



ВИКОРИСТАННЯ МЕТОДУ АНАЛІЗУ ІЄРАРХІЙ ДЛЯ ОПТИМІЗАЦІЇ СТРУКТУРИ БАЗИ ДАНИХ ВУЗЛА У КОРПОРАТИВНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

УДК 004.658:652.3

DOI: <https://doi.org/10.35546/2313-0687.2020.27.23-35>

Микола Фісун,

доктор технічних наук, професор кафедри інженерії програмного забезпечення,
Чорноморський національний університет імені Петра Могили, м. Миколаїв, Україна,
E-mail: mykola.fisun@gmail.com, 0000-0003-1297-6230

Михайло Дворецький,

ст. викладач кафедри інженерії програмного забезпечення,
Чорноморський національний університет імені Петра Могили, м. Миколаїв, Україна,
E-mail: m.dvoretzkiy@gmail.com, 0000-0001-5913-6859

Світлана Дворецька,

ст. викладач кафедри інтелектуальних інформаційних систем,
Чорноморський національний університет імені Петра Могили, м. Миколаїв, Україна,
E-mail: svetag603@gmail.com, 0000-0001-5199-9430

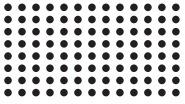
Євген Давиденко,

кандидат технічних наук, доцент кафедри інженерії програмного забезпечення,
Чорноморський національний університет імені Петра Могили, м. Миколаїв, Україна,
E-mail: genik.davydenko@gmail.com, 0000-0002-0547-3689

Мета статті. Метою статті є вирішення задачі багатокритеріальної оптимізації структури бази даних вузла у корпоративних інформаційних системах із використанням методу аналізу ієрархій для підвищення рівня доступності даних та ефективності використання розподілених та територіально розосереджених комп'ютерних систем.

Методи дослідження. Методи досліджень базуються на основних засадах теорії реляційних баз даних (у тому числі розподілених) та методах багатокритеріальної оптимізації. Зокрема, на: методах реляційної алгебри, теорії реляційної моделі даних, багатовимірній моделі даних – при побудові моделі SQL-запиту; методі аналізу ієрархій – при вирішенні задачі вибору найкращої альтернативи рівня представленості даних.

Основні результати дослідження. Введено поняття маркеру представленості даних, що відображає рівень необхідності представлення даних на вузлі РКІС. Для кожного елемента значення маркеру приймається за одне із множини значень {необхідно, бажано, не потрібно}, що визначає ступінь необхідності представлення даних того чи іншого типу робочого місця, ролі користувача або застосунку. Сформульовані критерії оптимальності структури БД дозволяють говорити про задачу багатокритеріальної оптимізації, вирішення якої виконується із використанням методу аналізу ієрархій. метод аналізу ієрархій (MAI), що є загальною методологією розв'язання широкого класу задач прийняття



рішень, дозволяє поєднати порівняно простий математичний апарат зі знаннями та досвідом ЛПР. Запропонована 4-х рівнева ієрархічна модель наступного вигляду: ціль – зацікавлені особи – критерії оптимальності – альтернативи. Серед особливостей використання методу слід відмітити: різний набір критеріїв оптимальності для різних зацікавлених осіб; квантування безперервної ознаки маркеру представленості даних із виділенням 5 альтернатив та автоматичне попереднє заповнення матриці попарних порівнянь на останньому рівні ієрархії.

Наукова новизна. Вперше уведено 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; 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.

Практична значимість. Розв'язання задачі багатокритеріального аналізу та вибір найкращої альтернативи дозволяє визначити оптимальний рівень значення маркеру представленості даних, що у свою чергу дозволяє класифікувати атрибути та кортежі відношень БД згідно їх представлення на вузлі РКІС. Проведені розрахунки для однієї з предметних областей впровадження результатів дослідження, результати яких наведені у відповідних таблицях дослідження, дозволяють говорити про підвищення ефективності БД вузла РКІС на 25% порівняно із представленням лише критичних даних, та на 11% порівняно із представленням всіх необхідних даних центральної БД відповідно.

Ключові слова: розподілена СКБД, корпоративна інформаційна система, багатокритеріальна оптимізація, реляційна модель даних, SQL-запит, метод аналізу ієрархій, матриця попарних порівнянь, вектор пріоритетів, маркер представленості даних.

Постановка проблеми. Широке використання інформаційних технологій багаторівневих, територіально розосереджених комп'ютерних інформаційних систем (КІС) на основі баз і сховищ даних, у тому числі із розподіленими базами даних, обумовлює необхідність розв'язання задачі автономності роботи користувачів, зниження навантаження на сервери БД та сервіси синхронізації даних шляхом оптимізації структури БД віддаленого вузла КІС. Комбінована стратегія розподілу даних є найбільш виправданою із точки зору можливості поєднання переваг стратегій з та без дублювання [1-3]. Але при її використанні, окрім задачі синхронізації дубльованої інформації, актуальною постає задача оптимального проектування структури БД з точки зору приналежності даних до категорії того чи іншого вузла мережі. Крім того, продуктивність системи напряму буде залежати від прийняття рішення щодо необхідності часткового або повного дублювання даних.

Аналіз останніх досліджень і публікацій. Питання оптимізації структури БД розглядається у роботах [4-10], але недостатній рівень уваги приділяється питанню підвищення продуктивності автоматизованих систем за рахунок оптимізації структури БД РКІС на базі статистики SQL-запитів. Так, у [4-6] при розгляді проектування автоматизованих систем управління та обробки

даних, організації сховищ даних та багатовимірних моделей не розглядаються питання використання комбінованої стратегії розподіленого представлення даних у КІС; а у роботах [7-10] автори розглядають питання підвищення продуктивності автоматизованих систем за рахунок використання матеріалізованих представлень, реструктуризації БД та денормалізації відношень. Однак аспект оптимальності структури окремого вузла РКІС залишається без уваги.

