

Н. Г. Кучук¹, В. Ю. Мерлак², В. В. Скородєлов¹

¹ Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

² Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна

МЕТОД ЗМЕНШЕННЯ ЧАСУ ДОСТУПУ ДО СЛАБКОСТРУКТУРОВАНИХ ДАНИХ

Анотація. На сьогодні актуальною є проблема зберігання та обробки слабкоструктурованої інформації. Суть проблеми є в тому, що слабкоструктурованість і великі обсяги даних призводять до порушення вимог QoS, зокрема, збільшується час доступу до даних і відповідно зростає час очікування реакції на запит. Крім того, істотно зростають витрати на зберігання даних. **Предметом дослідження** є методи опрацювання та зберігання слабкоструктурованих та неструктурованих даних. **Мета дослідження** – розроблення методу зменшення часу доступу до неструктурованих даних шляхом формування системи ієрархічних віджетів та планування доступу до системних ресурсів множинних запитів транзакцій окремих віджетів, що створюють черги при доступі до нереляційних сховищ даних. **Результати.** Проаналізовані варіанти структури сховищ даних для опрацювання та зберігання слабкоструктурованих даних. Обґрунтовано застосування системи ієрархічних віджетів при роботі із множинними запитами. Запропоновано метод перерозподілу черг до системних ресурсів. Головним критерієм при перерозподілі обрано вчасне та якнайшвидше завершення всіх транзакцій системи ієрархічних віджетів. Розроблений метод перепланування, котрий забезпечує зменшення часу обробки запитів транзакцій системи ієрархічних віджетів. **Висновки.** Показана необхідність динамічного перепланування процесу розподілу ресурсів, що необхідні для підтримки системи ієрархічних віджетів. Запропоновано метод розподілу ресурсів між транзакціями системи ієрархічних віджетів при обробці неструктурованою інформацією в нереляційних сховищах даних. Оптимізація проводилася за критерієм якнайшвидшого завершення всіх транзакцій за рахунок скорочення часу виконання найтриваліших запитів.

Ключові слова: віджет; слабкоструктурована інформація; сховище даних; транзакція.

Вступ

На сьогодні, загальний обсяг даних, що зберігаються у різних мережах та системах постійно збільшується, дані продовжують накопичуватися з величезною швидкістю практично у всіх сферах людської діяльності. Вважаючи на те, що до складу даних, що зберігаються, можуть входити різноманітні типи даних (фото, відео, аудіо, текст, дані геопозиціонування тощо) і ці дані часто оновлюються і доповнюються, то можна вважати, що **актуальною** є проблема зберігання та обробки слабкоструктурованої інформації. Суть проблеми є в тому, що слабкоструктурованість і великі обсяги даних призводять до порушення вимог QoS, зокрема, збільшується час доступу до даних і відповідно зростає час очікування реакції на запит. Крім того, істотно зростають витрати на зберігання даних.

Аналіз проблеми та наукових публікацій. Питання, котрі пов'язані з проблемами, що виникають при роботі з великими даними, особливо при зберіганні та обробці неструктурованої інформації, розглянуті в [1-5]. Різні підходи до вирішення даних проблем в різних спеціалізованих системах, запропоновані в [6-9]. Методи оптимального розподілу ресурсів при зростанні обсягів даних запропоновані в [10-21]. Однак в цих роботах не розглядаються питання зниження часу реакції на запити транзакцій з одночасним зниженням витрат на зберігання даних.

Постановка завдання. Метою статті є розроблення методу зменшення часу доступу до неструктурованих даних шляхом формування системи ієрархічних віджетів та планування доступу до системних ресурсів множинних запитів транзакцій окремих віджетів, що створюють черги при доступі до нереляційних сховищ даних.

Вибір варіанту структури сховищ даних для опрацювання та зберігання слабкоструктурованих даних

Для опрацювання та зберігання слабкоструктурованих даних розглянемо варіанти використання сховищ даних (СД), що мають реляційну (*SQL*) або нереляційну (*NoSQL*) структуру. Застосування реляційної структури для зберігання неструктурованих даних призводить до надмірності при виділенні зовнішньої пам'яті для зберігання інформації. Особливо негативно даний недолік позначається при роботі з Big Data, суттєво збільшується вартість зберігання даних.

Для роботи з неструктурованими даними нереляційні сховища даних пропонують гнучкі схеми розробки, котрі сприяють тому, що зайва пам'ять не резервується, а для прискорення доступу до даних в більшості випадків використовуються ієрархічні структури.

