

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧОРНОМОРСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ПЕТРА МОГИЛИ

Дворецький Михайло Леонідович



УДК 004.658.3

**МОДЕЛІ ТА ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ
ОПТИМІЗАЦІЇ СТРУКТУРИ БАЗИ ДАНИХ ВУЗЛА
У КОРПОРАТИВНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ**

05.13.06 – інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Миколаїв – 2020

Дисертацією є рукопис.

Робота виконана в Чорноморському національному університеті імені Петра Могили Міністерства освіти і науки України.

Науковий керівник:

доктор технічних наук, професор
Фісун Микола Тихонович,
Чорноморський національний університет
імені Петра Могили, м. Миколаїв,
професор кафедри інженерії
програмного забезпечення.

Офіційні опоненти:

доктор технічних наук, професор
Філатов Валентин Олександрович,
Харківський національний
університет радіоелектроніки,
завідувач кафедри штучного інтелекту;

кандидат технічних наук, професор
Савчук Тамара Олександрівна,
Вінницький національний
технічний університет,
професор кафедри комп'ютерних наук.

Захист відбудеться **22 грудня 2020 р. о 14⁰⁰** годині на засіданні спеціалізованої вченої ради Д 38.053.05 у Чорноморському національному університеті імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68 Десантників, 10.

З дисертацією можна ознайомитися в бібліотеці Чорноморського національного університету імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68 Десантників, 10 та за електронною адресою: <https://chmnu.edu.ua/disertatsiyi/>.

Автореферат розісланий 20 листопада 2020 року.

Вчений секретар
спеціалізованої вченої ради



Євген ДАВИДЕНКО

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність роботи. Широке використання інформаційних технологій багаторівневих, територіально розосереджених та розподілених корпоративних інформаційних систем (РКІС) на основі баз і сховищ даних, у тому числі розподілених, обумовлює необхідність забезпечення певної автономності роботи користувачів, зниження навантаження на сервери баз даних (БД) та сервіси синхронізації даних шляхом оптимізації структури БД віддаленого вузла корпоративної інформаційної системи (КІС).

Ключовим фактором, що впливає на надійність і доступність БД, є локалізація посилань. Про високий ступінь локалізації посилань свідчить представлення на поточному вузлі тих даних, що викликаються винятково його користувачем. Комбінована стратегія розподілу даних є найбільш привабливою, але за її використання, окрім задачі синхронізації дубльованої інформації, актуальною постає задача проектування оптимальної структури БД з точки зору приналежності даних до вузла РКІС.

Відношення БД представляються на вузлі РКІС після застосування операцій проєкції та вибірки. Тобто для оптимального представлення даних необхідно використати елементи вертикальної та горизонтальної фрагментації даних. Імітаційні моделі РКІС зазвичай дуже спрощені та не враховують ряд важливих факторів, що впливають на її поведінку. У свою чергу накопичення статистичних даних виконання SQL-запитів до розподіленої БД на реально працюючих серверах дає змогу врахувати всі особливості РКІС.

Наведені питання розглядалися у роботах таких вітчизняних вчених, як: Стогній А. О., Пасічник В. В., Шаховська Н. Б., Кунгурцев О. Б., Філатов В. О., Малахов Є. В., Скобцов Ю. А., Зіноватна С. Л., Новохатська С. Ю. та ін., а також розглянуті такими зарубіжними вченими, як Дейт К. Дж., Кодд Е. Ф., Гарсія-Моліна Г., Барсегян А. А., Ульман Дж. Д., Райордан Р. М., Бергер А. Б., Конноллі Т., Петкович Д. та ін. Однак, не дивлячись на певний прогрес у методах та інформаційних технологіях проектування структур розподілених баз даних, питання оптимізації структури БД вузлів РКІС шляхом нових підходів до моделювання та статистики SQL-запитів обумовлює *актуальність* теми цього дослідження.

Зв'язок роботи з науковими програмами, планами, темами. Дисертація виконана відповідно до завдань науково-дослідницьких робіт Чорноморського національного університету (надалі – ЧНУ) імені Петра Могили, в яких автор брав участь як виконавець:

- «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).

Метою дослідження є підвищення рівня доступності даних та ефективності використання розподілених та територіально розосереджених комп'ютерних систем шляхом реалізації інформаційної технології оптимізації структури БД вузла РКІС на основі статистики SQL-запитів.

Для досягнення поставленої мети в роботі сформульовано такі завдання:

- проаналізувати особливості задач використання інтегрованого комплексу територіально розосереджених інформаційних систем та методів їх оптимізації;
- побудувати модель SQL-запиту та формалізувати критерії оптимальності структури БД віддаленого вузла КІС;
- розробити інформаційну технологію парсингу та подальшого обліку SQL-запитів до реляційної БД, а також обчислення значень критеріїв оптимальності відповідно до результатів операцій зрізу та консолідації;
- сформулювати та розв'язати задачу багатокритеріальної оптимізації для визначення оптимального рівня представленості даних на вузлі РКІС;
- розробити інформаційну технологію обчислення вектору глобальних пріоритетів альтернатив із представленням результату у вигляді вектору нечітких чисел та приведенням до чіткого значення;
- виконати класифікацію нових даних БД відповідно до необхідності їх представлення на вузлі РКІС.

Об'єкт дослідження. Процес оптимізації структури БД розподілених корпоративних інформаційних систем, в основі яких лежать реляційні бази даних.

Предмет дослідження. Моделі SQL-запитів реляційних БД та методи вибору кращої альтернативи структури вузла РКІС.

Методи дослідження. Методи досліджень базуються на основних засадах теорії реляційних баз даних (у тому числі розподілених), методах парсингу тексту, багатокритеріальної оптимізації та шаблонного проєктування. Зокрема, на: методах представлення даних та обробки транзакцій у розподілених БД – при аналізі особливостей використання РКІС; методах реляційної алгебри, теорії реляційної моделі даних, багатовимірній моделі даних – у процесі побудови моделі SQL-запиту; методі аналізу ієрархій та апараті нечіткого логічного висновку – при вирішенні задачі вибору найкращої альтернативи рівня представленості даних; методі наївного Байєса – у процесі класифікації нових даних що надходять на вузол РКІС після останньої реструктуризації; методах парсингу тексту на основі формальних граматик – у процесі розв'язання задачі визначення складових SQL-запиту; методах шаблонного проєктування та підходах щодо розробки сучасних вебзастосунків при реалізації сервіс-орієнтованої інтернет-технології підтримки прийняття рішень при проєктуванні структури віддаленого вузла РКІС.

Наукова новизна отриманих результатів:

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 РКІС на основі статистики SQL-запитів.

2. *Вперше* побудовано модель залежності критеріїв оптимальності структури БД вузла РКІС від значення маркера представленості даних, що на відміну від

існуючих підходів дозволяє визначити оптимальне граничне значення маркера на основі статистики SQL-запитів.

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

4. Удосконалено модель SQL-запиту, яка, на відміну від існуючих, представлена моделлю багатовимірної БД в аналітиці множини атрибутів та кортежів відношення, а також інших характеристик запиту, що дозволяє провести оперативно-аналітичний аналіз, зважаючи на множинність вимірів з метою розрахунку рівня представленості атрибутів та кортежів відношення БД на вузлі РКІС.

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 РКІС 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. Практична значимість одержаних результатів полягає в розробленні інформаційної технології підтримки прийняття рішення при визначенні представленості даних на вузлі РКІС, що дозволяє оптимізувати структуру БД вузла.

Результати дослідження впроваджено на ППФ «Юнікс Трейд Ко» при проєктуванні структури БД віддаленого вузла РКІС, що дозволило виявити закономірності між значеннями критеріїв оптимальності структури БД та рівнем представленості даних інформаційної системи на базі статистики SQL-запитів. Ефект від впровадження зазначених результатів полягає у підвищенні ефективності використання БД вузла РКІС на 12 %.

Також розроблені моделі та інформаційна технологія впроваджені на «ТОВ Еліт Білдінг» для виявлення залежності критеріїв оптимальності структури БД та рівня представленості даних. Ефект від впровадження полягає у підвищенні швидкості виконання запитів на 14 % та доступності даних на 21 %.

Результати роботи також впроваджені в навчальний процес на кафедрі інженерії програмного забезпечення ЧНУ імені Петра Могили під час викладання дисциплін «Інформаційні технології OLTP, OLAP та DM на клієнт-серверній платформі», «Клієнт-серверні СКБД та аналітичні системи», «Database development».

Особистий внесок здобувача. Наукові положення, теоретичні розробки та практичні результати у представленій дисертаційній роботі отримані автором особисто. У наукових роботах, опублікованих у співавторстві, здобувачу належить таке. У [1] виконано практичну реалізацію задачі пошуку асоціативних правил засобами мови SQL та аналіз отриманих результатів. У [2] засобами SQL Server на

базі OLAP розроблено інформаційно-аналітичну систему для аналізу трендів продажів торговельного підприємства. У рамках [3] запропоновано та виконано апробацію алгоритму розрахунку моменту наступної синхронізації БД різнорідних інформаційних систем (ІС). У [13] запропоноване ієрархічне представлення даних при виконанні ABC-XYZ аналізу та реалізовані механізми подальшого поєднання результатів аналізу за різними критеріями. У ході [14] здобувачем виконано аналіз наявності інструментальних засобів та деяких показників продуктивності щодо автоматизації OLAP на базі СКБД Oracle та SQL Server. У [15] виконано розробку парсеру та багатовимірної моделі SQL-запиту, на базі якої розроблено СППР при оптимізації структури РБД. У рамках [16] здобувачу належить реалізація підсистем накопичення та аналізу статистики користувацьких запитів. У [17] розроблено технологію парсингу тексту запиту для оптимізації структури БД. У роботі [18] та [23] автору належить дослідження використання зворотного зв'язку користувацької активності в системах управління знаннями та реалізація відповідної технології. У рамках [20] виконано розробку системи управління знаннями на базі CMS Wordpress. У роботі [21] проаналізовано результати пошуку асоціацій на базі багатовимірної бази даних. У [22] реалізовано взаємодію бекенд та фронтенд фреймворків при розробці сервіс-орієнтованого вебсервісу. У [24] автор виконав порівняльний аналіз різних типів автоматизації облікових інформаційних систем та реалізацію відповідного вебзастосунку. У [25] здобувачу належить реалізація підсистеми визначення корисності кортежу на вузлі РКІС із використанням нейронних мереж. У рамках [26], [27] та [28] автором запропоновано реалізацію моделей та інформаційної технології оптимізації структури БД вузла РКІС.