У роботах [3, 11-12], автором на базі реляційної моделі даних та елементів теорії множин було формалізовано поняття зрізу даних множини відношень БД та побудована модель SQL-запиту. Також введено поняття маркеру представленості даних, що відображає рівень необхідності представлення даних на вузлі РКІС. Для кожного елемента значення маркеру приймається за одне із множини значень {необхідно, бажано, не потрібно}, що визначає ступінь необхідності представлення даних того чи іншого типу АРМ, ролі користувача або застосунку. Для визначення оптимального значення рівня маркету представленості введено декілька критеріїв оптимальності та побудовано математичну модель кожного з них, які дозволяють розрахувати їх значення в залежності від представленості даних у віддаленому вузлі територіально розосередженої КС.

Отже, отримано багатокритеріальну задачу, що має бути вирішена для визначення оптимального рівня представленості даних. Слід відмітити, що критерії оптимальності, моделі яких були визначені, є незалежними та монотонними і представлені на множині дійсних чисел в інтервалі [0;1]. Класичні методи Паретто та Слейтора [13-14] можуть дати результати лише на першому етапі моделювання, але при розрахунку оптимального рівня маркеру представленості виявляються малоефективними через зниження рівня одних критеріїв оптимальності при одночасному підвищенні інших. Вирішення задачі також ускладнюється тим фактом, що простір рішень визначається на множині дійсних чисел, а отже множина рішень містить велику кількість альтернатив.

Постановка завдання. Прийняття рішення передбачає вибір одного з можливих варіантів дій, що прийнято називати альтернативами. У деяких випадках альтернатив може виявитись забагато для людини, що приймає рішення (ЛПР), і вимагає залучення додаткових заходів підтримки вибору рішення. До цього випадку потрапляє також і задача визначення оптимального рівня маркету представленості даних, сформульована у [3,12], конкретне значення якого можна вважати однією з множини можливих альтернатив. Кожне значення рівня маркету представленості даних характеризується показниками їх привабливості для ЛПР, а саме такими критеріями оптимальності, як незалежність від центрального вузла БД, співвідношення розміру локальної БД до центральної, та показник рівня необхідності синхронізації даних. Всі вони слугують критеріями оптимальності рішення.

При вирішенні задачі багатокритеріальної оптимізації серед основних підходів слід згадати отримання суперкритерію шляхом згортки критеріїв, методи звуження множини альтернатив, групування критеріїв, а також методи головного критерію, ідеальної точки та послідовних поступок [15]. Методи звуження мно-

жини альтернатив у більшості випадків базуються на визначенні множини рішень, ефективних за Паретто (Стейтором, Смейлом) [13-14], що не може бути застосовано для поточної задачі. Методи групування критеріїв використовуються при великій кількості останніх, що у випадку 3-х критеріїв не є актуальним. Недоліками методів згортки є складність аргументування вибору конкретного методу та обґрунтування вибору вагових коефіцієнтів. Метод головного критерію веде до спрощення задачі та необхідності залучення експертів, а метод ідеальної точки дає різні результати при використанні різних методів визначення відстані [15].

У свою чергу, метод аналізу ієрархій (МАІ), що є загальною методологією розв'язання широкого класу задач прийняття рішень, дозволяє поєднати порівняно простий математичний апарат зі знаннями та досвідом ЛПР. Основою даного методу є представлення процесу рішення у вигляді багаторівневої ієрархії, яка має відображати всі складові задачі, що вирішується [16-17]. Основу методу складають принципи декомпозиції, парних порівнянь та ієрархічної композиції. Основними етапами методу є побудова ієрархії, оцінювання значимості та пріоритетів, перевірка узгодженості пріоритетів та синтез рішення.

Виклад матеріалу дослідження. У роботах [3, 12, 18] автором було введено такі критерії оптимальності структури БД вузла РКІС, як незалежність від центрального вузла БД, співвідношення розміру локальної БД до центральної, та показник рівня необхідності синхронізації даних. Рівень незалежності, або доступності даних визначався за

$$F_{\text{доступність}} = \frac{\sum_{i=1}^n F_{\text{доступність}}(Q_n)}{n}, \quad (1)$$

де $Q_n \in Q_{\text{node}}, Q_{\text{node}} = \{Q | F_{\text{приналежності}}(\text{Node}, Q) = 1\}$,

$$F_{\text{доступність}}(\text{Node}, Q) = \begin{cases} 1, \text{ якщо } \forall R'' \exists R_{\text{schema}}^{\text{remote}}, R'' \in R_{\text{schema}}^{\text{remote}} \cap \forall Q^{\text{inner}} F_{\text{доступність}}(Q^{\text{inner}}) = 1 \\ 0, \text{ якщо } \exists R'' \nexists R_{\text{schema}}^{\text{remote}}, R'' \in R_{\text{schema}}^{\text{remote}} \cup \exists Q^{\text{inner}} F_{\text{доступність}}(Q^{\text{inner}}) = 0 \end{cases}$$

Значення критерію розміру БД вузла визначається як

$$F_{\text{size}} = \frac{\sum_{i=1}^n \text{Size}R''_i}{\sum_{i=1}^n \text{Size}R_i^{\text{DBMS}}}, \quad (2)$$

де $\text{Size}R'' = \text{coef}_{\text{size}R} \times p \times \sum_{i=1}^n \text{Size}(\text{type}_i)$,

$$\text{coef}_{\text{size}R} = \frac{\text{Size}R_i^{\text{DBMS}} - \text{Size}R_0^{\text{DBMS}}}{p \times \sum_{i=1}^n \text{Size}(\text{type}_i)}$$

Показник рівня необхідності синхронізації даних розраховується, як

$$F_{synchro} = \frac{P_{node}^{modif}}{P_{node}}, \quad (3)$$

де P_{node}^{modif} – кардинальність відношення, що включає запити віддаленого вузла (згідно рішення про представленість), до яких входять значення атрибутів кортежів (комірки), що також входять до множини

R_{node}^{modif} , а P_{node} – кардинальність всіх запитів, атрибути та кортежі яких представлені у віддаленому вузлі.