Крім того, потрібно ще враховувати той факт, що при використанні реляційних структур доступ до об'єктів виконується за ключем, а в нереляційних СД доступ до документа може виконуватися за індексом, що істотно збільшує час пошуку потрібного об'єкту. На рис. 1. 2 наведені узагальнені результати експериментів, проведених з метою оцінки характеристик, які впливають на якість обслуговування та вартість системи.

При формалізації результатів використовувались такі позначення:

Vol – обсяг даних, що призначені для зберігання з подальшим опрацюванням, терабайт (Тб);

Vol_z – загальний розмір зайнятої зовнішньої пам'яті для зберігання даних обсягом Vol , Тб;

$Vol_{exc} = Vol_z - Vol$ – умовні непродуктивні витрати зовнішньої пам’яті, Тб;

T_{SQL}, T_{NoSQL} – загальний час обробки певної сукупності транзакцій комп’ютерної системи при використанні реляційних та нереляційних структур зберігання даних відповідно, год.;

$\zeta = T_{SQL} / T_{NoSQL}$ – відносна затримка часу виконання транзакцій, що розглядається як показник оцінки якості обслуговування (за часом) для порівняння реляційних та нереляційних структур;

Так, аналізуючи графіки, наведені на рис. 1, робимо висновок щодо суттєвого зростання непродуктивних витрат зовнішньої пам’яті при перевищенні загального обсягу даних, що призначені для зберігання, значення 40 Тб.

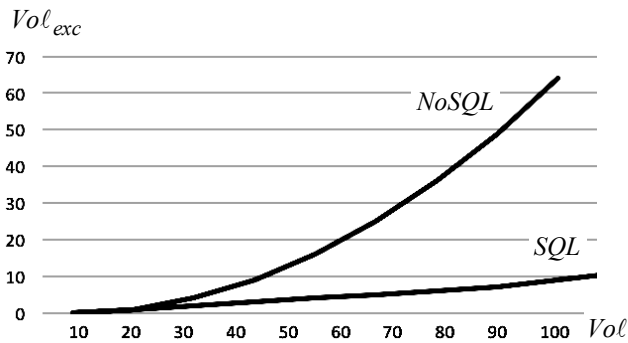


Рис. 1. Аналіз непродуктивних витрат зовнішньої пам’яті

Але індексний доступ до *NoSQL*-сховищ даних займає більше часу, ніж доступ за ключем до *SQL*-сховищ. Це наочно бачимо на рис. 2, причому чим більшим є виграш за зовнішньою пам’яттю, тим більшим є програш за часом доступу до даних, тобто збільшується значення показника оцінки якості обслуговування за часом. Це стає суттєвою проблемою, особливо при різних аспектах обробки Big Data.

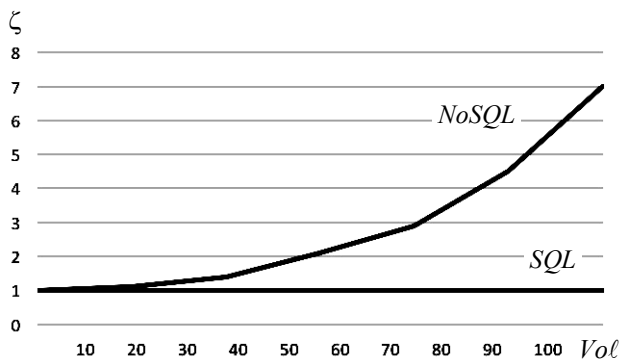
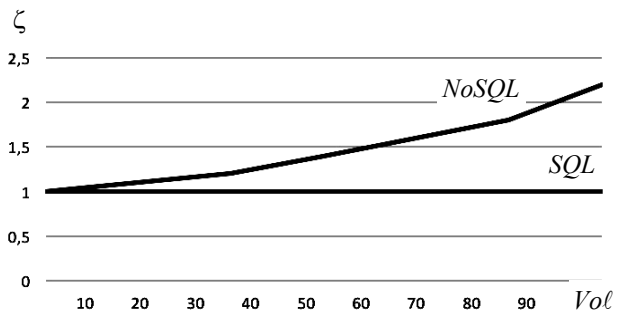


Рис. 2. Залежність відносної затримки часу виконання транзакцій від обсягу даних (Тб) V

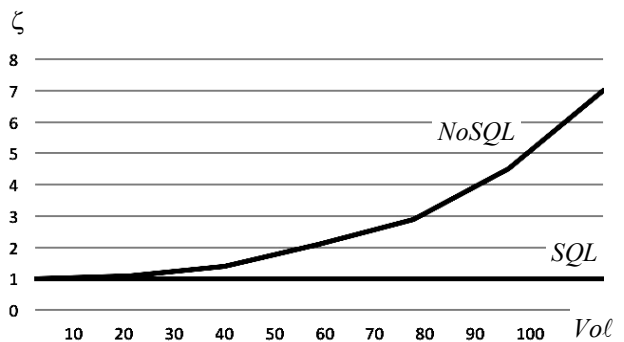
Один з існуючих підходів для вирішення даної проблеми передбачає використання ситуаційно-орієнтованих сховищ даних (СОСД). При цьому є суттєва різниця при обробці простих та множинних запитів, сформованих транзакціями комп’ютерної системи.