Апробація результатів дисертації. Основні положення і наукові результати дисертаційної роботи були представлені і отримали позитивну оцінку на 18 міжнародних, всеукраїнських та регіональних наукових конференціях, 4 з яких входять до наукометричної бази SCOPUS, а саме: Всеукраїнська науково-методична конференція «Могилянські читання» (2014–2019); Міжнародна науково-технічна конференція «Комп'ютерні науки: освіта, наука, практика» (2014); Міжнародна науково-практична конференція «Молодь у світі сучасних технологій» (2016); Міжнародна науково-практична конференція «Інформаційні технології та взаємодії» (2017); Міжнародна науково-практична конференція Ольвійський форум (2016–2019); 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems (2017); 12th & 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (2017–2018); 1st International Workshop on Information-Communication Technologies & Embedded Systems (2019).

Публікації. За темою дисертації було 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.

Структура й обсяг дисертації. Дисертація складається зі вступу, п'яти розділів, висновків, списку використаних джерел із 163 найменувань та чотирьох додатків. Загальний обсяг дисертації – 213 сторінок, з них – 4 додатки на 36 сторінках, 25 таблиць, 71 рисунок.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі наведено загальну характеристику роботи, обґрунтовано її актуальність, сформульовано мету та основні задачі дослідження, визначено методи вирішення поставлених задач, сформульовано наукову новизну роботи та практичну цінність отриманих результатів.

У першому розділі розглянуто підходи та методи підвищення швидкості роботи інформаційних систем (ІС), серед яких оптимізація запитів за рахунок реалізації загальних підходів модифікації тексту запитів, індексація даних, модифікація структури БД, контрольована денормалізація відношень та використання матеріалізованих представлень. Загальним недоліком проаналізованих підходів визначено необхідність модифікації самої ІС, що у випадку закритої реалізації останніх робить неможливим їх використання. Також розглянуто підходи, пов'язані із використанням концепції сховищ даних (СД), що полягає в розділенні БД для Online Transaction Processing (OLTP) систем і БД для виконання аналізу. Наведено ряд недоліків використання фізичних СД, а також їх віртуальних аналогів.

Розглядаючи проблеми використання «універсальних» або «комплексних» ІС, відмічається тенденція використання набору окремих спеціалізованих рішень із подальшою синхронізацією даних. Холдингова структура об'єкта автоматизації також може обумовити необхідність створення та використання БД розподіленої структури. Під корпоративною інформаційною системою розуміємо сукупність ІС окремих підрозділів підприємства, об'єднаних загальним документообігом, таких, що кожна система виконує частину задач з управління прийняттям рішення, а всі системи разом забезпечують функціонування підприємства.

Розглянуто різні концепції представлення даних у 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. Враховуючи закритість коду SQL-запитів РКІС, задача оптимального проєктування структури БД зводиться до пошуку оптимального розподілу даних по вузлах РКІС. Імітаційні моделі РКІС, що є вкрай складною системою, зазвичай дуже спрощені та не враховують ряд важливих факторів, що впливають на її поведінку.

Натомість запропоновано накопичення статистичних даних виконання SQL-запитів РКІС до розподіленої БД на реально працюючих серверах, що дає змогу врахувати всі особливості РКІС. Парсинг текстів SQL-запитів та розщеплення їх на окремі підзапити дозволить виявити множини відношень, атрибутів та кортежів БД, задіяних у кожному запиті. Подальший аналіз накопиченої та обробленої статистики в аналітиці доступних вимірів у вигляді програмного застосунка, хоста, користувача та ін. дозволить виявити частоту та характер звернень до окремих відношень, їх атрибутів і кортежів. Вводиться набір критеріїв оптимальності структури БД та формулюється задача багатокритеріальної оптимізації. Проаналізовано різні методи її розв'язання, серед яких виділяється метод аналізу ієрархій.

Зроблено висновок про доцільність досліджень, спрямованих на підвищення рівня доступності даних та ефективності використання розподілених та територіально розосереджених комп'ютерних систем шляхом реалізації інформаційної технології

оптимізації структури БД вузла РКІС на базі статистики SQL-запитів. Для виконання оптимізації структури БД сформульовано ряд завдань, 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 маркету представленості даних вузла; розв'язання задачі вибору найкращої альтернативи; класифікації нових записів відношень БД.

У другому розділі на основі реляційної моделі даних та операцій реляційної алгебри реалізовано модель SQL-запиту, що включає дані про застосунок, хост, користувача, а також набір відношень, атрибутів та кортежів БД, які входять до результуючого набору рядків (кортежів). Запит представляється у багатовимірній БД у такому базисі вимірів: $D = \{ДатаЧас, ТипАРМ, РозташуванняАРМ, РольКористувача, Застосунок, R, A, tup\}$.

У процесі аналізу щодо представленості даних для елементів вимірів вводиться характеристика «маркер представленості даних», що відображає рівень необхідності представлення даних на вузлі РКІС. Для кожного елементу значення маркера приймається одне з множини лінгвістичних значень {«необхідно», «бажано», «не потрібно»}, що визначає ступінь необхідності представлення даних того чи іншого робочого місця, ролі користувача або застосунка.

Визначення значення маркера при виконанні консолідації рядків таблиці фактів за значеннями $\langle R, A, tup \rangle$ виконується із використанням методу ковзного середнього із введенням вагових коефіцієнтів елементів вимірів. Тобто кожен атрибут виміру має значення, маркер представленості та ваговий коефіцієнт $A_{dim} = \{Val, Mrk, vol\}$, де $Mrk = \{\text{«обов'язково»}, \text{«необхідно»}, \text{«бажано»}, \text{«не потрібно»}, \text{«заборонено»}\}$, а vol – ваговий коефіцієнт.

Виконавши переведення нечислової лінгвістичної змінної маркерів у числове значення («обов'язково» – «2», «необхідно» – «1», «бажано» – «0», «не потрібно» – «-1», «заборонено» – «-2»), визначаємо для них функцію агрегації:

$$Aggregate_{i=1}^n Mrk_i = \begin{cases} 2, & \text{якщо } \exists Mrk_i = 2 \\ -2, & \text{якщо } \exists Mrk_i = -2 \wedge \nexists Mrk_i = 2 \\ \sum_{i=1}^n (Mrk_i \times \frac{Vol_i}{\sum_{i=1}^n Vol_i}) & \end{cases} \quad (1)$$

При прийнятті рішення з представленості даних на віддаленому вузлі виконується консолідація рядків таблиці фактів за $\langle R, A, tup \rangle$ та розраховується значення маркера для кожного її елемента за формулою (1), після чого приймається рішення про представленість за наступним критерієм:

$$Предст(Node, R, A, tup) = (Aggregate(R, A, tup)_{i=1}^n Mrk_i > koeff_{предст}^{node}) \quad (2)$$

Постає завдання формалізації критеріїв оптимальності граничного рівня маркера представленості даних $koeff_{предст}^{node}$. Виділено такі критерії: «незалежність БД», «розмір БД» та «необхідність синхронізації».

Для формалізації критерію «незалежність БД» вводимо функцію доступності даних для запиту Q :

$$F_{\text{доступність}}(\text{Node}, Q) = \begin{cases} 1, \text{ якщо } \forall R'' \in R_{\text{schema}}^{\text{remote}}, R'' \in R_{\text{schema}}^{\text{remote}} \wedge \\ \quad \forall Q^{\text{inner}} F_{\text{доступність}}(Q^{\text{inner}}) = 1 \\ 0, \text{ якщо } \exists R'' \notin R_{\text{schema}}^{\text{remote}}, R'' \in R_{\text{schema}}^{\text{remote}} \vee \\ \quad \exists Q^{\text{inner}} F_{\text{доступність}}(Q^{\text{inner}}) = 0 \end{cases} \quad (3)$$

Загальне значення визначається як середнє значення підмножини користувацьких запитів.

$$F_{\text{доступність}} = \frac{\sum_{i=1}^n F_{\text{доступність}}(Q_n)}{n}, \text{ де } Q_n \in Q_{\text{node}} \quad (4)$$

До множини Q_n включаються запити Q , для яких $F_{\text{приналежності}}(\text{Node}, Q) = 1$, $Q_n = \{Q \mid F_{\text{приналежності}}(\text{Node}, Q) = 1\}$, де

$$F_{\text{приналежності}}(\text{Node}, Q) = \begin{cases} 1, \text{ якщо } (\exists R'' \in R_{\text{set}}^{\prime\prime} \rightarrow \text{Aggregate}(R'')_{i=1}^n \text{Mrk}_i > -1) \vee \\ \quad (\exists Q^{\text{inner}} \in Q_{\text{set}}^{\text{inner}} \rightarrow F_{\text{приналежності}}(\text{Node}, Q^{\text{inner}}) = 1) \\ 0, \text{ якщо } (\forall R'' \in R_{\text{set}}^{\prime\prime} \rightarrow \text{Aggregate}(R'')_{i=1}^n \text{Mrk}_i \leq -1) \wedge \\ \quad (\forall Q^{\text{inner}} \in Q_{\text{set}}^{\text{inner}} \rightarrow F_{\text{приналежності}}(\text{Node}, Q^{\text{inner}}) = 0) \end{cases}$$

При формалізації критерію «розмір БД» визначається розмір кортежу, як $\text{Size}_R = \text{SizeR0}_i^{\text{DBMS}} + p \times \sum_{i=1}^n \text{Size}(\text{Type}_i)$. Також для кожного відношення визначається поправочний коефіцієнт $\text{coef}_{\text{sizeR}} = \frac{\text{SizeR}_i^{\text{DBMS}} - \text{SizeR0}_i^{\text{DBMS}}}{p \times \sum_{i=1}^n \text{Size}(\text{Type}_i)}$, використовуючи який, визначається розмір відношення R'' на вузлі РКІС як

$$\text{Size}_{R''} = \text{Coef}_{\text{sizeR}} \times (\text{SizeR0}_i^{\text{DBMS}} + p' \times \sum_{i=1}^{n'} \text{Size}(\text{Type}_i)). \quad (5)$$

Сума розмірів (5) всіх підмножин R'' визначає розмір БД віддаленого вузла. У подальшому значення критерію «розмір БД» вузла представлено у вигляді відношення до загального розміру БД центрального вузла КІС.

$$F_{\text{size}} = \frac{\sum_{i=1}^n \text{Size}_{R''_i}}{\sum_{i=1}^n \text{SizeR}_i^{\text{DBMS}}} \quad (6)$$

У процесі визначення критерію «необхідність синхронізації» вводиться модель SQL-запиту модифікації даних відношення $Q^{\text{modif}} = \{\text{Виміри}, R''^{\text{modif}}, \text{type}\}$, де R''^{modif} – підмножина на відношенні R , що зазнає змін внаслідок виконання операцій модифікації даних, $\text{type} = \{\text{insert}, \text{update}, \text{delete}\}$ – тип операції. Множина $R''^{\text{modif}}_{\text{node}}$

базового відношення R може бути визначена, як об'єднання підмножин R''^{modif} усіх запитів, що надійшли до БД з віддаленого вузла $R'''^{modif}_{node} = \bigcup_{i=1}^n R_i''^{modif}$.

