

12. Миронов В. В., Гусаренко А. С., Диметриев Р. Р., Сарваров М. Р. Создание персонализированных документов на основе ситуационно-ориентированной базы данных // Вестник Уфимского государственного авиационного технического университета. 2014. Т. 18. № 4 (65). С. 191–197. [[Mironov V. V., Gusarenko A. S. et al. "The personalized documents generating using DOM-objects in situation-oriented databases" // Vestnik UGATU, 2014, vol. 18, no. 4 (65), pp. 191-197. (In Russian).]]
13. Janga P., Davis K. C. "Mapping heterogeneous XML document collections to relational databases" // Lecture Notes in Computer Science. 2014. Vol. 8824. Pp. 86-99.
14. Миронов В. В., Гусаренко А. С., Тугузбаев Г. А. Извлечение семантической информации из графических схем // Информатика и автоматизация. 2021. Т. 20. № 4. С. 940–970. [[Mironov V. V., Gusarenko A. S., Tuguzbaev G. A. In: Informatika i Avtomatizatsiya, 2021, vol. 20, no. 4, pp. 940-970. (In Russian).]]
15. Миронов В. В., Юсупова Н. И., Гусаренко А. С. Ситуационно-ориентированные базы данных: современное состояние и перспективы исследования // Вестник Уфимского государственного авиационного технического университета. 2015. Т. 19. № 2 (68). С. 188–199. [[Mironov V. V., Yusupova N. I., Gusarenko A. S. "Situation-oriented databases: current state and prospects for research". In: Vestnik UGATU, 2015, vol. 19, no. 2 (68), pp. 188-199. (In Russian).]]
16. Миронов В. В., Гусаренко А. С., Юсупова Н. И. Структурирование виртуальных мультидокументов в ситуационно-ориентированных базах данных с помощью entry-элементов // Труды СПИИРАН. 2017. № 4 (53). С. 225–240. [[Mironov V. V., Gusarenko A. S., Yusupova N. I. "Structuring virtual multi-documents in situationally-oriented databases by means of entry-elements". In: Trudy SPIIRAN, 2017, no. 4 (53), pp. 255-240. (In Russian).]]
17. Миронов В. В. Выявление плагиата в графических конструкторских документах в ходе учебного проектирования // Системная инженерия. 2023. Т. 5. № 1 (10). С. 56–66. [[Mironov V. V. "Identify plagiarism in graphic design documents during educational design". In: Sistemnaja inzhenerija = System Engineering and Information Technologies, 2023, no. 1 (10), pp. 56-66. (In Russian).]]
18. Миронов В. В., Гусаренко А. С., Юсупова Н. И. Мониторинг просмотров видео YouTube в образовательной среде на основе ситуационно-ориентированной базы данных и веб-сервисов RESTful // Системная инженерия. 2021. Т. 3. № 1(5). С. 39–49. [[Mironov V. V., Gusarenko A. S., Yusupova N. I., "Monitoring youtube video views in the educational environment based on situation-oriented database and RESTful web services". In: Sistemnaja inzhenerija = System Engineering and Information Technologies, 2021, vol. 3, no. 1 (5), pp. 39-49. (In Russian).]]
19. Гусаренко А. С. Ситуационно-ориентированные базы данных: верификация электронных подписей гетерогенных документов в RESTful веб-сервисе // Моделирование, оптимизация и информационные технологии. 2022. Т. 10. № 4. С. 1–17. [[Gusarenko A. S., "Situation-oriented databases: verifying electronic signatures of heterogeneous documents in a RESTful web service". In: Modelirovanie, Optimizatsiya i Informatsionnye Tekhnologii = Modeling, Optimization and Information Technology, 2022, vol. 10, no. 4, pp. 1-17. (In Russian).]]
20. Гусаренко А. С. Ситуационно-ориентированные базы данных: обработка гетерогенных документов микросервисов в документо-ориентированном хранилище // Моделирование, оптимизация и информационные технологии. 2022. Т. 10. № 4. С. 1–16. [[Gusarenko A. S., "Situation-oriented databases: processing heterogeneous documents of microservices in a document-based storage". In: Modelirovanie, Optimizatsiya i Informatsionnye Tekhnologii = Modeling, Optimization and Information Technology, 2022, vol. 10, no. 4, pp. 1-16. (In Russian).]]
21. Миронов В. В., Гусаренко А. С., Юсупова Н. И. Ситуационно-ориентированные базы данных: обработка офисных документов // Моделирование, оптимизация и информационные технологии. 2022. Т. 10. № 2. С. 1–16. [[Mironov V. V., Gusarenko A. S., Yusupova N. I. "Situation-oriented databases: processing office documents". In: Modelirovanie, Optimizatsiya i Informatsionnye Tekhnologii = Modeling, Optimization, and Information Technology, 2022, vol. 10, no. 2, pp. 1-16. (In Russian).]]
22. Гусаренко А. С. Ситуационно-ориентированные базы данных: резервное копирование виртуальных мультидокументов модели динамических объектов обработки данных // Моделирование, оптимизация и информационные технологии. 2023. Т. 11. № 2. С. 1–19. [[Gusarenko A. S., "Situation-oriented databases: backing up virtual multi-documents of dynamic data processing objects model". In: Modelirovanie, Optimizatsiya i Informatsionnye Tekhnologii = Modeling, Optimization, and Information Technology, 2023, vol. 11, no. 2, pp. 1-19. (In Russian).]]
23. Papadopoulos A., Manolopoulos Y. "Parallel bulk-loading of spatial data" // Parallel Computing. 2003. vol. 29, no. 10, pp. 1419-1444.
24. Миронов В. В., Гусаренко А. С., Юсупова Н. И. Программное извлечение данных из word-документов на основе ситуационно-ориентированного подхода // Прикладная информатика. 2021. Т. 16. № 6. С. 66–83. [[Mironov V. V., Gusarenko A. S., Yusupova N. I. "Soft extract data from word-based documents situationally oriented approach". In: Prikladnaja informatika = Journal of Applied Informatics, 2021, vol. 16, no. 6, pp. 66-83. (In Russian).]]
25. Seera N. K., Taruna S. "Analyzing cost parameters affecting map reduce application performance" // International Journal of Information Technology and Computer Science. 2016. Vol. 8. No. 8.
26. Jiang D. et al. "The performance of MapReduce: an in-depth study" // Proc. VLDB Endow. 2010. Vol. 3. No. 1–2. Pp. 472-483.
27. Любченко М. А. Об одном опыте анализа данных и извлечения информации о программном продукте // Системная инженерия и информационные технологии. 2021. Т. 3. № 2 (6). С. 75–80. [[Lyubchenko M. A. "About one experience of data analysis and information extraction about a software product" // System Engineering and Information Technologies, 2021, vol. 3, no. 2 (6), pp. 75-80. (In Russian).]]
28. Закирова Э. Ф., Павлов С. В., Трубин В. Д., Христовуло О. И. Детализация пространственной информации для обеспечения защищенности баз данных в распределенных информационных системах // Системная инженерия и информационные технологии. 2022. Т. 4. № 1 (8). С. 20–26. [[Zakirova E. F., Pavlov S. V., Trubin V. D., Khristodulo O. I. "Detailing

