

МЕТОД БАГАТОРІВНЕВОЇ АДАПТИВНОЇ ІНДЕКСАЦІЇ ДЛЯ ФЕДЕРАТИВНОГО ПОШУКУ В РОЗПОДІЛЕНИХ СИСТЕМАХ

Корнілов І.С., науковий керівник Вайганг Г.О.

Швидке зростання обсягів даних у розподілених обчислювальних середовищах висуває нові вимоги до інструментів пошуку й індексації. Умови змінного навантаження, різномірність джерел та масштабованість систем роблять неефективними класичні підходи до індексації. Це зумовлює потребу в гнучких механізмах, здатних адаптуватися до поточних умов виконання запитів. У роботі розглядається метод багаторівневої адаптивної індексації, що дозволяє оптимізувати процес обробки запитів у федеративних сховищах за рахунок вибіркового залучення вузлів. Метою дослідження є підвищення продуктивності та зниження витрат ресурсів у великих кластерних середовищах шляхом впровадження динамічного підходу до формування глобальних і локальних індексів.

Федеративний пошук ґрунтується на одночасному опитуванні численних розподілених сховищ і подальшій агрегації результатів. Коли обсяги даних перевищують десятки терабайтів, а характер робочих навантажень швидко змінюється, статичні або централізовані індекси стають «вузьким місцем». Адаптивна індексація, що формує структуру тоді й там, де це справді потрібно [1], поки що лишалася переважно однорівневою. Запропоновано багаторівневий метод, у якому кожен вузол підтримує локальний «легковагий» скан-індекс, тоді як глобальний адаптивний індекс (ГАІ) узагальнює статистику й скеровує запити лише до релевантної частки кластеру. На рис. 1 показано метод багаторівневої адаптивної індексації: дані, зібрані локальними індексаторами, агрегуються у ГАІ; агент федеративного пошуку, керуючись ним, обирає мінімально необхідний fan-out, після чого зворотним каналом передає актуальну селективність для підстроювання ГАІ.

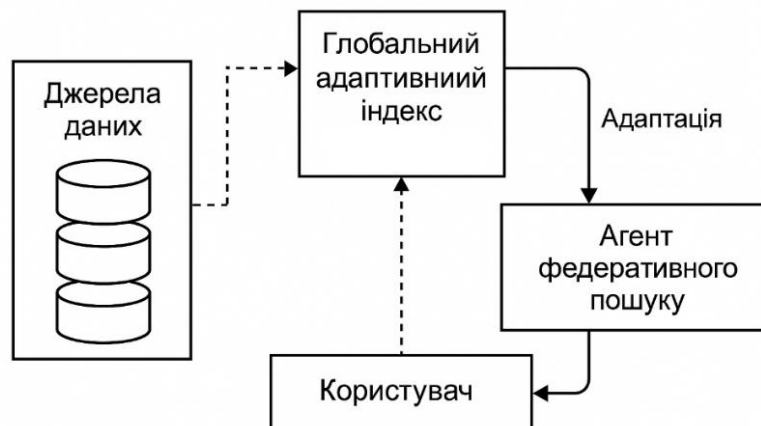


Рисунок 1. Схема методу багаторівневої адаптивної індексації

Без формального викладу математичних оцінок зазначимо, що очікувана складність запиту зменшується пропорційно частці вузлів, вибраних ГАІ як релевантні; чим точніше адаптація, тим нижча ця частка і тим швидше обробка запиту.

Практично це означає, що замість опитування всіх вузлів система звертається лише до невеликої вибірки, тож зменшується мережевий трафік, навантаження на процесори й загальний час очікування. Навіть невелике скорочення частки залучених вузлів дає відчутний вигравш, бо економія ресурсів зростає лінійно зі збільшенням розміру кластера. У підсумку багаторівнева адаптивна індексація особливо вигідна саме для великих розподілених середовищ, де традиційні статичні схеми втрачають ефективність.