Аналогічним чином визначається множина R'''^{modif}_{main} , що зазнає модифікацій на центральному вузлі або інших вузлах із подальшою синхронізацією із центральним. Перетин множин R'''^{modif}_{node} та R'''^{modif}_{main} визначає підмножину базового відношення, на якій можуть виникати конфлікти зміни, та значення атрибутів якої потребують застосування більш ресурсоємних алгоритмів синхронізації.

$$R'''^{modif}_{node} = R'''^{modif}_{node} \cap R'''^{modif}_{main} \quad (7)$$

Відповідно до приналежності даних до R'''^{modif}_{node} , виконується підрахунок кількості запитів віддаленого вузла та виводиться співвідношення до загальної кількості запитів, що і буде значенням критерію «необхідність синхронізації»:

$$F_{synchro} = \frac{p_{node}^{modif}}{p_{node}} \quad (8)$$

У процесі синхронізації даних віддаленого вузла період обміну не є чітко визначеним. Наявність механізмів обчислення моменту наступної синхронізації даних дозволяє підвищити актуальність даних та розвантажити ресурси серверів баз даних. За їх реалізації враховані такі фактори: 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 сервера БД поточними запитами.

Під час визначення моменту наступної синхронізації обчислюється значення рівня актуальності даних $k_{акт} = \sum_{i=1}^m (k_{R(i)} * (\sum_{j=1}^n k_{A(j)} * \sum_{q=1}^p (modif_{(j,q)} * k_{tup(q)})))$. Якщо $k_{акт}$ перевищує гранично припустиме значення $k_{гран} = f(\Delta t) * (\frac{k_{період}}{\Delta t} + l_{завант})$, то відбувається синхронізація. Гранично припустиме значення залежить від дати останнього оновлення даних та від поточного завантаження сервера, $k_{період}$ – коефіцієнт періодичності оновлення даних (є елементом налаштування підсистеми); Δt – інтервал, протягом якого дані не оновлювалися; $l_{завант}$ – рівень завантаженості сервера (обчислюється за допомогою спеціальних моніторингових програм (наприклад, SQL Profiler або performance monitor); $f(\Delta t)$ – порогова функція необхідності оновлення даних (приймає значення із множини $\{0;1\}$ залежно від перевищення Δt гранично допустимого інтервалу).

Отже, обчислення наступного моменту реплікації даних відбувається, базуючись на значенні функції:

$$f(t) = \begin{cases} 1, & \sum_{i=1}^m (k_{R(i)} * (\sum_{j=1}^n k_{A(j)} * \sum_{q=1}^p (modif_{(j,q)} * k_{tup(q)}))) - \\ & f(\Delta t) * (\frac{k_{період}}{\Delta t} + l_{завант}) > 0 \\ 0, & \sum_{i=1}^m (k_{R(i)} * (\sum_{j=1}^n k_{A(j)} * \sum_{q=1}^p (modif_{(j,q)} * k_{tup(q)}))) - \\ & f(\Delta t) * (\frac{k_{період}}{\Delta t} + l_{завант}) \leq 0 \end{cases} \quad (9)$$

У третьому розділі розроблено метод вибору найкращої альтернативи для граничного рівня маркера представленості даних. Використано метод аналізу ієрархій із автоматичною ініціалізацією матриці попарних порівнянь альтернатив згідно отриманих математичних моделей та нормалізації значень; із використанням елементів нечіткого логічного висновку для дефазифікації вектора глобальних пріоритетів; виконанням класифікації нових даних із використанням алгоритму наївного Байєса.

Під час побудови ієрархічної моделі процесу прийняття рішення вибору кращої альтернативи було використано 4 рівні ієрархії, а саме: «мета», «зацікавлені особи», «критерії», «альтернативи». З метою спрощення задачі було визначено 5 альтернатив рівня маркера представленості: «низький» (Н) – «-1», «нижче середнього» (НС) – «-0,5», «середній» (С) – «0», «вище середнього» (ВС) – «0,5» та «високий» (В) – «1» (рис. 1). Перелік критеріїв оптимальності моделі на рівні ієрархії «критерії» відрізняється для різних зацікавлених осіб. Використовуючи шкалу відносної важливості критеріїв та із залученням людини, що приймає рішення (ЛПР), якою на цьому етапі виступає зацікавлена особа «власник» об'єкта автоматизації, виконується побудова матриці попарних порівнянь для зацікавлених осіб.

На третьому рівні ієрархії складаються відповідні матриці попарних порівнянь за критеріями оптимальності за кожною зацікавленою особою. Тут у ролі ЛПР виступає зацікавлена особа, за якою заповнюється матриця.

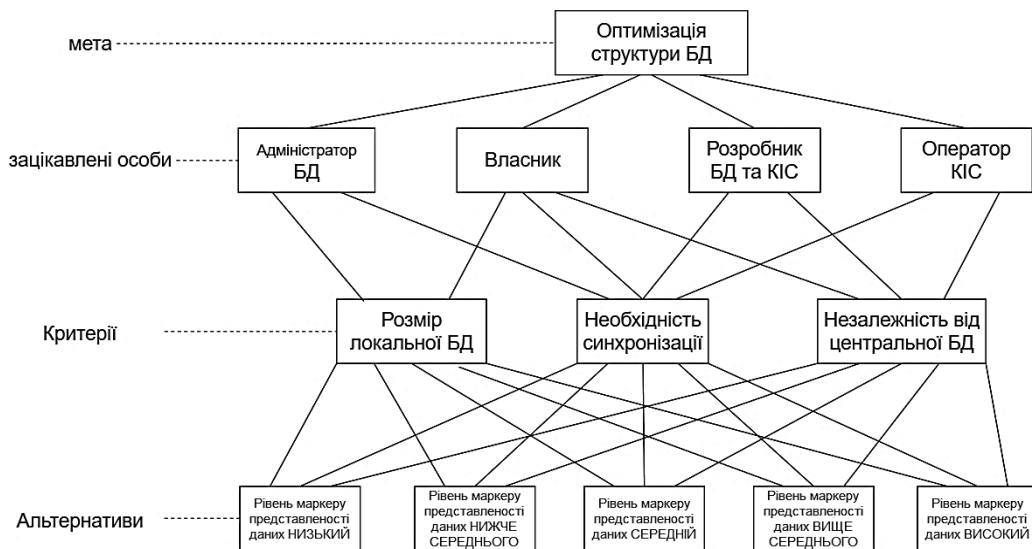


Рисунок 1 – Ієрархічна модель прийняття рішення при виборі кращої альтернативи маркера представленості даних

Наявність математичних моделей критеріїв оптимальності, сформульованих у (4), (6) та (8), дозволяє виконати розрахунок та початкову ініціалізацію матриць попарних порівнянь альтернатив на базі числових значень маркера представленості даних для кожної альтернативи. Отримана матриця подається на розгляд ЛПР для затвердження.

Так, у табл. 1 наведено приклад зміни критерію «розмір БД» вузла РКІС залежно від однієї з п'яти альтернатив для однієї з предметних областей впровадження результатів роботи.

Таблиця 1 – Залежність значення розміру БД від обраної альтернативи

Рівень маркера представленості даних	Низький	Ниже середнього	Середній	Вище середнього	Високий
Розмір локальної БД	0,02	0,24	0,47	0,55	0,75

Керуючись принципами парних порівнянь та аксіомою гомогенності, виконуємо нормування значень, наведених у табл. 1, використавши дещо модифіковану формулу природньої нормалізації:

$$W_i^{norm} = \frac{(W_i - \min_i W_i)}{(\max_i W_i - \min_i W_i)} \times (k - 1) + 1. \quad (10)$$

Після нормування згідно з (10) та округлення до цілого за математичними правилами, матриця попарних порівнянь альтернатив за критерієм «розміру БД» для зацікавленої особи «власник» виглядає наступним чином (табл. 2).

Таблиця 2 – Матриця переваг альтернатив за критерієм «розмір БД» для зацікавленої особи «власник»

Розмір БД	Низький	Ниже середнього	Середній	Вище середнього	Високий
Низький	1	3	6	7	9
Ниже середнього	1/3	1	2	2	3
Середній	1/6	1/2	1	1	2
Вище середнього	1/7	1/2	1	1	1
Високий	1/9	1/3	1/2	1	1

При розрахунку коефіцієнтів для альтернатив за критерієм «незалежність БД», враховуємо максимізацію критерію, як мету і, відповідно, (10) набуває такого вигляду $W_i^{norm} = \frac{(W_i^{\sim} - \min_i W_i^{\sim})}{(\max_i W_i^{\sim} - \min_i W_i^{\sim})} \times (k - 1) + 1$, де $W_i^{\sim} = (1 - W_i)$.

Відповідно до вимог ЛПР, у ролі яких виступають елементи рівня ієрархії «зацікавлені особи» (рис. 1), виконується обмеження області значень критеріїв оптимальності рівня маркера представленості даних на вузлі РКІС.

Так, за критерієм оптимальності «Незалежність БД» для однієї з предметних областей впровадження результатів роботи введено обмеження мінімального значення на рівні 0,4, тобто щонайменше 40 % запитів до БД вузла РКІС має оброблятися локально. Розрахунок значень критерію «Незалежність БД», виконаний на базі (4) для кожної альтернативи (аналогічно табл. 1), дозволяє визначити невідповідність альтернативи для інтервалу значення маркера представленості даних «Низький» введеному обмеженню для критерію оптимальності «Незалежність БД». Відповідно до цього, зазначена альтернатива виключається. Обмеження за критерієм «Розмір БД» було сформульовано на рівні максимально допустимого значення 0,65, відповідно до якого альтернатива маркера представленості даних «Високий» також виключається із множини альтернатив. Згідно з встановленими обмеженнями, матриця попарних порівнянь, наведена у табл. 2, зазнає таких змін (рис. 2).