При складанні ієрархії використано відношення між елементами сусідніх рівнів: ціль – зацікавлені особи – критерії – альтернативи. При відсутності можливості проведення аналізу із залученням декількох ЛПР, рівень ієрархії може бути зменшено до трьох: ціль – критерії – альтернативи (рис. 1).

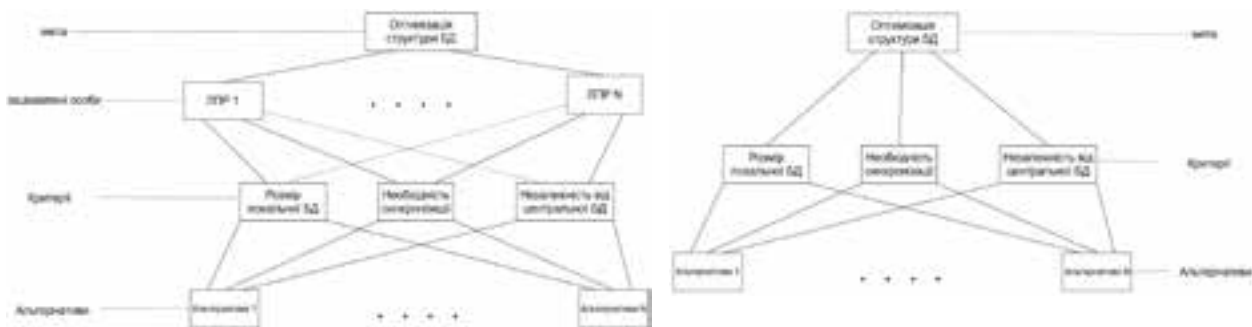


Рис. 1 – Загальне представлення ієрархії моделі для 4-х та 3-х рівнів відповідно

Значення маркеру представленості даних (альтернатива) є дійсним числом на інтервалі [-1, 1]. А наявність великої кількості альтернатив на 4-му рівні ієрархії веде до збільшення розмірності матриць переваг за критеріями та ускладнює вирішення задачі. Запропоновано спростити задачу, виконавши зменшення кількості аль-

тернатив до 5: «низький» (Н) – «-1», «нижче середнього» (НС) – «-0,5», «середній» (С) – «0», «вище середнього» (ВС) – «0,5», та «високий» (В) – «1». Рівень «зацікавлені особи» представлено елементами «Власник», «Розробник БД та КІС», «Адміністратор БД», «Розробник БД» та «Оператор КІС». Отримана ієрархічна модель наводиться на рис. 2.

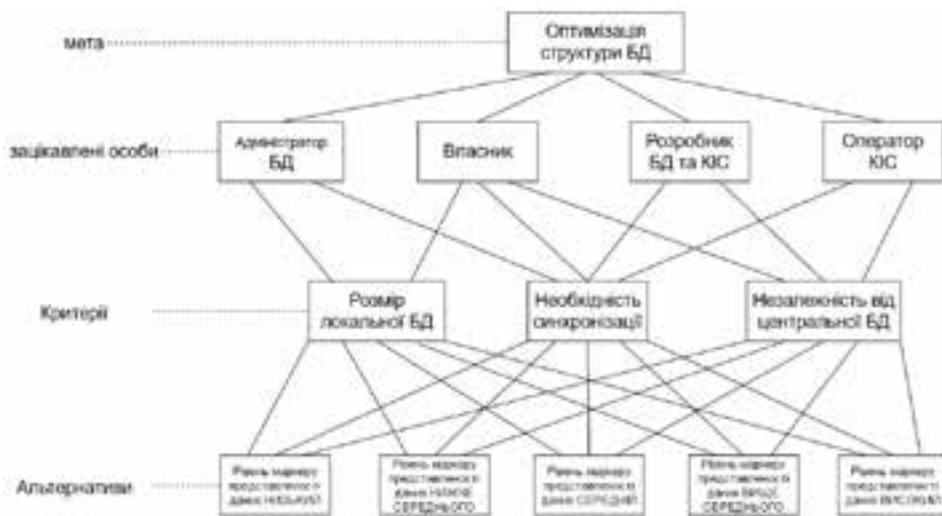


Рис. 2 – Ієрархічна модель задачі оптимізації структури вузла РКІС

Звернімо увагу, що перелік критеріїв оптимальності моделі відрізняється для зацікавленої особи проведенного аналізу. Так, для власника об'єкту автоматизації важливі всі три критерія (розмір БД, необхідність синхронізації та незалежність від центральної БД), оскільки від них залежить як якість роботи КІС, так і вартість обладнання. Для адміністратора БД важливими є критерії розміру БД та необхідності організації синхронізації даних. У свою чергу, для розробника БД та оператора КІС критерій розміру БД не є критичним. Зрозуміло, що відносна вага кожного з критеріїв для різних зацікавлених осіб також буде відрізнятися.