Як бачимо на рис. 3, а, швидкість обробки простих запитів транзакцій при цьому збільшується десь в 1,2-1,8 разів та наближається до часу запиту в реляційних структурах.

Але для множинних запитів покращення ситуації з великим часом обробки запитів не спостерігається (рис. 3, б).



а – прості запити



б – множинні запити

Рис. 3. Відносна затримка часу виконання запитів *NoSQL* у порівнянні із запитами *SQL*

Для зменшення сумарного часу отримання даних з використанням множинних запитів пропонується результати запитів зберігати в контейнері і відображати результат після того, як всі вони будуть виконані, тобто побудувати ієрархічну систему віджетів. Такий підхід суттєво зменшує час на отримання даних з використанням запитів *NoSQL* (рис. 4).

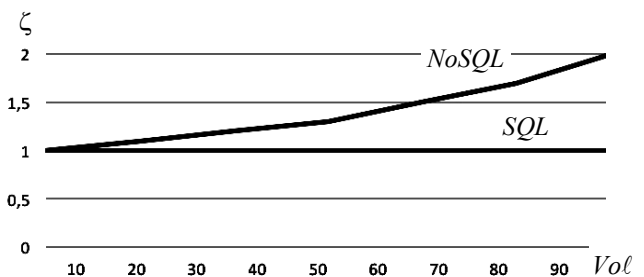


Рис. 4. Аналіз застосування ієрархічної системи віджетів

Але в цьому випадку швидкість починає спадати у випадку збільшення розмірності системи ієрархічних віджетів (*N*), що наочно представлено на рис. 5.

Основна причина цього – затримки в чергах до потрібних ресурсів. Тому необхідно перепланування запитів до ресурсів, котре зменшить часи чекання ресурсів.

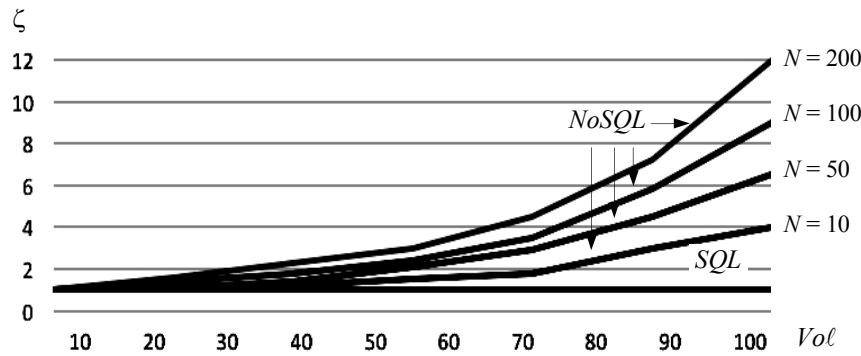


Рис. 5. Залежність відносної затримки часу виконання запитів NoSQL у порівнянні із запитом SQL від розмірності системи ієрархічних віджетів

Метод розподілу ресурсів між транзакціями системи ієрархічних віджетів

Для вирішення проблеми оптимального розподілу ресурсів між транзакціями системи ієрархічних віджетів розробимо математичну модель даного процесу. Зробимо такі припущення:

- процес моделюється на протязі фіксованого часового проміжку \aleph ;
- проведена декомпозиція запитів транзакцій на елементарні запити;
- проведена декомпозиція ресурсів на елементарні ресурсні одиниці таким чином, що одночасно один елементарний запит може звертатися до одного елементарного ресурсу.

Нехай на протязі часового проміжку \aleph необхідно виконати N запитів системи ієрархічних віджетів, n – номер поточного запиту, $n \in \overline{1, N}$. Дані запити потребують R різноманітних системних ресурсів, r – номер поточного ресурсу, $r \in \overline{1, R}$. Також формується бінарна прямокутна матриця поточної потреби запитів у системних ресурсах $Q = (q_{nr})$, $\text{card } Q = N \times R$, у якій кожен елемент q_{nr} є ненулевим, коли на даний момент часу n -й запит потребує ресурсу r . Зазначимо, що на початку часового проміжку \aleph у n -му рядку матриці Q одиничними є елементи у стовпцях з номерами, що відповідають потрібним для n -го запиту системним ресурсам. Відповідно, по завершенню проміжку \aleph даний рядок буде нульовим у випадку успішного завершення n -го запиту.

