

3.12.2020
N 1/6 - 05.102

ВІДГУК

офіційного опонента на дисертаційну роботу
Дворецького Михайла Леонідовича
«Моделі та інформаційна технологія оптимізації структури бази даних вузла у
корпоративних інформаційних системах»,
подану на здобуття наукового ступеня кандидата технічних наук
зі спеціальності 05.13.06 — інформаційні технології

Актуальність теми дисертації. Сучасні розподілені корпоративні інформаційні системи (РКІС) складаються з великої кількості неоднорідних територіально розосереджених апаратних і програмних компонентів, кожен з яких виконує свої функції. Однією з найбільш актуальних задач при побудові таких систем залишається задача забезпечення користувачів оперативним доступом до даних, синхронізація їх окремих частин та зниження навантаження на апаратне забезпечення. Одним із головних факторів, що впливає на надійність і доступність бази даних (БД), є локалізація посилань, яка досягається за рахунок оптимізації структури БД РКІС в цілому та її вузлів зокрема. На сьогоднішній день виконано численні дослідження стосовно оптимізації та вдосконалення РКІС, що базуються на використанні сучасних методів в галузі моделювання та інформаційних технологій, однак проблема побудови РКІС оптимальної структури й досі не отримала остаточного рішення. Таким чином, тема дисертаційного дослідження, що спрямована на підвищення рівня доступності даних та ефективності використання розподілених та територіально розосереджених комп'ютерних систем шляхом створення моделей і методів та реалізації інформаційної технології оптимізації структури БД вузла РКІС є *актуальною*.

Загальна характеристика роботи. Дисертація виконана в Чорноморському Національному університеті ім. Петра Могили і являє собою рукопис, що складається з вступу, п'яти розділів, загальних висновків, списку використаних джерел та додатків.

У вступі обґрунтовано актуальність обраної теми дисертації і її зв'язок із державними науковими програмами й темами. Сформульовано мету й задачі дослідження. Наведено характеристики наукової новизни, теоретичного і практичного значення отриманих результатів. Відзначено особистий внесок здобувача і показано ступінь апробації результатів.

У першому розділі проведено огляд і аналіз сучасного стану вирішення задачі моделювання і побудови РКІС, розглянуто підходи та методи підвищення швидкості та рівня доступності до даних на вузлі. Розглянуто різні концепції представлення даних в розподілених БД, серед яких виділено комбіновану, що поділяє базу даних на логічні фрагменти та дозволяє мати довільну кількість фізичних копій кожного фрагмента. Ключовим фактором, який впливає на надійність і доступність бази даних, є локалізація посилань.

Запропоновано накопичення статистичних даних виконання SQL-запитів РКІС до розподіленої БД, що дає змогу врахувати всі особливості РКІС. Парсинг текстів SQL-запитів дозволить виявити множини відношень, атрибутів

та кортежів БД, задіяних у кожному запиті. Вводиться набір критеріїв оптимальності структури БД та формулюється задача багатокритеріальної оптимізації.

З метою оптимізації структури БД вузла сформульовано ряд завдань для досягнення мети роботи, а саме: накопичення статистичних даних; розробка інформаційної технології парсингу тексту SQL-запитів; розробка реляційної та багатовимірної БД для зберігання та аналізу даних SQL-запитів; створення математичних моделей залежності критеріїв оптимальності структури БД вузла РКІС від граничного рівня маркеру представленості даних вузла; розв'язання задачі вибору найкращої альтернативи; класифікації нових записів відношень БД.

Проведено аналіз існуючих інструментальних засобів, моделей та технологій вирішення задачі побудови ефективної розподіленої інформаційної системи. Сформульовано постановку задачі дослідження.

Другий розділ присвячено моделі SQL-запиту, що включає дані про застосунок, хост, користувача, а також набір відношень, атрибутів та кортежів БД, які входять до результуючого набору. Запит представляється у багатовимірній БД у базисі вимірів, що включає «відношення», «атрибут» та «кортеж» БД.

При виконанні аналізу щодо представленості даних для елементів вимірів вводиться характеристика «маркер представленості даних», що відображає рівень необхідності представлення даних на вузлі РКІС і задається лінгвістичною змінною. Розроблено функцію агрегації для маркера представленості даних, що використано у процесі прийнятті рішення по представленості даних на віддаленому вузлі при консолідації рядків таблиці фактів та розрахунку значення маркеру для кожного її елементу. Визначено наступні критерії оптимальності: доступність та швидкість отримання даних, ступінь залежності від центрального вузла БД, розмір БД, ступінь достовірності даних, необхідність у подальшій синхронізації.