Використовуючи шкалу відносної важливості критеріїв [15] (табл. 1) та із залученням людини, що приймає рішення (ЛПР), якою на даному етапі виступає власник об'єкту автоматизації, виконуємо побудову матриці попарних порівнянь для зацікавлених осіб.

Таблиця 1 – Шкала відносної важливості критеріїв

a_j	Відносна вага критерію або альтернативи
1	Рівна можливість порівнюваних критеріїв (альтернатив)
3	Помірна (слабка) перевага одного над іншим
5	Сильна (істотна) перевага
7	Очевидна перевага
9	Абсолютна перевага
2, 4, 6, 8	Проміжні значення

Так, наприклад, для випадку очевидної переваги власнику над оператором КІС, сильної переваги над роз-

робником БД та помірної переваги над адміністратором БД; слабкої переваги адміністратора над розробником БД та помірної переваги над оператором КІС; а також помірної переваги розробника БД над оператором, маємо матрицю переваг наступного вигляду (табл. 2).

На третьому рівні ієрархії складаються відповідні матриці попарних порівнянь за критеріями оптимальності по кожній зацікавленій особі. Тут у ролі ЛПР виступає зацікавлена особа, за якою заповнюється матриця. Так, для зацікавленої особи «власник», у випадку очевидної переваги критерію незалежності над розміром БД та сильної переваги над необхідністю синхронізації, а також помірної переваги необхідності синхронізації над розміром БД, маємо наступну матрицю переваг критеріїв оптимальності (табл. 3).

Виконавши розрахунок середньої ваги за формулою Баклі [19], що базується на побудові середнього геометричного

$$W_j = \frac{r_j}{\sum_{i=1}^n r_i}, \text{ де } r_j = \left(\prod_{i=1}^m \alpha_{i,j} \right)^{\frac{1}{n}} \quad (4)$$

отримуємо для матриці переваг ЛПР «власник» наступний вектор відносної ваги критеріїв оптимальності моделі (5):

$$W^{крит} = \begin{bmatrix} 0,08 \\ 0,73 \\ 0,19 \end{bmatrix}. \quad (5)$$

Для перевірки відсутності конфліктів між елементами матриці, виконується розрахунок індексу

Таблиця 2 – Матриця переваг зацікавлених осіб

	Власник	Адмін. БД	Розробник БД та КІС	Оператор
КІС				
Власник	1	3	5	7
Адмін. БД	1/3	1	3	5
Розробник БД та КІС	1/5	1/3	1	3
Оператор КІС	1/7	1/5	1/3	1

Таблиця 3 – Матриця переваг критеріїв оптимальності для зацікавленої особи «Власник»

	Розмір БД	Незалежність	Необхідність синхронізації
Розмір БД	1	1/7	1/3
Незалежність	7	1	5
Необхідність синхронізації	3	1/5	1

узгодженості. Для даних таблиці 3 $U = 3,2\%$, що свідчить про допустимий рівень узгодженості (у випадку перевищення значення 10% виникає необхідність корегування значень таблиці через повторне опитування ЛПР).

Наступним кроком у класичному методі аналізу ієрархій є заповнення матриць попарних порівнянь альтернатив окремо по кожному критерію оптимальності [16], аналогічно до табл. 2 та 3. У нашому випадку,

наявність математичних моделей розрахунку значень критеріїв оптимальності, сформульованих у (1-3), дозволяє виконати розрахунок та початкову ініціалізацію даних матриць на базі числових значень маркеру представленості даних для кожної альтернативи. Далі заповнена матриця подається на розгляд ЛПР для затвердження. Так, наприклад, розмір БД локального вузла в залежності від однієї з п'яти альтернатив може змінюватись наступним чином (табл. 4).

Таблиця 4 – Залежність значення розміру БД від обраної альтернативи

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

Виходячи із наведених даних, розмір БД при низькому (min) та високому (max) рівні маркеру представленості даних відрізняється у $\frac{0,75}{0,02} = 37,5$ разів.

Керуючись принципами парних порівнянь та аксіомою гомогенності, яка говорить про те, що на кожному рівні ієрархії порівнювані елементи не мають сильно відрізнятися один від одного (згідно табл. 1 не більше, ніж у 9 разів), виконуємо нормування значень, наведених у табл. 4, використавши дещо модифіковану формулу природньої нормалізації:

$$W_i^{norm} = \frac{(W_i - \min_i W_i)}{(\max_i W_i - \min_i W_i)} \times (k - 1) + 1, \quad (6)$$

де W_i – значення критерію оптимальності для i -ї альтернативи, а $k = 9$, згідно табл. 1.

Пронормовані згідно (6) значення розміру локальної БД (табл. 4) представлені у таблиці 5.

Виконавши округлення до цілого за математичними правилами, будемо матрицю попарних порівнянь альтернатив для критерію розміру локальної БД (табл. 6).

Таблиця 5 – Пронормовані значення розміру БД від обраної альтернативи

Рівень маркеру представленості даних	Низький	Нижче середнього	Середній	Вище середнього	Високий
Пронормоване значення розміру локальної БД	1,00	3,41	5,93	6,81	9,00

Таблиця 6 – Матриця переваг альтернатив по критерію розмір БД

Розмір локальної БД	Низький	Нижче середнього	Середній	Вище середнього	Високий
Низький	1	3	6	7	9
Нижче середнього	0,33333	1	2	2	3
Середній	0,16667	0,5	1	1	2
Вище середнього	0,14286	0,5	1	1	1
Високий	0,11111	0,33333	0,5	1	1

Аналогічно до розрахунку матриці відносної ваги критеріїв оптимальності моделі, згідно до (4) виконуємо розрахунок матриці відносної ваги альтернатив за критерієм розміру локальної БД. Отриманий результат для даних табл. 6 матиме наступний вигляд:

$$W^{розмБД} = \begin{bmatrix} 0,570 \\ 0,190 \\ 0,095 \\ 0,081 \\ 0,063 \end{bmatrix}. \quad (7)$$

Далі, у табл. 7 наведено значення критеріїв необхідності синхронізації даних та рівня незалежності для кожної з п'яти альтернатив, отримані із багатовимірної БД користувачьких запитів.