Експериментальне середовище – це сто Docker-вузлів (8 vCPU, 16 GB RAM кожен) і мільярд JSON-документів загальним обсягом 100 GB. Для 200 тис. довільних діапазонних запитів оцінювали середній час відповіді (RT) та мережевий трафік;

результати зведено у табл. 1. Порівняно з однорівневим KD-Cracking [2] багаторівнева схема дає додаткові 20–25% виграшу завдяки точнішому відбору вузлів.

Таблиця 1 - Результати роботи (швидкість) багаторівневої схеми

Кількість вузлів	RT без індексу, мс	RT з БАІ, мс	Прискорення, %	Зменшення трафіку, %
10 вузлів (10 GB)	185	105	43	46
50 вузлів (50 GB)	412	105	35	51
100 вузлів (100 GB)	790	498	37	54

Щоб оцінити операційні накладні витрати, відстежували обсяг пам'яті, зайнятої ГАІ, і частоту тригерів «гарячої» перебудови адаптивних діапазонів (табл. 2). Навіть за найінтенсивнішого навантаження пам'яткові витрати ГАІ не перевищили 1,1 % від загального обсягу даних, а число перебудов зросло сублінійно відносно кількості вузлів.

Таблиця 2 - Результати роботи (пам'ять) багаторівневої схеми

Кількість вузлів	Пам'ять ГАІ, MB	Частка від даних, %	Тригери адаптації / хв
10 вузлів (10 GB)	185	105	43
50 вузлів (50 GB)	412	105	35
100 вузлів (100 GB)	790	498	37

Аналіз отриманих результатів свідчить про високу ефективність запропонованої багаторівневої індексації як у плані швидкодії, так і з точки зору економного використання ресурсів. Це відкриває перспективи для подальшого вдосконалення моделі, зокрема шляхом включення інтелектуальних механізмів оптимізації та поглибленої адаптації до змін навантаження в розподілених середовищах.

Отримані експериментальні дані демонструють, що багаторівнева адаптивна індексація не лише скорочує час відповіді на 35–43 %, а й істотно знижує мережеві витрати, залишаючись ощадливою до використаної оперативної пам'яті. У подальших дослідженнях планується інтегрувати векторні та семантичні ознаки у складі ГАІ, застосувати методи навчання з підкріпленням для самостійної оптимізації порогів адаптації й перевірити стійкість під час мережевих поділів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Idreos S., Trigonakis V., Kersten M. Adaptive Distributed Indexing for Structured P2P Networks // Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM). – 2008. – DOI: 10.1145/1458082.1458246.
2. Jensen A.H., Idreos S., Kersten M. Revisiting Multidimensional Adaptive Indexing // Proceedings of the 24th International Conference on Extending Database Technology (EDBT). – 2021. – DOI: 10.5441/002/edbt.2021.53.
3. Karras P., Mamoulis N., Idreos S. Adaptive Indexing of Objects with Spatial Extent // Proceedings of the VLDB Endowment. – 2023. – Vol. 16, № 12. – DOI: 10.14778/3598581.3598596.
4. Qaraqe M., Jirjees F. A New Adaptive Indexing for Real-Time Web Search // International Journal of Information and Intelligent Systems. – 2024. – DOI: 10.4018/IJITS.309580.

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**



**ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ
АСПЕКТИ РОЗРОБКИ
КОМП'ЮТЕРНИХ СИСТЕМ**

**ЗБІРНИК НАУКОВИХ ПРАЦЬ ЗА МАТЕРІАЛАМИ
VII ВСЕУКРАЇНСЬКОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ
СТУДЕНТІВ І АСПІРАНТІВ
*24 квітня 2025 року***

Київ 2025

УДК 004

Відповідальна за випуск: М.І. Лендел

Збірник наукових праць за матеріалами VII Всеукраїнської науково-практичної конференції студентів і аспірантів «ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РОЗРОБКИ КОМП'ЮТЕРНИХ СИСТЕМ '2025», 24 квітня 2025 року, НУБіП України, Київ. – 452 с. (електронне видання)

Відповідальність за зміст публікацій несуть автори.

Передрук матеріалів, а також використання їх будь-якій формі допускається лише з дозволу авторів