Вважаючи на те, що кожен запит має використовувати декілька елементарних ресурсів, які у загальному випадку взаємопов'язані, необхідно сформулювати множину графів $G = \{G_n\}$, $\text{card } G = N$, де G_n – граф, що відображає ієрархічну послідовність використання системних ресурсів. При цьому перший рівень відповідає цим ресурсам, використання яких не потребує попередників, а останній – ресурсам, які не є попередниками.

Крім того, важливим при розподілі ресурсів є вектор пріоритетів запитів

$$P = (p_n), \text{ card } P = N.$$

Також кожний ресурс r може бути охарактеризованим часом, що є потрібним на отримання умовної одиниці ресурсу t_r , дані значення складають вектор $T = (t_r)$. Відповідно складемо матрицю $S = (s_{nr})$, $\text{card } S = N \times R$, де s_{nr} – обсяг ресурсу r у умовних одиницях для потреб n -го запиту. Вважаючи на ймовірну конкуренцію між запитом за ресурс, визначимо матрицю значень часу очікування у черзі за ресурс $\mathfrak{S} = (\tau_{nr})$, $\text{card } \mathfrak{S} = N \times R$, де τ_{nr} – час очікування n -го запиту у черзі ло ресурсу r .

Для опису методу розподілу необхідно обрати критерій якості розподілу. Головним критерієм визначимо вчасне та якнайшвидше завершення всіх транзакцій системи ієрархічних віджетів. Це значить, що останній запит повинен бути завершеним якомога раніше. Відносний час завершення n -го запиту може бути розрахований таким чином:

$$t_{n_rel_end} = \sum_{r=1}^R (s_{nr} \cdot t_r + \tau_{nr}). \quad (1)$$

Тоді вищенаведений критерій можна формалізувати таким чином:

$$\max_n (t_{n_rel_end}) \rightarrow \min_{\tau_{nr}} \quad (2)$$

або

$$\max_n \left(\sum_{r=1}^R (s_{nr} \cdot t_r + \tau_{nr}) \right) \rightarrow \min_{\tau_{nr}}. \quad (3)$$

Зменшити час виконання запиту, що виконується найдовше, можливо за рахунок зменшення чекання ресурсів у чергах. Для цього пропонується така послідовність дій.

Нехай у момент часу (відносно значення t_{η_rel}) розглядається черга конкуруючих запитів до ресурсу з номером r_{η} . Черга складається із декількох запитів, що складають тимчасовий вектор

$$V_{\eta} = (n_{\xi_1}, \dots, n_{\xi_{\eta}}), \text{ card } V_{\eta} = \xi_{\eta}. \quad (4)$$

Для кожного елемента вектора V_{η} розрахуємо його вагу, котра визначить місце у черзі, зокрема,

тій запит, що буде отримувати ресурс першим. Для цього розраховуємо для кожного запиту час, що буде потрібним після отримання ресурсу r_{η} :

$$t_{\xi_i_rel} = \sum_{r \neq r_{\eta}} (q_{\xi_i r} \cdot s_{\xi_i r} \cdot t_r), \quad \xi_i \in \overline{\xi_1, \xi_{\eta}} \quad (5)$$

При визначенні ваги врахуємо також пріоритети запитів, наведені у векторі $P = (p_n)$:

$$\begin{aligned} v_{\xi_i_rel} &= p_{\xi_i} \cdot t_{\xi_i_rel} = \\ &= p_{\xi_i} \cdot \sum_{r \neq r_{\eta}} (q_{\xi_i r} \cdot s_{\xi_i r} \cdot t_r), \quad \xi_i \in \overline{\xi_1, \xi_{\eta}} \end{aligned} \quad (6)$$

Тоді першим для отримання ресурсу з номером r_{η} буде запит n_{η} , для якого виконана рівність

$$v_{n_{\eta}_rel} = \max_{\xi_i \in \overline{\xi_1, \dots, \xi_{\eta}}} v_{\xi_i_rel} \quad (7)$$

Результати експериментів

Для дослідження можливостей запропонованого методу було проведено експерименти з різними системами ієрархічних віджетів. Розглядалися варіанти різної розмірності розв'язуваної оптимізаційної задачі. Узагальнені результати експерименту показані на рис. 6.

Аналіз графіків рис. 6 свідчить про доцільність використання запропонованого методу при роботі з великими обсягами слабкоструктурованої інформації.