Відповідно до (6) виконуємо нормування та складаємо матрицю попарних порівнянь альтернатив для критеріїв незалежності (табл. 8) та необхідності синхро-

нізації (табл. 9). При розрахунку коефіцієнтів для альтернатив по критерію незалежності, враховуємо максимізацію критерію, як мету, і, відповідно (6) набуває наступного вигляду:

$$W_i^{norm} = \frac{(W_i^{\sim} - \min_i W_i^{\sim})}{(\max_i W_i^{\sim} - \min_i W_i^{\sim})} \times (k - 1) + 1, \\ \text{де } W_i^{\sim} = (1 - W_i)$$

Згідно до (4) виконуємо розрахунок матриці відносної ваги альтернатив за критеріями незалежності та необхідності синхронізації. Отриманий результат для даних табл. 8 та 9 представлений у (8) та (9) відповідно.

$$W^{незалежн} = \begin{bmatrix} 0,036 \\ 0,082 \\ 0,164 \\ 0,328 \\ 0,328 \end{bmatrix}. \quad (8)$$

Таблиця 7 – Залежність значення критеріїв рівня незалежності та необхідності у синхронізації від обраної альтернативи

Рівень маркеру представленості даних	Низький	Нижче середнього	Середній	Вище середнього	Високий
Незалежність	0,35	0,71	0,92	0,96	0,97
Необхідність синхронізації	0,15	0,18	0,13	0,10	0,07

Таблиця 8 – Матриця переваг альтернатив по критерію незалежності від центрального вузла БД

Незалежність	Низький	Нижче середнього	Середній	Вище середнього	Високий
Низький	1	0,5	0,2	0,11111	0,11111
Нижче середнього	2	1	0,5	0,25	0,25
Середній	5	2	1	0,5	0,5
Вище середнього	9	4	2	1	1
Високий	9	4	2	1	1

Таблиця 9 – Матриця переваг альтернатив по критерію необхідності синхронізації

Необхідність синхронізації	Низький	Нижче середнього	Середній	Вище середнього	Високий
Низький	1	1	1	0,5	0,14286
Нижче середнього	1	1	0,5	0,33333	0,11111
Середній	1	2	1	0,5	0,2
Вище середнього	2	3	2	1	0,33333
Високий	7	9	5	3	1

$$W^{сінхр} = \begin{bmatrix} 0,082 \\ 0,063 \\ 0,101 \\ 0,184 \\ 0,550 \end{bmatrix} \quad (9)$$

Глобальні пріоритети для альтернатив визначаємо за формулою

$$W_i = W_1^{крит} \times W_i^{розмБД} + W_2^{крит} \times W_i^{незалежн} + W_3^{крит} \times W_i^{сінхр} \quad (10)$$

Отже, виходячи із (5), (7), (8) та (10), і відповідно до (10) маємо значення глобальних пріоритетів альтернатив, наведені у табл. 10 та на (11).

Таблиця 10 – Розрахунок глобальних пріоритетів альтернатив для зацікавленої особи «Власник»

Альтернативи (рівень маркету представленості)	Критерії			Глобальні пріоритети
	Розмір БД	Незалежність	Необхідність синхронізації	
	0,08	0,73	0,19	
Низький	0,5701	0,0365	0,0824	0,0883
Нижче середнього	0,1900	0,0821	0,0629	0,0872
Середній	0,0950	0,1642	0,1012	0,1468
Вище середнього	0,0814	0,3285	0,1843	0,2813
Високий	0,0633	0,3285	0,5497	0,3487

$$W_{Власник}^{глоб} = \begin{bmatrix} 0,09 \\ 0,09 \\ 0,15 \\ 0,30 \\ 0,37 \end{bmatrix} \quad (11)$$

Склавши матриці попарних порівнянь критеріїв оптимальності для інших зацікавлених осіб (табл. 11–13), згідно даних математичних моделей (1–3) та відповідно до (6) виконуємо розрахунки векторів глобальних пріоритетів за вказаними особами (12–14).

$$W_{АдмінБД}^{глоб} = \begin{bmatrix} 0,41 \\ 0,15 \\ 0,10 \\ 0,12 \\ 0,23 \end{bmatrix} \quad (12)$$

$$W_{РозробникБД}^{глоб} = \begin{bmatrix} 0,04 \\ 0,08 \\ 0,17 \\ 0,33 \\ 0,38 \end{bmatrix} \quad (13)$$

Таблиця 11 – Матриця переваг критеріїв оптимальності для зацікавленої особи «Адміністратор БД»

	Розмір БД	Необхідність синхронізації
Розмір БД	1	2
Необхідність синхронізації	1/2	1

Таблиця 12 – Матриця переваг критеріїв оптимальності для зацікавленої особи «Розробник БД»

	Незалежність	Необхідність синхронізації
Незалежність	1	7
Необхідність синхронізації	1/7	1

Таблиця 13 – Матриця переваг критеріїв оптимальності для зацікавленої особи «Оператор КІС»

	Незалежність	Необхідність синхронізації
Незалежність	1	3
Необхідність синхронізації	1/3	1

$$W_{\text{ОператорКІС}}^{\text{глоб}} = \begin{bmatrix} 0,05 \\ 0,08 \\ 0,16 \\ 0,31 \\ 0,40 \end{bmatrix} \quad (14)$$

Використавши отримані результати векторів глобальних пріоритетів за зацікавленими особами (11-14) та матрицю переваг зацікавлених осіб (табл. 2), аналогічно до (10) розраховуємо вектор глобальних пріоритетів альтернатив.