При синхронізації даних вузла РКІС запропоновано використання динамічного визначення періоду обміну даними. Наявність механізмів обчислення моменту наступної синхронізації даних дозволяє підвищити актуальність даних, та розвантажити ресурси серверів баз даних. При їх реалізації враховані такі фактори, як кількість змін у журналі транзакцій; якість цих змін (на основі частоти використання даних у користувацьких запитах визначаються коефіцієнти впливу на роботу вузла РКІС); значення точки актуальності даних; завантаженість серверу БД поточними запитами.

У третьому розділі обрано та удосконалено метод аналізу ієрархій (МАІ) для вибору найкращої альтернативи з множини граничних рівнів маркеру представленості даних. Використано автоматичну ініціалізацію матриці попарних порівнянь альтернатив з урахуванням отриманих математичних моделей та нормалізації значень; елементи нечіткого логічного виведення для дефазифікації вектору глобальних пріоритетів; класифікацію нових даних із використанням наївного алгоритму Байєса. З метою підвищення точності результату, запропоновано представити вектор глобальних пріоритетів

альтернатив у вигляді вектору нечітких чисел для маркеру представленості даних.

Четвертий розділ присвячено створенню технології обліку та аналізу SQL-запитів, як складової інформаційної технології оптимізації структури БД вузла РКІС. При цьому автором роботи запропоновано та реалізовано підсистему парсингу тексту SQL-запиту з метою виділення множини відношень, атрибутів та вкладених підзапитів.

Завдання визначення множин кортежів запиту вирішено за рахунок виконання додаткового запиту для кожної таблиці із колекції відношень базового SQL-запиту, список атрибутів у якому замінюється на первинний ключ відношення. Результат запиту запам'ятовується як множина кортежів таблиці, що використовується поточним запитом. Для обробки статистичних даних, що отримуються шляхом парсингу, та, враховуючи множинність вимірів, дані SQL-запитів представлено у вигляді багатовимірної бази даних (ББД), що дозволяє використовувати методи OLAP для визначення оптимального рівня маркеру представленості даних згідно до встановлених ЛПР зрізів даних. Для реалізації таблиці фактів та таблиць вимірів у вигляді схеми «зірка», на рівні реляційної БД ІС обліку користувацьких запитів уведено ряд представлень. Даний підхід підвищує прозорість структури БД та є своєрідним інтерфейсом при взаємодії багатовимірної моделі із реляційною.

У п'ятому розділі представлено програмну складову інформаційної технології визначення оптимального рівня маркеру представленості даних на вузлі РКІС, що виконана у вигляді сервіс-орієнтованого вебзастосунку на основі трьохланкової архітектури.

При розрахунку значення ефективності використання вузла РКІС, було виведено співвідношення отриманого результату до витрачених ресурсів на базі раніше сформульованих критеріїв оптимальності структури БД, та отриманого у ході вирішення задачі вибору кращої альтернативи вектору відносної ваги критеріїв оптимальності. Отримані результати дозволяють зробити висновок про підвищення ефективності у межах від 16 % до 30 % у порівнянні з розміщенням на вузлі РКІС всіх необхідних даних або лише критичних даних відповідно.

У висновках сформульовано основні наукові та практичні результати дисертаційної роботи.

У додатках наведено: акти впровадження результатів дисертаційної роботи у ППФ «Юнікс Трейд Ко», ТОВ «Еліт-Білдінг», а також у навчальному процесі та науково-дослідницькій роботі Чорноморського національного університету імені Петра Могили; список публікацій за темою дисертації, лістинг скрипту створення структури БД обліку користувацьких запитів та граматики для опису мови T-SQL у вигляді форм Бекуса-Наура (БНФ).

Наукова новизна одержаних результатів полягає в тому, що:

1. Вперше введено поняття маркеру представленості даних на вузлі РКІС для елементів вимірів моделі SQL-запиту та розроблена функція агрегації, що дозволяє визначити рівень необхідності атрибутів та кортежів відношення БД

на вузлі PKIC на основі статистики SQL-запитів.

2. Вперше побудовано модель залежності критеріїв оптимальності структури БД вузла PKIC від значення маркера представленості даних, що на відміну від існуючих дозволяє визначити оптимальне граничне значення маркера на основі статистики SQL-запитів.

3. Вперше розроблено інформаційну технологію розрахунку критеріїв оптимальності структури БД вузла PKIC, що дозволяє визначити граничний рівень маркера представленості даних і прийняти рішення про представленість даних на вузлі PKIC та, на відміну від існуючих, використовує аналітичні дані та результати парсингу тексту SQL-запитів до серверів PKIC.