of spatial information to ensure database security in distributed information systems" // System Engineering and Information Technologies, 2022, vol. 4, no. 1 (8), pp. 20-26. (In Russian).]]

29. Климова А. В., Ковтуненко А. С., Парфенов Д. В. Формирование учебного плана с использованием технологии RDF-хранилищ // Системная инженерия и информационные технологии. 2022. Т. 4. № 2 (9). С. 49–55. [[Klimova A. V., Kovtunencko A. S., Parfenov D. V. "Curriculum development using RDF-storage technology" // System Engineering and Information Technologies, 2022, vol. 4, no. 2 (9), pp. 49-55. (In Russian).]]

30. Махмутов А. Р., Вульфин А. М., Миронов К. В. Исследование возможностей автономной работы конечных устройств интернета вещей // Системная инженерия и информационные технологии. 2022. Т. 5. № 1 (10). С. 41–47. [[Makhmutov A. R., Vulfin A. M., Mironov K. V. "Study of the possibilities of autonomous operation of end devices of the Internet of Things" // System Engineering and Information Technologies, 2022, vol. 5, no. 1 (10), pp. 41–47. (In Russian).]]

Поступила в редакцию 25 июля 2023 г.

МЕТАДАННЫЕ / METADATA

Title: Query performance to heterogeneous sources in situation-oriented databases.

Abstract: In the field of database development, active movement continues towards the creation of non-relational NoSQL databases. The issues of the efficiency of relational database functioning were considered in modern published works in terms of query optimization, measuring the execution time of SQL expressions, software tools for testing were already implemented and known here, and types of tests for measuring DBMS performance were defined. These tools and tests are easily ported from one relational DBMS to another to measure and evaluate query performance. At the same time, for NoSQL databases, only the volumes of data for which it is required to measure the processing time by the core of a non-relational DBMS can be common. In this regard, it is difficult to find universal and easily portable tools for testing NoSQL DBMS. At a similar stage of development are situationally oriented databases - a subspecies of NoSQL databases created within the framework of the SODB project. At the same time, it is possible to measure performance by using specialized profiling tools on the PHP platform by embedding them into the SODB program code. This work continues a series of studies on SODB and considers the issues of measuring and evaluating performance using examples of working with heterogeneous data sources in the SODB microservice, document-oriented storage, office documents, archives, relational DBMS, and archives. Especially for this, variants of examples of the use of SODB were developed or adapted. All obtained measurement results are accompanied by tables and graphic protocols of graphs for calling SODB program functions for methods of loading data volumes in whole (Bulk Load) and in portions (Portion Load). The results obtained allow us to conclude that the results are comparable in terms of data processing with other NoSQL databases and move further towards improving the efficiency of the SODB.

Key words: situationally oriented database; built-in dynamic model; heterogeneous sources of documents; virtual multidocuments; dynamic data processing objects; RESTful services; performance tests

Язык статьи / Language: русский / Russian.

Об авторе / About the author:

ГУСАРЕНКО Артем Сергеевич

ФГБОУ ВО «Уфимский университет науки и технологий», Россия.
Доцент. каф. автоматизированных систем управления. Дипл. информатик-экономист (Уфимск. гос. авиац. техн. ун-т, 2010).
Канд. техн. наук по матем. и прогр. обесп. выч. машин, комплексов и комп. сетей (там же, 2013). Иссл. в обл. иерархических моделей и ситуационного управления.
E-mail: gusarenko@ugatu.su
ORCID: <https://orcid.org/0000-0003-4132-6106>
URL: elibrary.ru/author_profile.asp?authorid=692073

GUSARENKO Artem Sergeevich

Ufa University of Science and Technology (UUST), Russia.
Docent, Dept. of Automated Control Systems. Dipl. Informatic economist (Ufa State Aviation Technical University, 2010).
Cand. of Tech. Sci. on mathematical and software support for computers, complexes, and computer networks (ibid, 2013). Research in the field of hierarchical models and situational management.
E-mail: gusarenko@ugatu.su
ORCID: <https://orcid.org/0000-0003-4132-6106>
URL: elibrary.ru/author_profile.asp?authorid=692073