Таблиця 14 – Розрахунок глобальних пріоритетів альтернатив

	Власник	Оператор КІС	Адмін.БД	Розробник БД	Глобальні пріоритети
	0,563	0,055	0,263	0,117	
Низький	0,0900	0,0500	0,4100	0,0400	0,17
Нижче середнього	0,0900	0,0800	0,1500	0,0800	0,10
Середній	0,1500	0,1600	0,1000	0,1700	0,14
Вище середнього	0,3000	0,3100	0,1200	0,3300	0,26
Високий	0,3700	0,4000	0,2300	0,3800	0,34

Висновки. Кожне значення рівня маркету представленості даних характеризується показниками їх привабливості для ЛПР, а саме таким критеріями оптимальності, як незалежність від центрального вузла БД, розмір локальної БД, та показник рівня необхідності синхронізації даних.

Задача є добре структурованою, оскільки всі залежності можуть бути представлені у чисельному вигляді, а критерії незалежні. Однак кращі поєднання критеріїв не можуть бути визначені на основі об'єктивної інформації, наявної в розпорядженні дослідника, а наявність декількох критеріїв призводить до того, що основою для вироблення рішення мають стати суб'єктивні переваги ЛПР. Рішення хоч і буде суб'єктивним, але в процесі вирішення будуть використані об'єктивні моделі на базі багатокрите-

ріальних методів, що дозволяє ЛПР уникнути поверхневих рішень та схильності до спрощення задачі.

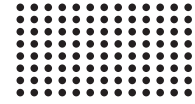
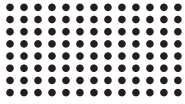
Вирішення багатокритеріальної задачі та знаходження оптимального рівня представленості даних на віддаленому вузлі дозволяє підвищити рівень доступності даних та ефективності використання розподілених та територіально розосереджених комп'ютерних систем. Ефективністю є співвідношення результату та ресурсів, тому враховуючи вектор відносної ваги критеріїв оптимальності моделі (5), розраховуємо ефективність, як

$$Eff = \frac{F_{\text{доступність}} \times W_1^{\text{крит}}}{F_{\text{size}} \times W_0^{\text{крит}} + F_{\text{synchro}} \times W_2^{\text{крит}}} \quad (15)$$

Порівняння отриманих результатів для БД вузла КІС предметної області наведено у табл. 15.

Таблиця 15 – Порівняння ефективності структури БД при різних стратегіях та рівнях представленості даних на вузлі РКІС

		Викор-ня центрального вузла	Предст-ня лише критичних даних	Предст-ня всіх необхідних даних	Повне дублювання даних	Оптимальний рівень маркера предст-ті
Незалежність	0,730	0	0,35	0,97	1	0,97
Розмір БД	0,081	0	0,02	0,75	1	0,63
Необхідність синхронізації	0,188	0	0,15	0,07	1	0,08
Ефективність БД вузла		–	8,5589	9,5892	2,7126	10,7257
Підвищення ефективності, %		–	25,32%	11,85%	295,41%	–



Отже, результати дослідження дозволяють підвищити ефективність використання певного вузла РКІС розглянутою предметною областю на 25% порівняно із представленням лише критичних даних, та на 11% порівняно із представленням всіх необхідних даних центральної БД відповідно. В якості перспектив подальшого розвитку роз-

глядається можливість представлення отриманого вектору глобальних пріоритетів у вигляді набору нечітких множин однієї змінної із кусочно-лінійними функціями приналежності для подальшої дефазифікації отриманих результатів та отримання більш точного числового значення оптимального рівня маркеру представленості даних на вузлі РКІС.

СПИСОК ЛІТЕРАТУРИ:

1. Иванов А.Ю. Мобильные распределенные базы данных втоматизированных информационно-управляющих систем МЧС России : монография / подред. В.С. Артамонова. СПб. : Санкт-Петербургский ун-т ГПС МЧС России, 2008. 152 с.
2. Новиков Б.А. Основы технологий баз данных : учеб. пособие / Б.А. Новиков, Е.А. Горшкова; под ред. Е.В. Погова. – М. : ДМК Пресс, 2019. – 240 с.
3. Mykhailo Dvoretzkyi, Svitlana Dvoretzka, Yuriy Nezdoliy, Svitlana Borovlova. Data Utility Assessment while Optimizing the Structure and Minimizing the Volume of a Distributed Database Node. Proceedings of the 1st International Workshop on Information-Communication Technologies & Embedded Systems (ICTES 2019) Mykolaiv, Ukraine, November 14–15, 2019. P. 128–137.
4. Пасічник В.В. Організація баз даних та знань/ Пасічник В.В., Резніченко В.А. – К. : Видавнича група ВНУ, 2006. – 384 с.
5. Малахов Є.В. Основи проектування баз даних : навч. посіб. для студ. вищих навч. закладів / Є.В. Малахов. – О. : Наука і техніка, 2006. – 156 с.
6. Пасічник В.В. Сховища даних: підручник/ Пасічник В.В. Шаховська Н.Б. – Львів : Магнолія 2006, 2008. – 492 с.
7. Кунгурцев А.Б., Возовиков Ю.Н. Поиск закономерностей в распределении запросов для управления материализованными представлениями // Тр. Одесск. политехн. ун-та. Одесса, 2008. – 2(30). – С. 135–140.
8. Кунгурцев А.Б. Модель реструктуризации реляционной базы данных путем денормализации схемы отношений / А.Б. Кунгурцев, С.Л. Зиноватная // Труды Одесского политехнического университета. – Одесса : Вид-во Одесский национальный политехнический университет. – 2006. – 2(26). – С. 105–111.
9. Филатов В.А. Методы и средства проектирования информационных систем и распределенных баз данных / В.А. Филатов, Р.В. Семенец // Вестник Херсонского национального технического университета No 4(27). – 2007. – С. 203–207.
10. Лаздынь С.В. Оптимизация распределенных корпоративных информационных сетей с использованием генетических алгоритмов и объектного моделирования / С.В. Лаздынь, С.Ю. Землянская // Наукові праці ДонНТУ. – 2009. – No 147. – С. 83–95.
11. Дворецький М. Л. Проектування структури розподіленої БД на базі парсингу SQL-запитів / М.Л. Дворецький, С.Ю. Боровльова, Є.О. Давиденко // Наукові праці [Чорноморського державного університету імені Петра Могили комплексу «Києво-Могилянська академія»]. Серія: Комп'ютерні технології. – 2016. – Т. 287. – Вип. 275. – С. 53–61.
12. Дворецький М. Л. Інформаційна технологія визначення корисних даних при оптимізації структури та мінімізації обсягів вузла розподіленої бд. / М. Л. Дворецький, С. В. Дворецька, Є. О. Давиденко // Вісник черкаського державного технологічного університету. – 4/2019. – Черкаси : Черкаський державний технологічний університет, 2019. – С. 26–35.
13. Коваленко І.І. Парето-оптимальний вибір при формуванні портфеля замовлень ІТ-проектів / І. І. Коваленко, Є. О. Давиденко // Наукові праці : науково-методичний журнал. – Вип. 161. – Т. 173. Комп'ютерні технології. – Миколаїв : Вид-во ЧДУ ім. П. Могили, 2011. – С. 44–48.
14. Подиновский В.В., Ногин В. Д. Парето-оптимальные решения многокритериальных задач. – М. : Наука. Главная редакция физико-математической литературы, 1982. – 256 с.
15. А.В. Лотов, И.И. Поспелова. Многокритериальные задачи принятия решений : учебное пособие. – М. : МАКС Пресс, 2008. – 197 с.
16. Ногин В.Д. Принятие решений в многокритериальной среде: количественный подход. – М. : ФИЗМАТЛИТ, 2002. – 144 с.
17. Ногин В.Д. Методы оптимальных решений [Учебное пособие]. – СПб. : филиал ГУ-ВШЭ, Издательство «Юстас», 2006. – 108 с.
18. Fisun M., Dvoretzkyi M., Shved A. and Davydenko Ye. “Query parsing in order to optimize distributed DB structure”, Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Bucharest, Romania, vol. 1, pp. 172–178, September 2017. DOI: 10.1109/idaacs.2017.8095071
19. Buckley J.J. Fuzzy hierarchical analysis // Fuzzy sets and systems. 1985. Vol. 17, No 3. P. 233–247.



USING THE METHOD OF ANALYZING HIERARCHIES TO OPTIMIZE THE DATABASE STRUCTURE OF THE NODE IN CORPORATE INFORMATION SYSTEMS

Mykola Fisun,

Dr. Sc., Professor, Professor, Associate Professor of IPS Department,
Mykolaiv National University named by P. Mohyla, Mykolaiv, Ukraine,
e-mail: mykola.fisun@gmail.com, 0000-0003-1297-6230

Mykhailo Dvoretzkyi,

lecturer of IPS department,
Mykolaiv National University named by P. Mohyla, Mykolaiv, Ukraine,
e-mail: m.dvoretzkiy@gmail.com, 0000-0001-5913-6859

Svitlana Dvoretzka,

Lecturer of IPS Department, Mykolaiv National University named by P. Mohyla, Mykolaiv, Ukraine,
e-mail: svetag603@gmail.com, 0000-0001-5199-9430

Yevhen Davydenko,

Ph.D., Associate Professor, Associate Professor of IPS Department,
Mykolaiv National University named by P. Mohyla, Mykolaiv, Ukraine,
e-mail: genik.davydenko@gmail.com, 0000-0002-0547-3689

Abstract. The goal of the article. The aim of the article is to solve the problem of multicriteria optimization of the structure of a node database in corporate information systems using the method of hierarchy analysis to increase the level of data availability and the efficiency of using distributed and geographically dispersed computer systems.

Research methods. Research methods are based on the principles of the relational databases theory (including distributed ones) and methods of multicriteria optimization. In particular, on: methods of relational algebra, theory of a relational data model, multidimensional data model – when building a model of an SQL query; the method of analyzing hierarchies – when solving the problem of choosing the best alternative to the level of data representation.

Main results of research. The concept of a data representation marker is introduced, reflecting the level of need for data presentation at the DCIS node. For each element, the value of the marker is taken as one of the set of values {necessary, preferably, not necessary}. It determines the degree to which it is necessary to present data for a particular type of workplace, user role or application. The formulated optimality criteria for the structure of the database allow to speak about the problem of multi-criteria optimization. Its' solution is carried out using the analytic hierarchy process. The analytic hierarchy process (AHP), which is a general methodology for solving a wide class of decision-making problems, allows to combine a relatively simple mathematical apparatus with the knowledge and experience of decision makers. A 4-level hierarchical model of the following type is proposed: goal – stakeholders – criteria of optimality – alternatives. Among the features of using the method, it should be noted: a different set of optimality criteria for various stakeholders; splitting into intervals the data representation marker with making 5 alternatives and automatic pre-filling of the matrix of pairwise comparisons at the last level of the hierarchy.