Розмір локальної БД	Низький	Нижче середнього	Середній	Вище середнього	Високий	
Низький	1	3	6	7	9	→
Нижче середнього	1/3	1	2	2	3	
Середній	1/6	1/2	1	1	2	→
Вище середнього	1/7	1/2	1	1	1	
Високий	1/9	1/3	1/2	1	1	

Рисунок 2 – Фільтрація матриці попарних порівнянь альтернатив

Розрахунок вектора глобальних пріоритетів для однієї з предметних областей впровадження результатів роботи, виконані згідно наведеної на рис. 1 ієрархічної моделі та складених матриць попарних порівнянь різних рівнів ієрархії дають такий результат:

$$W^{\text{глоб}} = \begin{bmatrix} 0,000 \\ 0,273 \\ 0,334 \\ 0,393 \\ 0,000 \end{bmatrix} \quad (11)$$

Згідно (11) оптимальним граничним рівнем маркера представленості даних може бути обране значення 0,5, що відповідає альтернативі «вище середнього». З метою підвищення точності результату запропоновано представити вектор глобальних пріоритетів альтернатив у вигляді вектора нечітких чисел для маркера представленості даних. Для отримання числового значення оптимального рівня маркера представленості даних виконується об'єднання результатів та дефазифікація.

Відповідно до отриманого результату (рис. 3), маємо оптимальне значення маркера представленості даних вузла РКІС для предметної області впровадження результатів роботи на рівні 0,2.

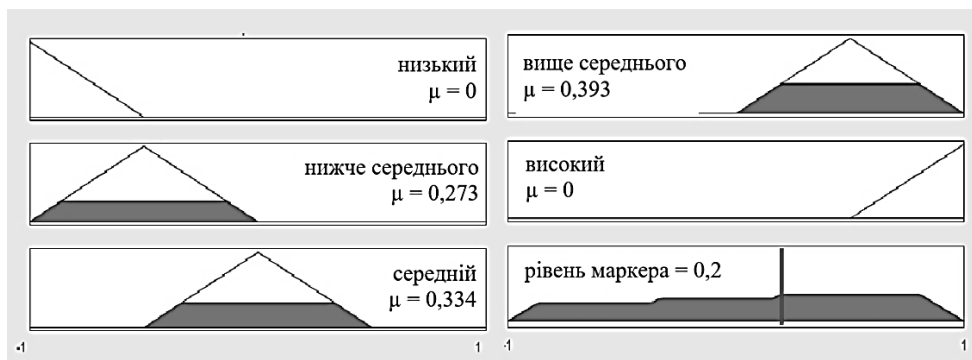


Рисунок 3 – Виконання агрегації та дефазифікації отриманих результатів

Для перевірки отриманих результатів було виконано табулювання інтервалу допустимих значень маркера представленості даних $[-1, 1]$ із кроком 0,1, що дає 21 альтернативу на останньому рівні ієрархії моделі (рис. 1). Далі виконано спрощений розрахунок вектора глобальних пріоритетів із використанням матриць

попарних порівнянь для 2 та 3 рівнів ієрархії, отриманих від ЛПР, та автоматично заповнених матриць для 4 рівня (розмірність 21×21) згідно (4), (6) та (8). Отриманий результат, що наведено у табл. 3, співпадає із результатом для 5 альтернатив при такому ж спрощеному розрахунку, але із додатковим етапом дефазифікації вектора пріоритетів альтернатив (рис. 4).

Таблиця 3 – Вектор глобальних пріоритетів для 21 альтернативи

Альтернатива	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
Маркер	-1,0	-0,9	-0,8	-0,7	-0,6	-0,5	-0,4	-0,3	-0,2	-0,1	0,0
Пріоритет	0,008	0,012	0,009	0,013	0,010	0,012	0,009	0,01	0,013	0,021	0,034
Альтернатива	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	
Маркер	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	
Пріоритет	0,05	0,054	0,063	0,076	0,082	0,101	0,117	0,111	0,103	0,093	

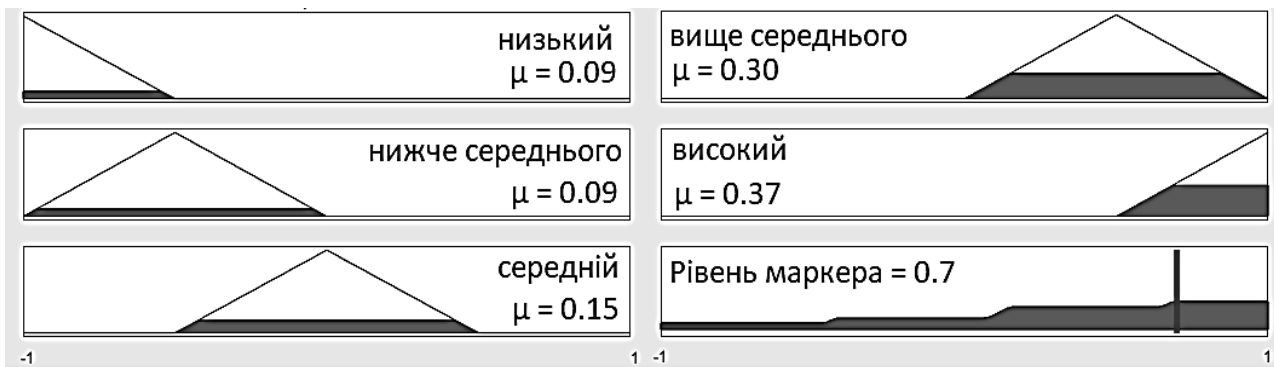


Рисунок 4 – Дефазифікація результатів, отриманих для 5 альтернатив

Виконання повного аналізу щодо необхідності представлення атрибутів та кортежів таблиць БД на вузлі РКІС є досить ресурсоємкою операцією та не може виконуватись часто. Відповідно до цього виникає завдання визначення корисності для вузла нових даних РКІС, доданих після проведення описаного вище аналізу.

Оскільки на рівень представленості під час аналізу користувацьких запитів до БД впливає комбінація значень атрибутів кортежу відношення, запропоновано представити завдання визначення рівня необхідності представлення нових даних у вигляді задачі класифікації. На вході маємо відношення та перелік значень атрибутів кортежу, а на виході – рішення про його корисність для БД вузла РКІС.

Промислові реляційні БД у більшості випадків перебувають у 3-й нормальній формі, що свідчить про відсутність транзитивних залежностей у межах відношення. Базуючись на цьому, використано наївний Байєсівський алгоритм. Отже, для кожного нового кортежу X розраховуються ймовірність «потрібності» на вузлі РКІС:

$$P(\text{потрібен} | X) = \frac{\prod_{i=1}^n P_{x_i}(\text{потрібен} | x_i)}{\prod_{i=1}^n P_{x_i}(\text{потрібен} | x_i) + \prod_{i=1}^n P_{x_i}(\text{непотрібен} | x_i)}, \quad (12)$$

де x_i – значення i -го атрибуту кортежу X .

Якщо значення, отримане за (12), перевищує оптимальне граничне значення маркеру представленості даних, приймаємо рішення про потрібність нового кортежу на віддаленому вузлі РБД.

$$F_X^{\text{Потрібен}} = \begin{cases} \text{істина, } \frac{(coef_{\text{предст}}^{node} + 1)}{2} > P(\text{потрібен} | X) \\ \text{хибність, } \frac{(coef_{\text{предст}}^{node} + 1)}{2} < P(\text{потрібен} | X) \end{cases}$$

Четвертий розділ присвячено створенню інформаційної технології обліку та аналізу SQL-запитів. Для можливості отримання набору відношень, атрибутів та кортежів SQL-запиту реалізовано підсистему парсингу тексту SQL-запиту. Для цього розроблено граматику мови SQL, що дозволило із використанням програмного продукту ANTLR виділити множини відношень, атрибутів та вкладених підзапитів. На рис. 5 наведено приклад дерева граматичного розбору запиту із одним вкладеним підзапитом та двома атрибутами (рис. 5).

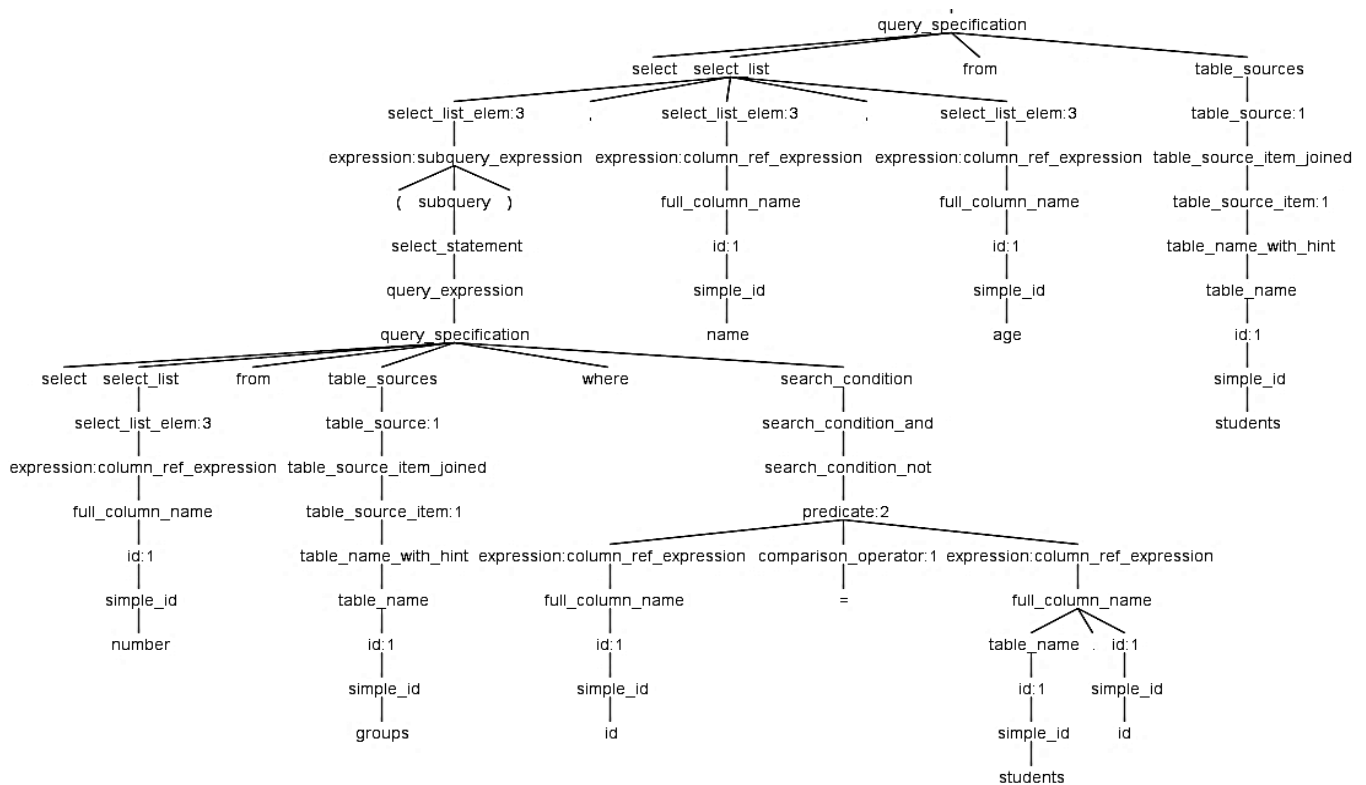


Рисунок 5 – Дерево граматичного розбору

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.