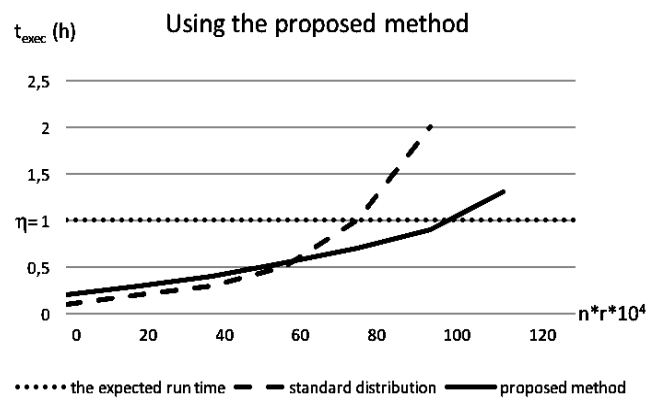


Рис. 6. Відносний час завершення системи ієрархічних віджетів

Висновок

Були проведені дослідження, які обґрунтовують використання системи ієрархічних віджетів для обробки великих даних. Показана необхідність динамічного перепланування процесу розподілу ресурсів, що необхідні для підтримки системи ієрархічних віджетів. Запропоновано метод розподілу ресурсів між транзакціями системи ієрархічних віджетів при обробці неструктурованою інформації в нереляційних базах даних. Оптимізація проводилася за критерієм якнайшвидшого завершення всіх транзакцій за рахунок скорочення часу виконання найтриваліших запитів.

Напрямок подальших досліджень – реалізація запропонованого методу зі зняттям обмежень.

СПИСОК ЛІТЕРАТУРИ

- Kuchuk G., Kovalenko A., Komari I.E., Svyrydov A., Kharchenko V. Improving big data centers energy efficiency: Traffic based model and method. *Studies in Systems, Decision and Control*, vol 171. Kharchenko, V., Kondratenko, Y., Kasprzyk, J. (Eds.). Springer Nature Switzerland AG, 2019. Pp. 161-183. DOI: http://doi.org/10.1007/978-3-030-00253-4_8
- Кучук, Г.А. Синтез стратифікованої інформаційної структури інтеграційної компоненти гетерогенної складової Єдиної АСУ Збройними Силами України / Г.А. Кучук, О.П. Давікоза // *Наука і техніка Повітряних Сил Збройних Сил України: науково-технічний журнал*. – Х.: ХУ ПС, 2013. – № 3(12). – С. 154-158.
- Коваленко А.А., Кучук Г.А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: <https://doi.org/10.20998/2522-9052.2018.1.04>
- Kuchuk G., Nechausov S., Kharchenko, V. Two-stage optimization of resource allocation for hybrid cloud data store. *International Conference on Informp.ation and Digital Technologies*. 2015. P. 266-271.
- Кучук Г.А. Розрахунок навантаження мультисервісної мережі / Г.А. Кучук, Я.Ю. Стасєва, О.О. Болюбаш // *Системи озброєння і військова техніка*. – 2006. – № 4 (8). – С. 130 – 134.
- Кучук Г. А. Метод синтезу інформаційної структури зв'язного фрагменту корпоративної мультисервісної мережі / Г.А. Кучук // *Збірник наукових праць Харківського університету Повітряних сил*. – 2013. – № 2(35). – С. 97-102.
- Шматков С. І. Модель інформаційної структури гіперконвергентної системи підтримки електронних обчислювальних ресурсів університетської e-learning / С. І. Шматков, Н. Г. Кучук, В. В. Донець // *Системи управління, навігації та зв'язку: науковий журнал*. – Полтава: ПНТУ, 2018. – Вип. 2 (48). – С. 97-100.
- Donets V., Kuchuk N., Shmatkov S. Development of software of e-learning information system synthesis modeling process. *Сучасні інформаційні системи*. 2018. Т. 2, № 2. С. 117–121. DOI: <https://doi.org/10.20998/2522-9052.2018.2.20>
- Kuchuk N. Method for calculating of R-learning traffic peakedness / N. Kuchuk; O. Mozhaiev, M. Mozhaiev; H. Kuchuk // 2017 4th International Scientific-Practical Conference Problems of Infocommunications Science and Technology, PIC S and T 2017. – 2017. – P. 359 – 362. URL : <http://dx.doi.org/10.1109/INFOCOMMST.2017.8246416>.
- Merlac V. Resources Distribution Method of University e-learning on the Hyperconvergent platform / V. Merlac, S. Smatkov, N. Kuchuk // *DESSERT'2018*. May 24-27, 2018. – P. 136-140. – URL : <http://dx.doi.org/10.1109/DESSERT.2018.8409114>
- Зиков І. С., Кучук Н. Г., Шматков С. І. Синтез архітектури комп'ютерної системи управління транзакціями e-learning. *Сучасні інформаційні системи*. 2018. Т. 2, № 3. С. 60–66. DOI: <https://doi.org/10.20998/2522-9052.2018.3.10>
- Кучук Г.А. Управління трафіком мультисервісної розподіленої телекомунікаційної мережі / Г.А. Кучук // *Системи управління, навігації та зв'язку*. – К.: ЦНДІ НіУ, 2007. – Вип. 2. – С. 18-27.
- Кучук Г. А., Можасєв О. О., Воробйов О. В. Метод агрегування фрактального трафіка. *Радіоелектронні та комп'ютерні системи*. 2006. № 6 (18). С. 181–188.
- Кучук Г. А. Минимизация загрузки каналов святы вычислительной сети / Г.А. Кучук // *Системи обробки інформації*. – Х.: НАНУ, ПАНМ, ХВУ, 1998. – Вип. 1(5). – С. 149-154..