The scientific novelty. For the first time, the concept of a marker of data representation at the DCIS node was introduced for the dimension elements of the SQL query model. An aggregation function was developed, which allows determining the level of need for attributes and tuples of the database relation on the DCIS node based on the statistics of SQL queries. The analytic hierarchy process was further developed due to the automatic initialization of the matrix of alternatives pairwise comparisons according to the obtained mathematical models and the normalization of values and the presentation of the result in the form of a vector of fuzzy numbers.

The practical significance. Solving the problem of multicriteria analysis and choosing the best alternative allows to determine the optimal level of the value of the data representation marker. In turn, it allows to classify the attributes and tuples of DB relations according to their presentation on the DCIS node. The calculations performed for one of the subject areas of



the research results application, were given in the corresponding research tables. They allow to speak of an increase in the efficiency of the DB of the DCIS node by 25% compared to the presentation of only critical data, and by 11% compared to the presentation of all necessary data of DB respectively.

Keywords: *distributed DBMS, computer information system, multicriteria optimization, relational data model, SQL query, hierarchy analysis method, matrix of pairwise comparisons, priority vector, data representation marker.*

ИСПОЛЬЗОВАНИЕ МЕТОДА АНАЛИЗА ИЕРАРХИЙ ДЛЯ ОПТИМИЗАЦИИ СТРУКТУРЫ БАЗЫ ДАННЫХ УЗЛА В КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Николай Фисун,

доктор технических наук, профессор кафедры инженерии программного обеспечения,
Черноморский национальный университет имени Перта Могилы, г. Николаев, Украина,
e-mail: mykola.fisun@gmail.com, 0000-0003-1297-6230

Михаил Дворецкий,

ст. преподаватель кафедры инженерии программного обеспечения,
Черноморский национальный университет имени Перта Могилы, г. Николаев, Украина,
e-mail: m.dvoretskiy@gmail.com, 0000-0001-5913-6859

Светлана Дворецкая,

ст. преподаватель кафедры инженерии программного обеспечения,
Черноморский национальный университет имени Перта Могилы, г. Николаев, Украина,
e-mail: svetag603@gmail.com, 0000-0001-5199-9430

Евгений Давиденко,

кандидат технических наук, доцент кафедры инженерии программного обеспечения,
Черноморский национальный университет имени Перта Могилы, г. Николаев, Украина,
e-mail: genik.davydenko@gmail.com, 0000-0002-0547-3689

Цель статьи. Целью статьи является решение задачи многокритериальной оптимизации структуры базы данных узла в корпоративных информационных системах с использованием метода анализа иерархий для повышения уровня доступности данных и эффективности использования распределенных и территориально рассредоточенных компьютерных систем.

Методы исследования. Методы исследований базируются на основных принципах теории реляционных баз данных (в том числе распределенных) и методах многокритериальной оптимизации. В частности, на: методах реляционной алгебры, теории реляционной модели данных, многомерной модели данных – при построении модели SQL-запроса; методе анализа иерархий – при решении задачи выбора лучшей альтернативы уровня представленности данных.

Основные результаты исследования. Введено понятие маркера представленности данных, отражающее уровень необходимости представления данных на узле РКИС. Для каждого элемента значение маркера принимается за одно из множества значений {необходимо, желательно, не нужно}, что определяет степень необходимости представления данных того или иного типа рабочего места, роли пользователя или приложения. Сформулированные критерии оптимальности структуры БД позволяют говорить о задаче многокритериальной оптимизации, решение которой выполняется с использованием метода анализа иерархий. метод анализа иерархий (МАИ), что является общей методологией решения широкого класса задач принятия решений, позволяет совместить сравнительно простой математический аппарат со знаниями и опытом ЛПП. Предложена 4-х уровневая иерархическая модель следующего вида: цель – заинтересованные лица – критерии оптимальности – альтернативы. Среди особенностей использования метода следует отметить: разный набор критериев оптимальности для различных заинтересованных лиц; разбиение



на интервалы непрерывного признака маркера представленности данных с выделением 5 альтернатив и автоматическое предварительное заполнение матрицы попарных сравнений на последнем уровне иерархии.

Научная новизна. Впервые введено понятие маркера представленности данных на узле РКИС для элементов измерений модели SQL-запроса и разработана функция агрегации, что позволяет определить уровень необходимости атрибутов и кортежей отношения БД на узле РКИС на основе статистики SQL-запросов; получил дальнейшее развитие метод анализа иерархий за счет автоматической инициализации матрицы попарных сравнений альтернатив согласно полученных математических моделей и нормализации значений и представление результата в виде вектора нечетких чисел с приведением к четкому значению.

Практическая значимость. Решение задачи многокритериального анализа и выбор наилучшей альтернативы позволяет определить оптимальный уровень значения маркера представленности данных, в свою очередь позволяет классифицировать атрибуты и кортежи отношений БД согласно их представлению на узле РКИС. Проведенные расчеты для одной из предметных областей применения результатов исследования, результаты которых приведены в соответствующих таблицах исследования, позволяют говорить о повышении эффективности БД узла РКИС на 25% по сравнению с представлением только критических данных, и на 11% по сравнению с представлением всех необходимых данных центральной БД соответственно.

Ключевые слова: *распределенная СУБД, корпоративная информационная система, многокритериальная оптимизация, реляционная модель данных, SQL-запрос, метод анализа иерархий, матрица попарных сравнений, вектор приоритетов, маркер представленности данных.*