У процесі проєктування структури БД обліку SQL-запитів на концептуальному рівні можна виділити такий перелік сутностей: «місце розташування»; «робоча станція»; «програмне забезпечення»; «користувачський запит»; «відношення»; «атрибут»; «кортеж»; «транзакція». Після приведення схеми БД до третьої нормальної форми, до схеми додаються такі відношення, як «тип робочої

станції»; «програмне забезпечення робочої станції»; «журнал транзакцій»; «вкладені запити»; «список відношень запиту»; «список атрибутів запиту»; «список кортежів запиту».

Враховуючи великі обсяги статистичних даних, які фіксують виконання операцій користувачів із БД РКІС, а також необхідність виконання подальшого аналізу накопичених даних з точки зору множинності вимірів, дані SQL-запитів представлено у вигляді багатовимірної бази даних (ББД).

Для реалізації таблиці фактів та таблиць вимірів у вигляді схеми «зірка», на рівні реляційної БД ІС обліку користувацьких запитів вводиться ряд представлень. Цей підхід підвищує прозорість структури БД для аналітиків та є своєрідним інтерфейсом при взаємодії багатовимірної моделі із реляційною. За умови використання такого підходу стає можливим внесення змін у реляційну модель БД без необхідності виконувати відповідні модифікації структури ББД.

ББД реалізовано на базі SQL Analysis Server із використанням пакету SQL Server Development Tools for Business Intelligence. Структура всіх вимірів багатовимірного кубу є типовою. Вона містить один описовий атрибут «назва» та унікальний ідентифікатор об'єкту, що є первинним ключем для відношення. На базі створених вимірів та двох таблиць фактів, окремо для атрибутів та кортежів відношення, створюється два OLAP-куба, що мають в своїй основі схему «зірка» – по одній таблиці фактів та по одній таблиці на кожний вимір ББД. Схему ББД наведено на рис. 6.

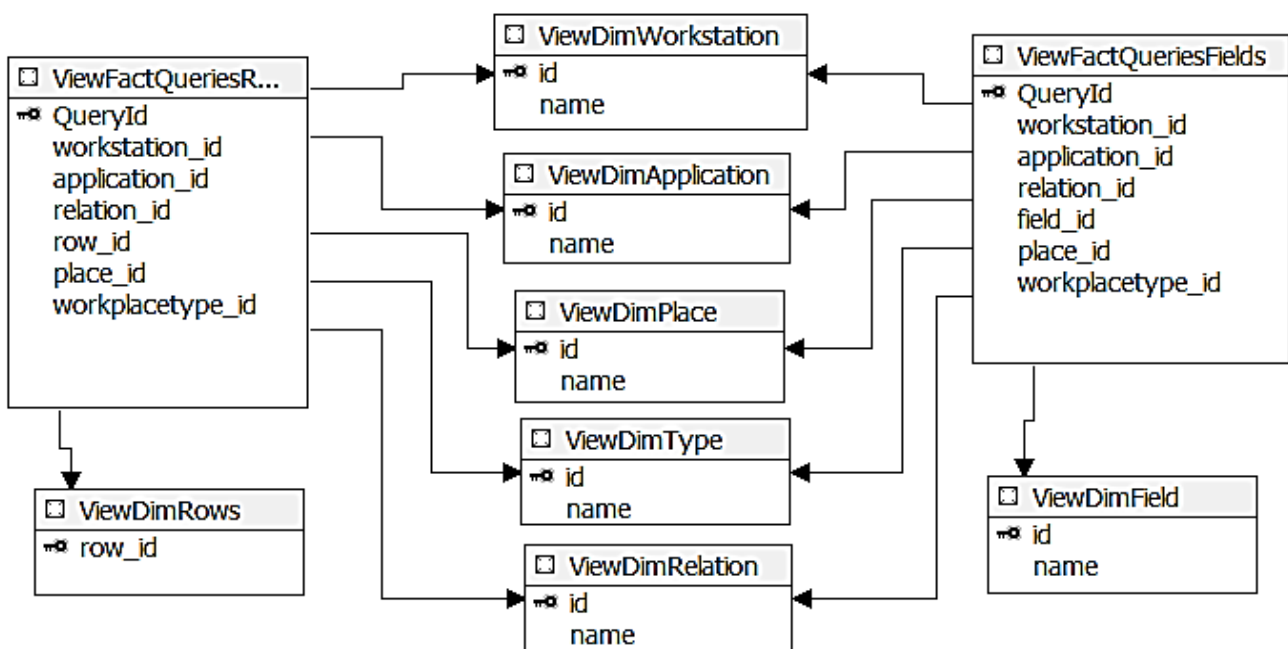


Рисунок 6 – Data Source View підсистеми оперативно-аналітичного аналізу запитів

У якості клієнтського застосунку в межах поточного дослідження використано MS Excel, що має механізми з підключення та отримання даних із SQL Analysis Services та надає достатній рівень функціональних можливостей для побудови зрізів даних, необхідних ЛПР для прийняття рішення щодо подальшої фільтрації даних під час розрахунку значень критеріїв оптимальності структури вузла РКІС.

Розроблена ББД дозволяє ЛПР виконувати оперативно-аналітичний аналіз даних з метою виявлення оптимального набору необхідного на вузлі РКІС програмного забезпечення. Виконання онлайн операцій консолідації, деталізації та зрізу дозволяє визначити множини атрибутів та кортежів відношень БД, до яких виконуються SQL-запити вузла РКІС та визначити їх кількість.

У п'ятому розділі представлено програмну реалізацію інформаційної технології визначення оптимального рівня маркеру представленості даних на вузлі РКІС, що виконана у вигляді сервіс-орієнтованого вебзастосунку на основі трьох-ланкової архітектури. У якості сервера БД обрано SQL Server, бекенд складова (application server) реалізована на базі фреймворку Symfony із використанням скриптової мови PHP. Взаємодія із БД реалізована із використанням Object-Relational Mapper (ORM) Doctrine, що дозволяє оперувати із даними БД за допомогою об'єктної моделі. Даний прошарок взаємодіє із сервером БД за допомогою команд мови SQL та повертає екземпляри класів сутностей у середовище бекенд. Фронтенд складова реалізована на базі фреймворку Angular та мови програмування TypeScript (рис. 7).

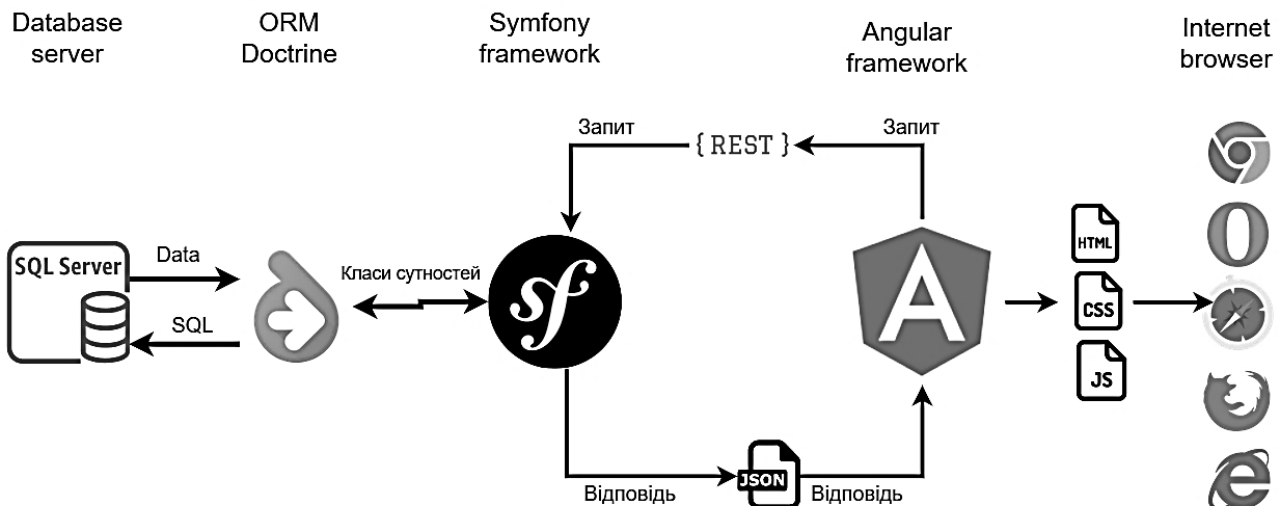


Рисунок 7 – Взаємодія елементів розробленої сервіс-орієнтованої інформаційної технології

Бекенд та фронтенд компоненти взаємодіють через REST-API із обміном даними у json-форматі. Наведена архітектура є досить гнучкою та дозволяє за необхідності змінювати СКБД та підключатись до бекенд Application Programming Interface (API), використовуючи стороннє програмне забезпечення.

Для взаємодії із БД через ORM Doctrine на рівні бекенд компоненти реалізовані класи-сутності «Робоча станція», «Тип робочої станції», «Застосунок», «Місце розташування», «Критерій оптимальності», «Матриця» (рис. 8). Класи розширюють інтерфейс JsonSerializer для можливості експорту даних у json-форматі. Бекенд компонента реалізує маршрути для отримання необхідних даних щодо перерахованих сутностей. У табл. 4 наведено приклад описання маршруту для встановлення обмежень за критерієм оптимальності.

Фронтенд складова реалізована у вигляді набору компонент та сервісів. Так, робота із критеріями оптимальності реалізована у вигляді компоненти

CriteriaChartsComponent, що відповідає за логіку обробки та виведення даних, та двох сервісів: CriteriaService та CompareMatrixService, у яких реалізована логіка взаємодії із бекенд частиною застосунку. Діаграму класів для фронтенд складової реалізації задачі встановлення обмежень за критеріями оптимальності наведено на рис. 9.