15. Коваленко А. А. Оптимальное управление трафиком мультисервисной сети на основе методов последовательного улучшения решений / А.А. Коваленко, Г.А. Кучук // Системи озброєння і військова техніка. – 2016. – № 3(47). – С. 59-63.
16. Amin Salih M., Potrus M.Y. A Method for Compensation of TCP Throughput Degrading During Movement Of Mobile Node, *ZANCO Journal of Pure and Applied Sciences*. 2015. Vol. 27, No 6. P. 59–68.
17. Mozhaev O. Multiservice network security metric / O. Mozhaev, H. Kuchuk, N. Kuchuk, M. Mozhaev, M. Lohvynenko // IEEE Advanced information and communication technologies-2017. Proc. of the 2th Int. Conf. – Lviv, 2017. – P. 133-136. – DOI: <https://doi.org/10.1109/AIACT.2017.8020083>
18. Коваленко А.А. Использование временных шкал при аппроксимации длины очередей компьютерных сетей / А.А. Коваленко, Г.А. Кучук, И.В. Рубан // Сучасний стан наукових досліджень та технологій в промисловості. – 2018. – № 2 (4). – С. 12–18. – DOI: <http://doi.org/10.30837/2522-9818.2018.4.012>
19. Кучук Г. А. Метод параметрического управления передачей данных для модификации транспортных протоколов беспроводных сетей / Г.А. Кучук, А.С. Мохаммад, А.А. Коваленко // Системи обробки інформації. – 2011. – № 8(98). – С. 211-218.
20. Кучук Г.А. Метод мінімізації середньої затримки пакетів у віртуальних з'єднаннях мережі підтримки хмарного сервісу / Г.А. Кучук, А.А. Коваленко, Н.В. Лукова-Чуйко // Системи управління, навігації та зв'язку. – Полтава. ПНТУ, 2017. – Вип. 2(42). – С. 117.-120.
21. Свиридов А. С., Коваленко А. А., Кучук Г. А. Метод перерозподілу пропускної здатності критичної ділянки мережі на основі удосконалення ON/OFF-моделі трафіку. *Сучасні інформаційні системи*. 2018. Т. 2, № 2. С. 139–144. DOI: <https://doi.org/10.20998/2522-9052.2018.2.24>

