

4. Xian, L., Bi, S., Quan, Z., Wang, H. (2021). Online Cognitive Data Sensing and Processing Optimization in Energy-harvesting Edge Computing Systems. URL: <https://arxiv.org/pdf/2106.14113.pdf>

5. Capponi, A., Fiandrino, C., Kliazovich, D., Bouvry, P., Giordano, S. (2017). A Cost-Effective Distributed Framework for Data Collection in Cloud-Based Mobile Crowd Sensing Architectures. DOI: 10.1109/TSUSC.2017.2666043/.

**УДК 004.657:004.33(043.2)**

*Зінченко Б. В., здобувач вищої освіти 2 курсу ОС «Магістр» спеціальності 122 Комп'ютерні науки,*

*Римар П. В., старший викладач кафедри інформаційних технологій*

## **ОПТИМІЗАЦІЯ БАЗИ ДАНИХ ІЗ ДОПОМОГОЮ ВИКОРИСТАННЯ COLUMNSTORE-ІНДЕКСІВ**

*Донецький національний університет імені Василя Стуса, м. Вінниця*

У сучасному світі обробка великої кількості даних у реальному часі стала ключовим аспектом багатьох бізнес-процесів. Оптимізація баз даних стає необхідністю для забезпечення швидкодії та ефективності систем обробки інформації. У цьому контексті використання Columnstore-індексів виявляється перспективним напрямом для оптимізації роботи баз даних [1].

У процесі дослідження буде проведений порівняльний аналіз продуктивності баз даних із використанням Columnstore-індексів та традиційні реляційні індекси. В роботі розглядаються конкретні сценарії, – як-от вибірка даних за певними умовами, сортування, агрегація та з'єднання таблиць. Для ілюстрації порівнянь виконаємо кілька запитів на обидві системи та проаналізуємо час виконання. Одним із методів цього дослідження є ретельний аналіз та порівняння Columnstore-індексів із традиційними реляційними базами даних із погляду ефективності, швидкодії та витрат ресурсів. Розглянемо види оптимізації [2].

*Оптимізація запитів:* цей вид оптимізації спрямований на поліпшення виконання SQL-запитів, містить у собі використання індексів, оптимізацію структури запиту, аналіз та розробку планів виконання.

*Оптимізація структури бази даних:* орієнтована на вдосконалення самої структури бази даних. Це може містити нормалізацію таблиць, вибір оптимальних типів даних, розподіл даних між таблицями для зменшення дублювання інформації.

*Оптимізація індексів:* забезпечує ефективність використання індексів для швидкодії вибірки даних, містить у собі створення правильних індексів, їх видалення або переструктуризацію для підвищення продуктивності.

*Оптимізація запитів із використанням кешування:* зменшення навантаження на базу даних шляхом зберігання результатів, часто використовуваних запитів у кеші. Це допомагає швидше давати відповіді на однакові або схожі запити.

*Оптимізація конфігурації бази даних:* встановлення оптимальних параметрів та налаштувань бази даних, як-от розмір буферного пулу, параметри кешування, налаштування паралельного виконання запитів та ін.

*Оптимізація завдань та процесів:* вдосконалення та оптимізація розкладу й виконання завдань, як-от індексація, резервне копіювання, планування оптимізованих завдань обслуговування бази даних.

*Вибірка даних за певними умовами.* Розглянемо випадок, коли компанія має базу даних клієнтів і хоче витягти інформацію про клієнтів, які здійснили покупки на суму більше 1 000 доларів протягом останнього місяця. Виконаємо цей запит на обидві системи та порівняємо час виконання [3].

```
SELECT CustomerName, PurchaseAmount
FROM Customers
WHERE PurchaseAmount > 1000 AND PurchaseDate >= DATEADD(MONTH, -1, GETDATE());
```

*Рисунок 1. Програмний код запиту*

Після виконання запитів було порівняно час виконання на обох системах. Як і очікувалось, запит на базі даних із Columnstore-індексами виконається швидше через їх здатність ефективно обробляти великі об'єми даних [4, 5].

На підставі проведеного дослідження можна визначити конкретні області, де використання Columnstore-індексів демонструє значні покращення продуктивності. Наприклад, у великих таблицях, де велика кількість колонок, та в операціях агрегації чи аналітичних запитах, де необхідно опрацювати великий обсяг даних, Columnstore-індекси можуть значно прискорити вибірку та обробку інформації.

### **Список використаних джерел**

1. Farber L., West M., (2015). Columnstore index: SQL Server's answer to a columnar storage format. Microsoft White Paper.
2. Graefe, G. (2016). Data Management for General Data Analytics. *ACM SIGMOD Record*, 45(4), 27–32.
3. Baklarz E., Zikopoulos P., (2015). Big Data and Analytics: Infonomics and the Business of Managing Data. IBM Press.
4. Kandiraju G., Shegalov G., Bilke A., (2013). SQL Server Columnstore Index. Apress.

5. Nikolic B., Ivanovic M., Milošević D. (2017). A Comparative Study of Columnar Storage and Indexing Techniques in Relational Database Management Systems. *Journal of Computer Science and Technology*, 17(4), 407–416.

#### **УДК 004.62**

*Колосова К. К., здобувачка I курсу спеціальності 122 Комп'ютерні науки,  
Огороднік М. О., здобувачка I курсу спеціальності 122 Комп'ютерні науки,  
Ніколюк П. К., д-р фіз.-мат. наук, професор, професор кафедри інформаційних технологій*

### **ПОРІВНЯННЯ РІЗНИХ МЕТОДІВ ЗНАХОДЖЕННЯ ЗНАЧЕННЯ КОРЕЛЯЦІЇ МІЖ НАБОРАМИ ЗМІННИХ**

*Донецький національний університет імені Василя Стуса, м. Вінниця*

В основі багатьох кількісних досліджень лежить встановлення зв'язку між даними, а саме знаходження кореляції для розуміння, як пов'язані дві чи більше змінних. Дослідники зауважують, що кореляція не вказує на причинно-наслідкові зв'язки, особливо, коли цей критерій застосовується до перехресних даних. Припущення про вимірювання елементів тільки однією факторною величиною призводить до можливого завищення величини кореляції між змінними. Це обмеження викликало появу методів кореляції другого покоління, заснованих на моделюванні структурних рівнянь, як-от підтверджувальний факторний аналіз та дослідницьке моделювання структурних рівнянь. Точна оцінка кореляції між змінними має вирішальне значення для достовірності результатів дослідження. Коли дослідники отримують точні результати кореляції, це гарантує, що зв'язки між змінними представлені якомога ближче до реальності. Ця точність важлива, оскільки вона дає змогу дослідникам робити обґрунтовані та значимі висновки на основі своїх даних.

Якщо результати кореляції неточні або необ'єктивні, це може призвести до неправильної інтерпретації зв'язків між змінними. Наприклад, якщо кореляція завищена, це може свідчити про сильніший зв'язок між змінними, ніж існує насправді, що є результатом неправильних висновків і потенційно помилкових втручань, заснованих на помилкових доказах. У кількісних дослідженнях у сферах освіти та психології дослідники використовують різні методи для оцінки кореляції між змінними. Деякі приклади цих методів містять двофакторну кореляцію, підтверджувальний факторний аналіз та моделювання на основі дослідницьких структурних рівнянь.