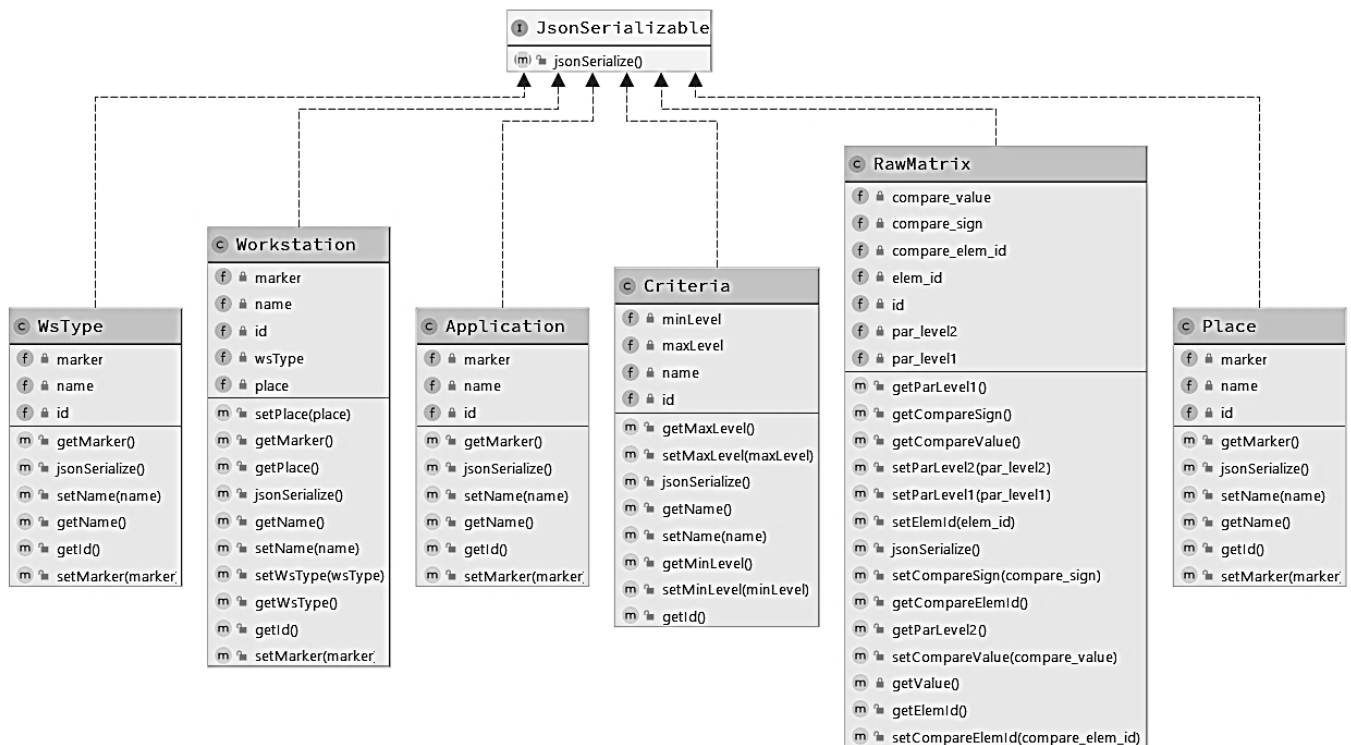


Рисунок 8 – Класи-сутності для взаємодії із створеною БД

Таблиця 4 – Маршрут встановлення обмежень за критерієм оптимальності

URI	Метод / Призначення	Параметри / тіло запиту	Структура відповіді (json)	Опис відповіді
/criteria/set-min-max	POST / Зміна значень мінімального та максимального рівня критерію оптимальності	Body: {code: number, min: number, max: number} – ідентифікатор критерію оптимальності, мінімальне та максимальне допустимі значення	{value: [{alt:number, disable: boolean}]}	Масив альтернатив для рівня маркера представленості даних із зазначенням допустимості альтернативи відповідно до діючих обмежень

Користувачський інтерфейс відображає отримані дані у вигляді набору графіків, по одному на кожен критерій оптимальності. Графік відображає ряд значень критерію відповідно до кожної альтернативи. Через елементи користувацького інтерфейсу ЛПР встановлює мінімальний та максимальний рівень за кожним критерієм (рис. 10).

Роботу з іншими сутностями БД реалізовано аналогічним чином.

Користувач інформаційної технології має змогу переглянути всі етапи формування глобального вектора пріоритетів альтернатив. Це дає змогу оцінити вплив кожної складеної матриці попарних порівнянь на всіх рівнях ієрархії моделі та, відповідно, приймати більш осмислене рішення щодо визначення оптимального рівня маркера представленості даних на вузлі РКІС. На рис. 11 наведено вектор

глобальних пріоритетів альтернатив у вигляді діаграми із урахуванням звуження множини альтернатив відповідно до встановлених обмежень за критеріями оптимальності.

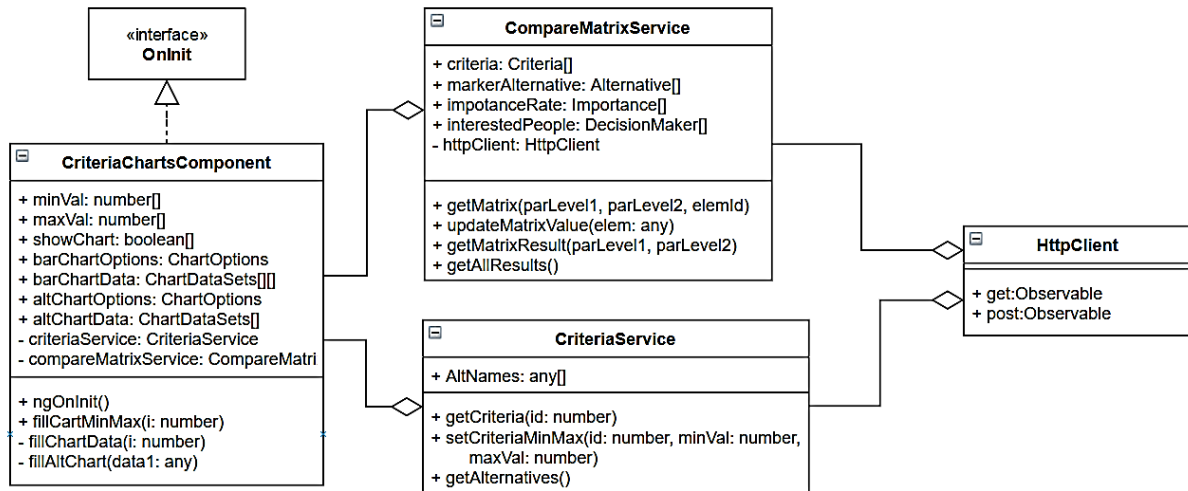


Рисунок 9 – Діаграма класів. Компонента CriteriaChartsComponent та сервіси CriteriaService та CompareMatrixService

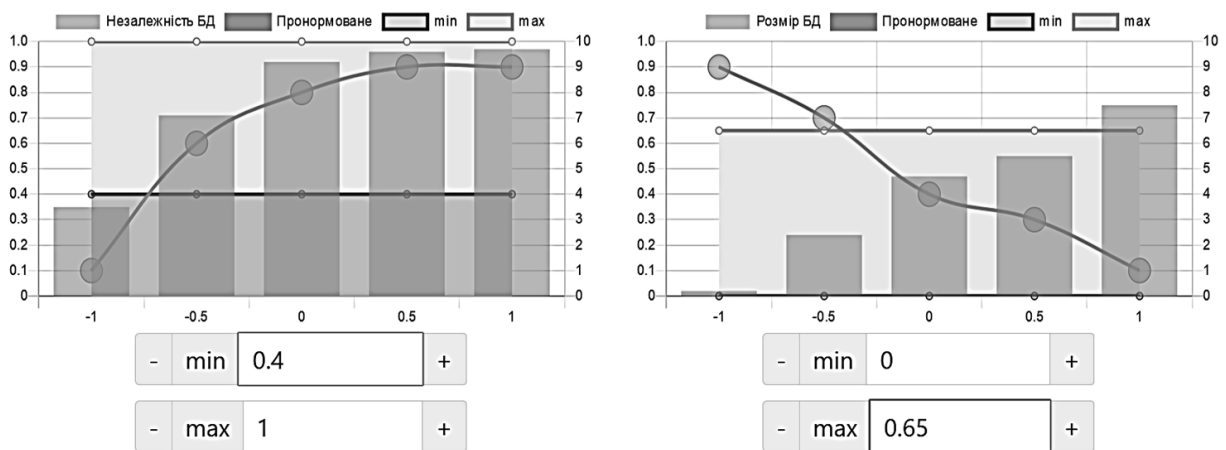


Рисунок 10 – Дані за критеріями оптимальності «Незалежність БД» та «Розмір БД»

Глобальний вектор пріоритетів

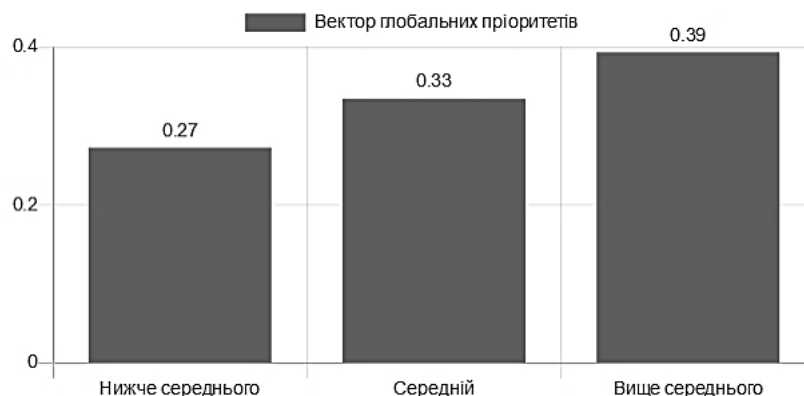


Рисунок 11 – Діаграма глобальних пріоритетів альтернатив

Далі також у вигляді діаграм відображено пріоритети за критеріями оптимальності по зацікавленим особам. Під час обробки та виведенні даних враховується особливість ієрархічної моделі відносно несуттєвості деяких критеріїв оптимальності для окремих зацікавлених осіб.

Етап обробки отриманого результату полягає у представленні вектора глобальних пріоритетів альтернатив у вигляді набору нечітких множин однієї змінної із подальшим об'єднанням результатів та дефазифікацію результату для отримання числового значення оптимального рівня маркера представленості даних. Цей етап виконано у середовищі MathLab із використанням модуля Fuzzy Logic Toolbox, що дозволяє розробляти та виконувати симуляцію для систем нечіткого логічного висновку. Вихідний сигнал визначаємо на множині із п'яти інтервалів значення маркера представленості даних у вигляді нечітких чисел трикутної форми. Використання інструмента Rule Viewer Fuzzy Logic Toolbox дозволяє отримати дефазифіковане значення вихідної змінної після виконаної агрегації набору наведених правил. Для предметної області впровадження результатів роботи отримано оптимальне значення маркера представленості даних на вузлі РКІС на рівні 0,2.