REFERENCES

1. Kuchuk, G., Kovalenko, A., Komari, I.E., Svyrydov, A. and Kharchenko, V. (2019), “Improving big data centers energy efficiency: Traffic based model and method”, *Studies in Systems, Decision and Control*”, vol 171, Kharchenko, V., Kondratenko, Y., Kasprzyk, J. (Eds.), Springer Nature Switzerland AG, pp. 161-183, DOI: http://doi.org/10.1007/978-3-030-00253-4_8.
2. Kuchuk, G.A. and Davikoza, O.P. (2013), “Synthesis of the stratified information structure of the integration component of the heterogeneous component of the Unified ACS of the Armed Forces of Ukraine”, *Science and Technology of the Air Force of the Armed Forces of Ukraine*, KhNUPS, Kharkiv, No. 3(12), pp. 154-158.
3. Kuchuk, G., Nechausov, S. and Kharchenko, V. (2015), “Two-stage optimization of resource allocation for hybrid cloud data store”, *International Conference on Information and Digital Technologies*, Zilina, pp. 266-271.
4. Kovalenko, A. and Kuchuk H. (2018), “Methods for synthesis of informational and technical structures of critical application object’s control system”, *Advanced Information Systems*, 2018, Vol. 2, No. 1, pp. 22–27, DOI: <https://doi.org/10.20998/2522-9052.2018.1.04>.
5. Kuchuk, G.A., Staseva Ya.Yu. and Bolyubash O.O. (2006), “Розрахунок навантаження мультисервісної мережі”, *Системи озброєння і військова техніка*, No. 4(8), pp. 130-134.
6. Kuchuk, G.A. (2013), “Method of Synthesis of Information Structure of Connected Fragment of Corporate Multiservice Network”, *Collection of Scientific Papers of Kharkov University of Air Force*, No. 2(35), pp. 97-102.
7. Shmatkov S.I., Kuchuk, N.G. and Donets V.V. (2018), “Model of information structure of the hyperconvergent system of support of electronic computing resources of university e-learning”, *Control systems, navigation and communication*, PNTU, Poltava, No. 2 (48), pp. 97-100.
8. Donets, V., Kuchuk, N. and Shmatkov, S. (2018), “Development of software of e-learning information system synthesis modeling process”, *Advanced Information Systems*, Vol. 2, No 2, pp. 117–121, DOI: <https://doi.org/10.20998/2522-9052.2018.2.20>.
9. Kuchuk, N., Mozhaiev, O., Mozhaiev, M. and Kuchuk, H. (2017), “Method for calculating of R-learning traffic peakedness”, *4th International Scientific-Practical Conference Problems of Infocommunications Science and Technology, PIC S and T 2017*, pp. 359–362. URL : <http://dx.doi.org/10.1109/INFOCOMMST.2017.8246416>.
10. Merlac, V., Smatkov, S., Kuchuk, N. and Nechausov A. (2018), “Resources Distribution Method of University e-learning on the Hyperconvergent platform”, *Conf. Proc. of 2018 IEEE 9th International Conference on Dependable Systems, Service and Technologies. DESSERT'2018*, Ukraine, Kyiv, May 24-27, pp. 136-140, – DOI: <http://dx.doi.org/10.1109/DESSERT.2018.8409114>
11. Zykov, I.S., Kuchuk, N.H. and Shmatkov S.I. (2018), “Synthesis of architecture of the computer transaction management system e-learning”, *Advanced Information Systems*, Vol. 2, No. 3, pp. 60-66, DOI: <https://doi.org/10.20998/2522-9052.2018.3.10>.
12. Kuchuk, G.A. (2007), "Traffic management of multiservice distributed telecommunication network", *Systems of control, navigation and communication*, No. 2, pp. 18–27.
13. Kuchuk G.A., Mozhaev, O.O. and Vorobyov O.V. (2006), “The method of aggregation of fractal traffic”, *Radio electronic and computer system*, No. 6 (18), pp. 181-188.
14. Kuchuk, G. A. (1998), “Minimize the load on the network holiness channels”, *Inf. Proc. Systems*, No. 1 (5), pp. 149-154.
15. Kovalenko, A.A. and Kuchuk, G.A. (2016), “Optimal traffic control of a multiservice network based on the methods of sequential improvement of solutions”, *Systems of armament and military equipment*, No. 3 (47), pp. 59-63.
16. Amin Salih M. and Potrus, M.Y. (2015), “A Method for Compensation of Tcp Throughput Degrading During Movement Of Mobile Node”, *ZANCO Journal of Pure and Applied Sciences*, Vol. 27, No 6, pp. 59–68.
17. Mozhaev, O., Kuchuk H., Kuchuk, N., Mozhaev, M. and Lohvynenko M. (2017), “Multiservice network security metric”, *IEEE Advanced information and communication technologies-2017*, Proc. of the 2th Int. Conf, Lviv, pp. 133-136, DOI: <https://doi.org/10.1109/AIACT.2017.8020083>.
18. Kovalenko, A.A., Kuchuk, G.A. and Ruban I.V. (2018), “The use of timelines for approximating the length of queues of computer networks”, *The current state of research and technology in industry*, No. 2(4), pp. 12–18. – DOI: <http://doi.org/10.30837/2522-9818.2018.4.012>.
19. Kuchuk G.A., Mohammad A.S. and Kovalenko, A.A. (2011), “The parametric data transmission control method for modifying the transport protocols of wireless networks”, *Information Processing Systems*, No. 8 (98), pp. 211-218.

20. Kuchuk, G.A., Kovalenko, A.A. and Lukova-Chujiko, N.V. (2017), "A method for minimizing the average latency of packets in the virtual connections of the cloud service support network", *Control, navigation and communication systems*, PNTU, Poltava, No. 2 (42), - pp. 117–120.
21. Sviridov, A., Kovalenko, A. and Kuchuk, H. (2018), "The pass-through capacity redevelopment method of net critical section based on improvement ON/OFF models of traffic", *Advanced Information Systems*, Vol. 2, No. 2, pp. 139–144, DOI: <https://doi.org/10.20998/2522-9052.2018.2.24>

Received (Надійшла) 25.11.2019

Accepted for publication (Прийнята до друку) 20.01.2020

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Кучук Ніна Георгіївна – кандидат педагогічних наук, доцент, доцент кафедри обчислювальної техніки та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;

Nina Kuchuk – Candidate of Pedagogical Sciences, Associate Professor, Associate Professor of Computer Science and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;

e-mail: nina_kuchuk@ukr.net; ORCID ID: <http://orcid.org/0000-0002-0784-1465>.