4. Удосконалено модель обліку SQL-запиту, яка, на відміну від існуючих, дозволяє проводити оперативно-аналітичний аналіз, зважаючи на множинність вимірів з метою розрахунку рівня представленості атрибутів та кортежів відношення БД на вузлі PKIC.

5. Отримав подальшого розвитку метод аналізу ієрархій за рахунок автоматичної ініціалізації матриці попарних порівнянь альтернатив згідно отриманих математичних моделей та нормалізації значень та представлення результату у вигляді вектору нечітких чисел із приведенням до чіткого значення, що дозволило підвищити ефективність вузла PKIC на 4%.

6. Отримав подальшого розвитку метод асинхронної реплікації даних за рахунок динамічного визначення моменту наступної синхронізації даних на основі статистики SQL-запитів, що дозволяє підвищити актуальність даних та розвантажити ресурси серверів баз даних.

Практичне значення отриманих результатів.

1. Отримано закономірності між значеннями критеріїв оптимальності структури БД та рівнем представленості даних на базі статистики SQL-запитів.

2. Використання модифікованого методу ієрархій дозволяє отримувати оптимальне рішення структури бази віддаленого вузла PKIC.

3. Результати експериментів, а також інформаційна технологія, побудована на базі моделі обліку SQL-запитів і нового підходу до оптимізації структури бази даних вузла PKIC, впроваджено на двох підприємствах із територіально-розподіленою структурою та мало позитивний результат щодо підвищення ефективності використання БД, що зафіксовано у відповідних актах впровадження. Так, на ППФ «Юнікс Трейд Ко» підвищення ефективності використання БД вузла PKIC склало 12%. У ТОВ «Еліт білдінг» в результаті впровадження запропонованої інформаційної технології підвищено швидкість виконання запитів на 14 %, а рівень доступності даних - на 21 %.

4. Результати дисертаційної роботи використано при виконанні наступних держбюджетних НДР:

- «Інтеграція інформаційних технологій баз даних, баз знань та Data Mining» (номер держ. реєстрації № 0112U007749);
- «Інтелектуалізація інформаційних систем та СППР за рахунок впровадження методів ситуаційного моделювання, сценарного аналізу, Data Mining та OLAP технологій» (номер держ. реєстрації

№ 0114U004402);

- «Розроблення найсучаснішого інтерактивного навчально-тренажерного та аналітично-консультативного комплексу військово-цивільного призначення» (номер держ. реєстрації № 0118U000193).

5. Наукові результати, отримані в дисертаційній роботі, використані в курсах лекцій, лабораторних роботах та в практичних заняттях по дисциплінах «Інформаційні технології, OLTP, OLAP та DM на клієнт-серверній платформі», «Клієнт-серверні СКБД та аналітичні системи», «Data base development», а також у курсовому та дипломному проектуванні для студентів спеціальності 121 «Інженерія програмного забезпечення» на кафедрі інженерії програмного забезпечення факультету комп'ютерних наук Чорноморського національного університету ім. Петра Могили.

Ступінь обґрунтованості та достовірності наукових положень, висновків і рекомендацій. Достовірність наукових положень, висновків і рекомендацій підтверджується коректним застосуванням математичного апарату, результатами комп'ютерного моделювання, системою методик і перетворень, що не містить принципових помилок, та дозволяють визначити рівень представленості даних на вузлі РКІС на основі критеріїв оптимальності структури БД.

Рецензована дисертаційна робота характеризується комплексним підходом до вирішення актуальних питань, у процесі реалізації якої здобувач вміло використовує сучасні методи, прийоми та інструменти наукових досліджень.

Обґрунтованість і достовірність отриманих результатів підтверджується позитивними апробаціями на міжнародних та всеукраїнських науково-практичних конференціях, публікаціями у фахових наукових виданнях та впровадженням у виробництво.

Все це дає підстави стверджувати, що висновки, пропозиції та рекомендації, сформульовані в дисертації, є достатньо обґрунтованими і достовірними.

Повнота викладення наукових результатів дисертації в опублікованих працях. Достатня кількість публікацій здобувача, що пройшли апробацію на міжнародних науково-технічних та науково-практичних конференціях, відображає основні наукові положення, рекомендації та висновки, що отримані в результаті дисертаційних досліджень.

За темою дисертації опубліковано 28 наукових робіт, з них: 17 – у фахових виданнях; 5 публікацій, що реферуються у Scopus; 6 – у збірниках матеріалів міжнародних та всеукраїнських науково-технічних конференцій.

