

## МЕТОД СИНХРОНІЗАЦІЇ БАЗ ДАНИХ ЗА ДОПОМОГОЮ ХЕШ-ФУНКЦІЙ НА ОСНОВІ НЕЛІНІЙНОГО РОЗБИТТЯ ТАБЛИЦЬ

Мельник А.М.<sup>1)</sup>, Гарах О.О.<sup>2)</sup>

<sup>1)</sup> Тернопільський національний економічний університет, викладач кафедри комп'ютерних наук

<sup>2)</sup> Тернопільський національний економічний університет, магістр

### І. Постановка проблеми

З появою інформаційних систем з розподіленою архітектурою з'явилася проблема підтримки синхронного стану даних на різних вузлах. На вирішення цієї проблеми направлені системи реплікації [1]. В роботі [2] представлено два способи виявлення змін: реплікація поточного стану та дельта-реплікація. Реплікація поточного стану визначає такий спосіб реплікації, при якому постачальник даних ще до початку обміну даними знає, які дані необхідно направити клієнту. При такому способі реплікації відбувається значне навантаження на мережу, яка дуже часто впливає на загальну ефективність цього процесу. Для вирішення цієї проблеми можна застосувати механізм хеш-функцій. Якщо хеш-функції повертають різні значення для двох вхідних записів, то ці записи є різними, а якщо хеш-значення не співпадають, то можна стверджувати, що і записи співпадають.

### II. Мета роботи

Метою дослідження є реалізація методу синхронізації баз даних за допомогою хеш-функцій на основі нелінійного розбиття таблиць.

### III. Особливості реалізації алгоритму синхронізації баз даних за допомогою хеш-функцій на основі нелінійного розбиття таблиць

Базу даних розглядаємо як лінійну структуру, де операції здійснюються із записами повністю, тобто або запис передається на віддалений вузол або не передається. Таблицю розбиваємо на деякі інтервали, для кожного з яких розраховуємо значення хеш-функції. Для інтервалів в яких хеш-значення на вузлі відправнику і на вузлі одержувачу не співпадають – здійснюємо подрібнення інтервалів і повторюємо вказані операції. На практиці виникає необхідність налаштування синхронізованих процесів між базами даних, які не проектувалися для використання в розподілених системах і для яких неможливо змінити існуючу схему. Часто такі бази даних не були відповідним чином нормалізовані, що призводить до збільшення довжини відповідних записів. Таким чином, при зміні одного значення розміром в один біт, описаний алгоритм вимагатиме передавання всього запису або набору записів. Розглянемо метод, який може вирішити цю проблему.

Оскільки реляційна база даних оперує із таблицями, і таблиця не є одновимірним представленням як файл, а двовимірним, то отримаємо наступний алгоритм: 1) визначаємо множину загальних сутностей на вузлі відправнику і на вузлі одержувачі (сутності можуть бути різними, але повинні мати однакові ідентифікатори); 2) розбиваємо множину загальних сутностей на області заданої величини. Область характеризується множиною ідентифікаторів записів і множиною номерів колонок; 3) розраховуємо значення хеш-функції для заданих областей і передаємо їх значення віддалений вузол; 4) розраховуємо значення хеш-функції на віддаленому вузлі для відповідних областей і порівнюємо їх із значеннями, які отримані на вузлі відправнику; 5) області, які відрізняються одна від одної за показниками, що отримані на попередньому кроці, розбиваємо і для нових областей повторюємо третій крок; 6) цей процес повторюємо доки розмір розбитої області не буде дорівнювати одній колонці або одному запису або не досягне деякого порогового значення.

### Висновок

У роботі запропоновано метод, що дозволяє ефективно здійснювати синхронізацію погано спроектованих баз даних без модифікації їх структури. Описаний алгоритм порівняння реалізований на основі хеш-функцій.

### Список використаних джерел

1. Белоусов В.Е. Алгоритмы репликации данных в распределенных системах обработки информации. Диссертация на соискание ученой степени кандидата технических наук: 05.13.01 / Белоусов Всеволод Евгеньевич – Пенза, 2005 – 184 с.
2. Tridgell A. The Rsync Algorithm. Technical Report TR-CS-96-05 / A. Tridgell, P. Mackerras // Department of Computer Science, The Australian National University, Canberra, Australia, 1996.
3. Рябков Н.С. Аналитический обзор методов репликации и синхронизации баз данных / Н.С. Рябков // Качество, инновации, образование. – 2006. – №4. – С. 56–63.