Мерлак Вікторія Юрївна – аспірантка кафедри комп'ютерних систем, мереж та кібербезпеки, Національний аерокосмічний університет імені М. С. Жуковського «ХАІ», Харків, Україна;

Viktoriia Merlak – graduate Student of Computer Systems, Networks and Cyber security Department, National Aviation University "Kharkiv Aviation Institute", Kharkiv, Ukraine;

e-mail: v.merlak@csn.khai.edu; ORCID ID: <http://orcid.org/0000-0003-4096-0797>.

Скороделов Володимир Васильович – кандидат технічних наук, доцент, професор кафедри обчислювальної техніки та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;

Vladimir Skorodelov – Candidate of Technical Sciences, Associate Professor, Professor of Computer Science and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;

e-mail: Skorodelov@ukr.net; ORCID ID: <http://orcid.org/0000-0003-2629-6527>.

Метод снижения времени доступа к слабоструктурированным данным

Н. Г. Кучук, В. Ю. Мерлак, В. В. Скороделов

Аннотация. На сегодня актуальна проблема хранения и обработки слабоструктурированной информации. Суть проблемы заключается в том, что слабоструктурированность и большие объемы данных приводят к нарушению требований QoS, в частности, увеличивается время доступа к данным и соответственно увеличивается время ожидания реакции на запрос. Кроме того, существенно возрастают расходы на хранение данных. **Предметом исследования** являются методы обработки и хранения слабоструктурированных и неструктурированных данных. **Цель исследования** – разработка метода уменьшения времени доступа к неструктурированным данным путем формирования системы иерархических виджетов и планирования доступа к системным ресурсам множественных запросов транзакций отдельных виджетов, которые создают очереди при доступе к нереляционным хранилищам данных. **Результаты.** Проанализированы варианты структуры хранилищ данных для обработки и хранения слабоструктурированных данных. Обосновано применение системы иерархических виджетов при работе с множественными запросами. Предложен метод перераспределения очередей к системным ресурсам. Главным критерием при перераспределении выбрано своевременное и скорейшее завершение всех транзакций системы иерархических виджетов. Разработан метод перепланировки, который обеспечивает уменьшение времени обработки запросов транзакций системы иерархических виджетов. **Выводы.** Показана необходимость динамической перепланировки процесса распределения ресурсов, необходимых для поддержки системы иерархических виджетов. Предложен метод распределения ресурсов между транзакциями системы иерархических виджетов при обработке неструктурированной информации в нереляционных хранилищах. Оптимизация проводилась по критерию скорейшего завершения всех транзакций за счет сокращения времени выполнения длительных запросов.

Ключевые слова: виджет; слабоструктурированная информация; хранилище данных; транзакция.

A method of reducing access time to poorly structured data

N. Kuchuk, V. Merlac, V. Skorodelov

Abstract. Today, the problem of storage and processing of poorly structured information is relevant. The essence of the problem is that poorly structured and large amounts of data lead to violation of QoS requirements. Also, the data access time increases and, accordingly, the waiting time for a response to a request increases. In addition, data storage costs increase significantly. **The subject of the research** is the processing and storage of poorly structured and unstructured data. **The purpose of the study** is to develop a method for reducing access time to unstructured data by creating a hierarchical widget system and planning access to system resources of multiple transaction requests of individual widgets, create queues when accessing non-relational data stores. **Results.** Variants of the structure of data warehouses for processing and storage of poorly structured data are analyzed. The use of a hierarchical widget system when working with multiple queries is justified. A method for redistributing queues to system resources is proposed. The main criterion for redistribution is the timely and speedy completion of all transactions of the hierarchical widget system. The developed redevelopment method, which provides a reduction in the processing time of transaction requests of the hierarchical widget system. **Conclusions.** The need for dynamic redevelopment of the resource allocation process necessary to support the hierarchical widget system is shown. A method for distributing resources between transactions of a hierarchical widget system while processing unstructured information in non-relational databases is proposed. Optimization was carried out according to the criterion of speedy completion of all transactions by reducing the execution time of long requests.

Keywords: widget; poorly structured information; data store; transaction.