Відповідність змісту автореферату та основних положень дисертації. Усі положення та висновки дисертаційної роботи повною мірою стисло відображені у авторефераті. Представлений автореферат відповідає вимогам нормативних документів МОН України до кандидатських дисертацій.

Автореферат дисертації є цілком ідентичним основним положенням дисертаційної роботи та повністю передає її зміст. Матеріали автореферату викладено чітко та послідовно.

Відповідність спеціальності. Дисертаційна робота Дворецького М.Л. відповідає паспорту спеціальності 05.13.06 – інформаційні технології за пунктом 3 напрямків досліджень: «дослідження та побудова інформаційних технологій для розроблення і впровадження баз і сховищ даних систем комп'ютерної підтримки рішень в автоматизованих системах і мережах; створення інформаційних технологій для системного аналізу дослідження, розроблення архітектури та методів побудови багаторівневих, територіально розосереджених комп'ютерних систем і мереж із розподіленими базами даних та знань зокрема комерційного призначення; дослідження, розроблення і впровадження Інтернет-технологій для побудови сервіс-орієнтованих систем, а також для організації та реалізації систем розподіленої обробки інформації».

Зауваження щодо змісту дисертаційної роботи та автореферату

1. У розділі 1 при аналізі підходів щодо підвищення швидкості роботи інформаційних систем на основі БД не наведено заходи, пов'язані із оптимізацією налаштувань СКБД та аналізом необхідної потужності апаратного забезпечення.

2. У підрозділі 2.3 при формалізації критерію необхідності синхронізації даних не розглядається питання якості каналів зв'язку між центральним та віддаленим вузлами РКІС.

3. У підрозділі 3.2 при введенні обмежень області значень критеріїв оптимальності не враховано наявність значень, що надходять від різних експертів.

4. У підрозділі 3.4 не наведено, яка кількість нових кортежів БД має вести до повторного розрахунку оптимального рівня маркеру представленості даних.

5. У підрозділі 4.5 при аналізі статистики SQL-запитів, автор не використовує переваг методів інтелектуального аналізу даних, зокрема пошуку асоціативних правил при зверненні до відношень БД у межах одного запиту.

6. У підрозділі 5.1 не аргументовано вибір стеку технологій на користь наведених фреймворків розробки фронтенд- та бекенд-складової, а також системи керування базами даних.

7. У розділі 5, таблиці, що містять опис маршрутів бекенд-складової інформаційної технології, доцільно було б представити у окремому додатку до роботи.

Однак, зазначені зауваження не носять принципового характеру і не знижують цінності проведеного здобувачем дослідження, актуальності, новизни та практичної значущості дисертаційної роботи.

Загальний висновок по дисертаційній роботі щодо її відповідності встановленим вимогам. Дисертаційну роботу Дворецького М. Л. «Моделі та інформаційна технологія оптимізації структури бази даних вузла у корпоративних інформаційних системах» можна визнати завершеним науковим дослідженням, результати якого дозволяють вирішити актуальне науково-практичне завдання щодо підвищення рівня доступності даних та ефективності використання розподілених та територіально розосереджених комп'ютерних систем на основі статистики SQL-запитів.

Дисертація Дворецького М. Л. за змістом та якістю теоретичних і методичних розробок відповідає рівню дисертацій на здобуття наукового ступеню кандидата технічних наук. Вибрану тему дисертаційної роботи необхідним чином розкрито; поставлену мету досягнуто; завдання дисертаційної роботи в цілому виконано.

Дисертація Дворецького М. Л. на тему «Моделі та інформаційна технологія оптимізації структури бази даних вузла у корпоративних інформаційних системах» відповідає спеціальності 05.13.06 – «інформаційні технології», а також нормативним вимогам до кандидатських дисертацій технічного профілю.

Автореферат дисертації необхідним чином розкриває основні її положення та висновки, є ідентичним дисертації за структурою та змістом та не містить інформації, що є відсутньою у дисертаційній роботі.

Вважаю, що за актуальністю обраної теми, обсягом та рівнем виконаних досліджень, достовірністю та обґрунтованістю висновків, фаховим спрямуванням, теоретичною та практичною значимістю, новизною досліджень, значенням для науки і практики дисертаційна робота «Моделі та інформаційна технологія оптимізації структури бази даних вузла у корпоративних інформаційних системах» відповідає вимогам п.п.9, 11 та 12 «Порядку присудження наукових ступенів та присвоєння вченого звання старшого наукового співробітника», а її автор, Дворецький Михайло Леонідович, заслуговує на присудження наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – «Інформаційні технології».

Офіційний опонент, професор кафедри
комп'ютерних наук Вінницького
національного технічного університету
кандидат технічних наук, професор



Т. О. Савчук