При розрахунку значення ефективності використання вузла РКІС, було виведено співвідношення отриманого результату до витрачених ресурсів на базі сформульованих у (4), (6) та (8) критеріїв оптимальності структури БД і отриманого у ході вирішення задачі вибору кращої альтернативи вектора відносної ваги критеріїв оптимальності.

$$Eff = \frac{F_{\text{доступність}} \times W_1^{\text{крит}}}{F_{\text{size}} \times W_0^{\text{крит}} + F_{\text{synchron}} \times W_2^{\text{крит}}} \cdot \quad (13)$$

Обчислені за (13) значення для випадків представлення на вузлі РКІС лише критичних даних; всіх необхідних даних, а також даних, що відповідають граничному рівню маркера представленості даних 0,2 та 0,5 відповідно, наведено у табл. 5.

Таблиця 5 – Порівняння ефективності вузла РКІС для різних значень граничного рівня представленості даних вузла РКІС

		Лише критичні дані	Всі необхідні дані	Рівень маркера 0,2	Рівень маркера 0,5
Незалежність	0,73	0,35	0,97	0,85	0,92
Розмір БД	0,08	0,02	0,75	0,34	0,47
Необх. синхронізації	0,19	0,15	0,07	0,15	0,13
Ефективність БД вузла		8,56	9,59	11,13	10,75
Підвищення ефективності		30,07 %	16,10 %	—	—

Отримані результати дозволяють зробити висновок про підвищення ефективності у межах від 16 % до 30 % у випадку вибору значення граничного рівня маркера представленості даних на рівні 0,2 порівняно з розміщенням на вузлі РКІС всіх необхідних даних або лише критичних даних відповідно.

ВИСНОВКИ

У дисертації на основі виконаних теоретичних досліджень вирішено актуальну науково-технічну задачу створення моделей та інформаційної технології оптимізації структури БД вузла у корпоративних інформаційних системах.

У ході виконання дослідження:

1. Виконано аналіз особливостей задач використання інтегрованого комплексу територіально розосереджених ІС та методів їх оптимізації.
2. Побудовано модель SQL-запиту та формалізовано критерії оптимальності структури БД віддаленого вузла КІС.
3. Розроблено інформаційну технологію парсингу та подальшого обліку SQL-запитів до реляційної БД, а також обчислення значень критеріїв оптимальності відповідно до результатів операцій зрізу та консолідації.
4. Сформульовано та вирішено задачу багатокритеріальної оптимізації для визначення оптимального рівня представленості даних на вузлі РКІС.
5. Розроблено інформаційну технологію обчислення вектора глобальних пріоритетів альтернатив із представленням результату у вигляді вектора нечітких чисел та приведенням до чіткого значення.
6. Реалізовано класифікацію нових даних БД відповідно до необхідності їх представлення на вузлі РКІС.

Наукова новизна отриманих результатів:

1. Вперше уведено поняття маркера 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 РКІС на основі статистики SQL-запитів.
2. Вперше побудовано модель залежності критеріїв оптимальності структури БД вузла РКІС від значення маркера представленості даних, що на відміну від існуючих підходів дозволяє визначити оптимальне граничне значення маркера на основі статистики SQL-запитів.
3. Вперше розроблено інформаційну технологію розрахунку критеріїв оптимальності структури БД вузла РКІС, що дозволяє визначити граничний рівень маркера представленості даних і прийняти рішення про представленість даних на вузлі РКІС, та на відміну від існуючих використовує аналітичні дані та результати парсингу тексту SQL-запитів до серверів РКІС.
4. Удосконалено модель SQL-запиту, яка, на відміну від інших, представлена моделлю багатовимірної БД в аналітиці множини атрибутів та кортежів відношення, а також інших характеристик запиту, що дозволяє проведення оперативно-аналітичного аналізу зважаючи на множинність вимірів з метою розрахунку рівня представленості атрибутів та кортежів відношення БД на вузлі РКІС.
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 РКІС на 4%.
6. Отримав подальшого розвитку метод асинхронної реплікації даних за рахунок динамічного визначення моменту наступної синхронізації даних на

базі статистики SQL-запитів, що дозволяє підвищити актуальність даних, та розвантажити ресурси серверів БД.

Результати дослідження впроваджено на ППФ «Юнікс Трейд Ко» при проєктуванні структури БД віддаленого вузла РКІС. Ефект від впровадження зазначених результатів полягає у підвищенні ефективності використання БД вузла РКІС на 12 %. Розроблені моделі та інформаційна технологія впроваджені на ТОВ «Еліт Білдінг» для виявлення залежності критеріїв оптимальності структури БД та рівня представленості даних. Ефект від впровадження полягає у підвищенні швидкості виконання запитів на 14 % та доступності даних на 21 %.

Здійснено впровадження результатів дисертаційного дослідження в навчальний процес на кафедрі «Інженерії програмного забезпечення» при викладанні дисциплін «Інформаційні технології OLTP, OLAP та DM на клієнт-серверній платформі», «Клієнт-серверні СКБД та аналітичні системи», «Database development».

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Фісун М. Т., Дворецький М. Л. Пошук асоціативних правил засобами мови SQL при інтелектуальному аналізі даних. *Вісник Херсонського національного технічного університету*. 2005. № 1 (21). С. 185–189; **внесок автора:** реалізовано пошуку асоціативних правил засобами мови SQL та аналіз отриманих результатів; **база (и):** Google Scholar, CrossRef.
2. Фісун М. Т., Дворецький М. Л. Застосування OLAP-технологій в інформаційно-управляючих системах. *Інформаційні технології та комп'ютерна інженерія: Міжнародний науково-технічний журнал Вінницького національно-технічного університету*. 2006. №2(6). С. 128–134; **внесок автора:** засобами SQL Server на базі OLAP розроблено інформаційно аналітичну систему для аналізу трендів продажів торгівельного підприємства; **база (и):** Google Scholar.
3. Фісун М. Т., Дворецький М. Л. Синхронізація оновлення даних в гетерогенних інформаційних системах. *Наукові праці: Науково-методичний журнал Миколаївського державного гуманітарного університету ім. Петра Могили. Сер. Комп'ютерні технології*. 2006. Вип. 44. Т. 57. С. 61–66; **внесок автора:** запропоновано та виконано апробацію алгоритму розрахунку моменту наступної синхронізації БД різнорідних ІС; **база (и):** Google Scholar.
4. Дворецький М. Л. Реалізація задачі класифікації засобами мови Transact-SQL при інтелектуальному аналізі даних. *Системні технології: Регіональний міжвузівський збірник наукових праць. Дніпропетровськ, ДНВП «Системні технології»*. 2006. Випуск №6 (47). С. 102–111; **база (и):** Google Scholar, Ulrichsweb, Index Copernicus.
5. Дворецький М. Л. Створення перехресного набору даних (crosstab) засобами Transact SQL. *Вісник Херсонського національного технічного університету*. 2006. № 1 (24). С. 503–507; **база (и):** Google Scholar, CrossRef.
6. Дворецький М. Л. Розв'язання задачі класифікації в багатовимірних базах даних. *Вісник Херсонського національного технічного університету*. 2006. № 1 (30). С. 198–202; **база (и):** Google Scholar, CrossRef.
7. Дворецький М. Л. Інтелектуальний аналіз даних в 1С:8.0. *Наукові праці: Науково-методичний журнал Миколаївського державного гуманітарного університету ім. Петра Могили. Сер. Комп'ютерні технології*. 2007. Вип. 55. Т. 68. С. 141–149; **база (и):** Google Scholar.

8. Дворецький М. Л. Проектування та оцінка оптимальності структури сховища даних та багатовимірної БД. *Наукові праці: Науково-методичний журнал Миколаївського державного гуманітарного університету ім. П. Могили. Сер. Комп'ютерні технології*. 2008. Вип. 77. Т. 90. С. 216–221; **база (и):** Google Scholar.
9. Дворецький М. Л. Інтеграція підсистем обліку та оперативно-аналітичної обробки даних в інформаційній системі супермаркету. *Комп'ютерні науки: освіта, наука, практика. Миколаїв. Національний університет кораблебудування*. 2014. С. 58–60; **база (и):** Google Scholar, CrossRef.
10. Дворецький М. Л. Використання SQL/OLAP в T-SQL при автоматизації операцій консолідації та деталізації кросс-таблиць на базі реляційних джерел даних. *Могилянські читання: досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний аспекти: матеріали всеукр. наук.-метод. конф.* Миколаїв: Вид-во ЧНУ ім. П. Могили, 2014. С. 44–46.
11. Дворецький М. Л., Динамічне формування запиту із використанням SQL/OLAP в T-SQL при автоматизації кросс-таблиць на базі реляційних джерел даних. *Наукові праці: Науково-методичний журнал Чорноморського державного університету ім. П. Могили. Сер. Комп'ютерні технології*. 2014. Вип. 238. Т. 250. С. 32–37; **база (и):** Google Scholar, Ulrichsweb, Index Copernicus.
12. Дворецький М. Л. Задача синхронізації даних в РБД. Синхронне та асинхронне оновлення. *Могилянські читання: досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний аспекти: матеріали всеукр. наук.-метод. конф., 14-18 лист. 2016 р.* Миколаїв: Вид-во ЧНУ ім. Петра Могили, 2016. С. 105–108.
13. Дворецький М. Л., Кулаковська І. В. Порівняльний ABC-XYZ аналіз на базі різних факторів із використанням ієрархічних даних. *Проблеми інформаційних технологій Херсонського національного технічного університету*. 2016. № 19. С. 200–209; **внесок автора:** запропоноване ієрархічне представлення даних та реалізовані механізми подальшого поєднання результатів аналізу за різними критеріями; **база (и):** Index Copernicus, Google Scholar, Research Bible, Open Academic Journals Index, Directory of Open Access Journals..
14. Фісун М. Т., Дворецький М. Л., Юхатов А. В. Порівняльний аналіз методів побудови OLAP-систем із використанням засобів MS SQL SERVER та ORACLE. *Наукові праці: Науково-методичний журнал Чорноморського державного університету ім. П. Могили. Сер. Комп'ютерні технології*. 2016. Вип. 271. Т. 283. С. 36–42; **внесок автора:** виконано аналіз засобів та показників продуктивності щодо автоматизації OLAP на базі СКБД Oracle та SQL Server; **база (и):** Google Scholar, Ulrichsweb, Index Copernicus.
15. Дворецький М. Л., Давиденко Є. О., Боровльова С. Ю. Проектування структури розподіленої БД на базі парсингу SQL-запитів. *Наукові праці: Науково-методичний журнал Чорноморського державного університету ім. Петра Могили. Сер. Комп'ютерні технології*. 2016. Вип. 275. Т. 287. С. 53–61; **внесок автора:** виконано розробку парсеру та багатовимірної моделі SQL-запиту, розроблено СППР оптимізації структури РБД; **база (и):** Google Scholar, Ulrichsweb, Index Copernicus.
16. Дворецький М. Л., Дворецька С. В. Оптимізація механізмів пошуку та оцінка якості знань на базі статистичних даних користувацьких запитів. *Інформаційні технології та взаємодії* : матеріали міжнар. наук.-практ. конф.,

8–10 лист. 2017 р. Київ, 2017. С. 157–158; **внесок автора:** реалізація підсистем накопичення та аналізу статистики користувацьких запитів.

