**О. А. Тригуб**

студентка 2-го курсу навчальної групи ІК-31 кафедри інформаційних систем та технологій

Національний технічний університет КПІ ім. Ігоря Сікорського, Україна

**В.А. Нікітін**

доктор філософії, асистент кафедри інформаційних систем та технологій

Національний технічний університет КПІ ім. Ігоря Сікорського, Україна

**Є.В. Крилов**

к.т.н., доцент кафедри інформаційних систем та технологій

Національний технічний університет КПІ ім. Ігоря Сікорського, Україна

**Вступ**

 Розподілені бази даних (РБД) є ключовим елементом сучасних інформаційних систем, забезпечуючи високу доступність і масштабованість даних для широкого спектру застосувань, від фінансових платформ до масштабних соціальних мереж. Однак однією з головних проблем таких систем є забезпечення узгодженості даних. Оскільки інформація розподілена між великою кількістю вузлів, можуть виникати ситуації, коли збережені дані відрізняються в різних місцях. Ця проблема є критичною для точності та достовірності інформації, особливо у випадку збоїв у мережі або виходу з ладу окремих компонентів системи.

**Підходи узгодження даних**

 У сфері розподілених баз даних використовуються різні методи узгодження даних, і вибір конкретного підходу залежить від можливостей і вимог системи. Модель ACID (atomicity, consistency, isolation, durability - атомарність, узгодженість, ізоляція, довговічність), визначає набір характеристик, які гарантують коректне виконання транзакцій в реляційних базах даних. Альтернативою ACID в розподілених системах є підхід BASE (basic availability, flexible consistency and fault tolerance - базова доступність, гнучка узгодженість і відмовостійкість), який допускає тимчасові неузгодженості в даних, але гарантує високу доступність. Концепція PACELC базується на затримці та узгодженості за відсутності розривів в мережі. Вона розширює теорему CAP шляхом введення компромісу.

 Теорема CAP є фундаментальним принципом в області розподілених обчислень. Вона стверджує, що жодна система не може одночасно гарантувати три ключові властивості: узгодженість, доступність і толерантність до розділів. Це означає, що у випадку розриву зв'язку система повинна вибирати між забезпеченням доступності даних або забезпеченням їхньої узгодженості.

**Різновиди узгодженості даних в інформаційних системах**

 Є декілька ключових підходів до досягнення узгодженості даних у таких

системах:

 Сувора узгодженість означає, що всі вузли в системі завжди містять однакові дані, надаючи користувачам доступ до тієї ж інформації незалежно від того, до якого вузла вони підключаються. Однак такий підхід вимагає значних ресурсів і часто призводить до затримок у виконанні транзакцій.

 Послідовна узгодженість означає, що зміни даних вносяться в чітко визначеному порядку. Хоча ця система не забезпечує абсолютної узгодженості на всіх вузлах, вона гарантує, що операції застосовуються послідовно. Такий підхід забезпечує певну гнучкість у взаємодії між вузлами та підвищує продуктивність системи.

 Причинно-наслідкова узгодженість гарантує правильний порядок виконання взаємозалежних операцій. Водночас вона зменшує навантаження на систему, гарантуючи коректність лише в певних випадках, оскільки не всі операції потребують суворої синхронізації.

 Кінцева узгодженість означає, що всі вузли синхронізуються з часом, навіть якщо в системі виникають тимчасові неузгодженості даних. Цей підхід корисний, коли доступність даних є більш пріоритетною за миттєву узгодженість.

**Методи забезпечення кінцевої узгодженості**

 Деякі з найсучасніших методів забезпечення остаточної узгодженості

включають наступні:

 Пріоритетний час транзакцій покращує узгодженість у високонавантажених системах, забезпечуючи більш ефективну обробку запитів та зменшуючи затримки. Пріоритет транзакцій визначається на основі часу, щоб допомогти оптимізувати розподіл ресурсів. Наприклад, транзакції з низьким пріоритетом можуть бути викладені, якщо система перевантажена, таким чином забезпечуючи швидку обробку критично важливих запитів.

 Механізм транзакційного годинника заснований на об'єднанні декількох транзакцій в одну підсумкову транзакцію і надсиланні її на вузли розподіленої бази даних. Кількість операцій запису значно зменшується, оскільки транзакції, пов'язані з оновленням запису, об'єднуються в одну транзакцію. У випадку операції видалення всі інші транзакції, пов'язані з цим записом, знищуються і виконується лише операція видалення.

 Активний антиентропійний механізм з використанням модифікованого фільтра Блума та алгоритму хешування PH-2: цей метод орієнтований на швидке відновлення узгодженості в розподілених системах, де важливим аспектом є підтримка високої доступності даних в умовах можливих неузгодженостей. Алгоритм хешування PH-2 та фільтр Блума зменшують споживання пам'яті, забезпечують ефективну перевірку належності елемента до реплікованих даних та допомагають пришвидшити процес синхронізації.

 Алгоритм хешування PH-2 використовує математичну базу простих чисел, які можуть ефективно генерувати хеш-значення. Такий підхід гарантує високу толерантність до колізій, оскільки виключає виникнення колізій при обробці даних різного розміру. Крім того, використання РН-2 значно прискорює процес зіставлення даних, особливо в умовах високих навантажень і при використанні gossip-протоколів.

**Експериментальні результати**

 Експериментальні дослідження на прототипі розробленої інформаційної системи показали, що запропонований метод є більш ефективним при узгодженні даних порівняно з традиційними підходами. Використання розподіленої системи з годинником транзакцій та механізмом пріоритезації дозволило значно скоротити час синхронізації даних при високому навантаженні. Крім того, активні антиентропійні механізми сприяли швидкому відновленню узгодженості після відмови вузла системи.

**Висновки**

 Розподілені бази даних потребують особливих підходів для забезпечення узгодженості, а запропоновані методи демонструють значні переваги з точки зору підвищення їх ефективності та надійності. Використання цих методів дозволяє досягти оптимального балансу між доступністю та узгодженістю даних, що є важливим фактором для сучасних інформаційних систем. Крім того, комбінація розроблених методів покращує продуктивність системи в цілому за різних умов експлуатації та забезпечує гнучкість і стійкість до високих навантажень і збоїв у мережі.

Список використаних джерел

* Nikitin V.A. Методи підвищення ефективності узгодженості даних в інформаційних системах. Дисертація на здобуття ступеня доктора філософії. Національний технічний університет України «КПІ ім. Ігоря Сікорського», 2024.
* Nikitin V., Krylov E. Active anti-entropy mechanism based on Spectral Bloom filter and PH-2 hash algorithm for reconciliation of replicas of NoSQL distributed document-oriented databases. Information Technology and Society, №3(9), 2023, 63–67. **DOI:** <https://doi.org/10.32689/maup.it.2023.3.8>
* Nikitin V., Krylov E. Comparison of hashing methods for supporting consistency in distributed databases. Adaptive Systems of Automatic Control, №1(40), 2022, 48–53. **DOI:** <https://doi.org/10.20535/1560-8956.40.2022.261646>