17. Fisun M., Dvoretzkyi M., Shved A., Davydenko Y. Query parsing in order to optimize distributed DB structure. *9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Bucharest. 2017, P. 172–178. doi: 10.1109/IDAACS.2017.8095071; **внесок автора:** розроблено технологію парсингу тексту запиту для оптимізації структури БД; **база (и):** Google Scholar, Web of Science, SCOPUS, IEEE, DBLP..

18. Fisun M., Dvoretzkyi M., Horban H. The usage of the feedback with user activities in company knowlagde management system. *12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv. 2017. P. 143–146. doi: 10.1109/STC-CSIT.2017.8098755; **внесок автора:** використання зворотнього зв'язку користувацької активності в системах управління знаннями та реалізація відповідної технології; **база (и):** Google Scholar, SCOPUS, IEEE.

19. Дворецький М. Л. Підходи щодо підвищення швидкості роботи SQL-запитів при використанні індексів БД. *Ольвійський форум: стратегії країн Причорноморського регіону в геополітичному просторі: матеріали XII міжнар. наук.-практ. конф., 7-10 червня 2018 р. Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2018. С. 30–32.*

20. Дворецький М. Л., Дворецька С. В., Давиденко Є. О. Розробка системи управління знаннями організації на базі CMS WORDPRESS. *Проблеми інформаційних технологій Херсонського національного технічного університету*. 2018. №1 (023). С. 173–180; **внесок автора:** виконано розробку системи управління знаннями на базі CMS Wordpress; **база (и):** Index Copernicus, Google Scholar, Research Bible, Open Academic Journals Index, Directory of Open Access Journals.

21. Fisun M., Horban H., Dvoretzkyi M. Methods of Searching for Association Dependencies in Multidimensional Databases. *13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv. 2018, P. 88–93. doi: 10.1109/STC-CSIT.2018.8526737; **внесок автора:** виконано аналіз результатів пошуку асоціацій на базі багатовимірної бази даних; **база (и):** Google Scholar, SCOPUS, IEEE.

22. Дворецький М. Л., Дворецька С. В., Давиденко Є. О. Використання angular та symfony при реалізації web-орієнтованого застосунку автоматизації обліку торгової точки. *Наукові праці: Науково-методичний журнал Чорноморського національного університету ім. Петра Могили. Сер. Комп'ютерні технології*. 2018. Вип. 305. Т. 317. С. 70–77; **внесок автора:** реалізовано взаємодію бекенд та фронтенд фреймворків при розробці сервіс-орієнтованого вебзастосунок; **база (и):** Google Scholar, Ulrichsweb, Index Copernicus.

23. Fisun M., Dvoretzkyi M., Horban H., Komar M. Knowledge management applications based on user activities feedback. *International Journal of Computing, Ternopil*. 2019. 18 (1). P. 32–44; **внесок автора:** дослідження впливу зворотнього зв'язку активності користувачів систем управління знаннями; **база (и):** Google Scholar, SCOPUS.

24. Дворецький М. Л., Боровльова С. Ю., Дворецька С. В. WEB-застосунок складського обліку в неавтоматизованих торгових точках. *Наукові праці: Науково-методичний журнал Чорноморського національного університету ім. Петра Могили. Сер. Комп'ютерні технології*. 2018. Вип. 308 Т. 320. С. 45–52; **внесок**

автора: виконано аналіз різних типів автоматизації облікових систем та реалізовано відповідний web-застосунок; **база (и):** *Google Scholar, Ulrichsweb, Index Copernicus*.

25. Дворецький М. Л., Дворецька С. В. Використання brain.js при визначенні корисності кортежу для віддалено вузла РБД. *Могилянські читання: досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний аспекти*: матеріали XXII Всеукр. наук.-метод. конф., 11–16 лист. 2019 р. Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2019. С. 123–125; **внесок автора:** реалізація підсистеми визначення корисності кортежу на вузлі РКІС із використанням нейронних мереж.

26. Дворецький М. Л., Дворецька С. В., Давиденко Є. О. Інформаційна технологія визначення корисних даних при оптимізації структури та мінімізації обсягів вузла розподіленої БД. *Вісник Черкаського державного технологічного університету*. 2019. № 4. С. 26–35. doi: 10.24025/2306-4412.4.2019.184808; **внесок автора:** реалізація інформаційної технології оптимізації структури БД вузла РКІС; **база (и):** *Google Scholar, Crossref, Index Copernicus, Eurazian Scientific Journal Index, Directory of Open Access Journals*.

27. Dvoretzkyi M., Dvoretzka S., Nezdoliy Y., Borovlova S. Data Utility Assessment while Optimizing the Structure and Minimizing the Volume of a Distributed Database Node, *1st International Workshop on Information-Communication Technologies & Embedded Systems (ICTES)*, Mykolaiv, Ukraine, November 14-15. 2019. P. 128–137; **внесок автора:** вирішення задачі багатокритеріальної оптимізації при проєктуванні структури БД вузла РКІС; **база (и):** SCOPUS.

28. Фісун М. Т., Дворецький М. Л., Дворецька С. В. Побудова моделей для оптимізації структури бази даних вузла у корпоративних інформаційних системах. *Інформаційні технології та комп'ютерна інженерія: Міжнародний науково-технічний журнал Вінницького національно-технічного університету*. Vol 48, № 2. 2020. С. 52–60. doi: <https://doi.org/10.31649/1999-9941-2020-48-2-52-60>; **внесок автора:** реалізація моделей SQL-запиту та критеріїв оптимальності структури БД вузла РКІС; **база (и):** *Google Scholar*.

АНОТАЦІЯ

Дворецький М. Л. Моделі та інформаційна технологія оптимізації структури бази даних вузла у корпоративних інформаційних системах – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Чорноморський національний університет імені Петра Могили, Миколаїв, 2020.

Метою роботи є підвищення рівня доступності даних та ефективності використання розподілених та територіально розосереджених комп'ютерних систем шляхом реалізації інформаційної технології оптимізації структури БД віддаленого вузла на основі статистики SQL-запитів.

У роботі побудовано модель залежності критеріїв оптимальності структури БД віддаленого вузла РКІС від значення маркера представленості даних згідно зі статистичними даними SQL-запитів до БД. Розрахунок значень здійснюється на базі багатовимірної БД із виконанням відповідних зрізів даних згідно розробленої функції агрегації. Удосконалено модель SQL-запиту, яка представлена на базі багатовимірної

БД із можливістю виконання оперативно-аналітичного аналізу зважаючи на множинність вимірів. При побудові моделі запропоновано отримувати діапазони ключів запиту для визначення множини кортежів відношення результуючого набору рядків. При виборі найкращої альтернативи запропоновано модифікований метод аналізу ієрархій із автоматичною ініціалізацією матриці попарних порівнянь альтернатив згідно отриманих математичних моделей та нормалізації значень. Використано елементи нечіткого логічного висновку для дефазифікації вектора глобальних пріоритетів та виконано класифікацію нових даних із використанням алгоритму наївного Байєса. Розроблено інформаційну технологію розрахунку критеріїв оптимальності структури БД вузла РКІС на базі статистики SQL-запитів та визначення оптимального рівня представленості даних на вузлі РКІС. Виконано впровадження результатів роботи в діючі РКІС та отримано позитивний ефект.

Ключові слова: корпоративна інформаційна система, реляційна база даних, SQL-запит, парсинг тексту, метод аналізу ієрархій, вебсервіс.

ABSTRACT

Dvoretzkyi M. L. Models and information technology for optimizing the structure of the node database in corporate information systems. – Manuscript

The thesis for a Degree of Candidate of Science (Engineering) in specialty 05.13.06 – information technology. – Petro Mohyla Black Sea National University, Mykolaiv, 2020.

The aim of the work is to increase the level of data availability and efficiency of distributed computer systems use by implementing information technology to optimize the structure of the database of the remote node based on the statistics of SQL queries

The work creates the dependence model of the criteria of database optimal structure of CIS node on the value of the data representation marker according to the statistical data of SQL queries to the database. The values are calculated on the basis of a multidimensional database with obtaining of the corresponding data slices according to the developed aggregation function. The SQL query model has been improved, which is presented on the basis of a multidimensional database with the possibility of performing operational and analytical analysis according to the multiplicity of dimensions. When building the model, it is proposed to obtain ranges of query keys to determine the set of tuples of the relation of the resulting set of rows. A modified analytic hierarchy process with automatic initialization of the matrix of pairwise comparisons of alternatives according to the obtained mathematical models and normalization of values is proposed. Elements of fuzzy inference were used to dephase the vector of global priorities. The classification of new data was performed using the naive Bayesian algorithm. The information technology (IT) of calculation of optimality criteria of the CIS DB structure node is developed. The IT bases on the statistics of SQL-queries and calculates an optimum level of the data representation on DCIS node. The results of the research were implemented in the working DCIS and positive effect was obtained.

Keywords: corporate information system, relational database, SQL query, text parsing, analytic hierarchy process, web service.

Підп. до друку 17.11.2020.
Формат 60×84 ¹/₁₆. Папір офсетний.
Гарнітура «Times New Roman». Друк ризограф.
Умовн. друк. арк. 0,9. Тираж 100 пр. Зам. 6120.

54003, м. Миколаїв, вул. 68 Десантників, 10.
Тел.: (0512)50-03-32, (0512)56-55-81.
Свідоцтво суб'єкта видавничої справи ДК № 6124 від 05.04.2018.

