

Міністерство освіти і науки України  
Харківський національний університет імені В.Н. Каразіна

На правах рукопису

Жолткевич Галина Григоріївна

УДК 004.042/519.713.2

Математичні імітаційні моделі для забезпечення узгодженості  
для розподілених сховищ даних

01.05.02 — Математичне моделювання та обчислювальні методи

Дисертація на здобуття наукового ступеня  
кандидата технічних наук

Науковий керівник  
Рукас Кирило Маркович,  
доктор технічних наук, доцент

Харків — 2019

## ЗМІСТ

Вступ	3
Розділ 1. Балансування узгодженості	5

## ВСТУП

Актуальність теми. В наше сьогоднішнє інноваційні технології з'являються дуже швидко, а існуючі розвиваються з неймовірною швидкістю. Гіперлуп, багаточисленні дослідження космосу, наукові роботи в інших галузях, таких, як медицина, зелені мережі, а також, більш побутові, але все ще такі потрібні технології, такі, як комунікації, транспорт, розумні будинки... Не можна нехтувати тим фактом, що всі ці системи потребують більшої гнучкості, швидкості, надійності та засобів для зберігання інформації також надійно та швидко і доступно, а інколи навіть доступність має бути майже у будь-якій точці земної кулі. Тому одним з найважливіших компонентів для багатьох таких систем є швидке і надійне розподілене сховище. В 21 столітті термін "розподілене сховище" становиться вже звичним. Деякі сховища збільшують кількість вузлів, деякі - ні. Причиною цього є те, що багато з таких систем потребують сильної узгодженості даних. Але якщо збільшувати кількість вузлів для сховища, консистентність падає дуже швидко. А для деяких систем це важлива частина для їх стабільної роботи.

Бо є дуже відома CAP-теорема, яка стверджує, що неможливо одночасно задовільнити всі три характеристики для сховища, узгодженість (consistency), доступність (availability), стійкість до розділення (partition tolerance).

Ми не збираємося оскаржувати цю теорему, але робимо спробу обійти цю проблему. Механізм для цього і буде темою для цієї роботи.

Взаємозв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана згідно з планом науково-дослідницьких робіт Харківського національного університету імені В.Н. Каразіна в рамках теми "Математическое и компьютерное

моделирование информационных процессов в сложных естественных и технических системах" (номер государственной регистрации 0112U002098).

Мета і завдання дослідження. Метою роботи являється побудова імітаційних та математичних моделей для механізму підтримки сильної узгодженості у розподілених сховищах даних, проведення експериментів, оцінювання складності імітаційних моделей, побудова метрик, за якими можна дослідити складність даних моделей, а також розробка обчислювальних методів для сформованого механізму. Це дозволить оцінити, наскільки можна розширити будь-яку розподілену систему і сформує методи для коректної роботи за такими умовами.

Для досягнення цієї мети у роботі розв'язані наступні задачі:

## РОЗДІЛ 1

### БАЛАНСУВАННЯ УЗГОДЖЕНОСТІ

Наскільки нам відомо, що відповідно до CAP-теореми можна задовільнити тільки будь-які дві з трьох характеристик для розподіленого сховища даних. В цьому розділі розглядається можливість досягнути компромісу та забезпечити консистентні відповіді від бази даних, не втрачаючи доступності бази.

Ми розуміємо, що консистентність

In this chapter we make attempt to reach compromise and ensure or increase consistency value for any distributed datastore without losing availability and partition tolerance. We would like to introduce the idea first and then estimate its impact availability and partition tolerance.

So let us go deeper into a problem.

Imagine we have a distributed datastore that has  $N$  nodes. We do not take into account for now role of each node (master or slave) and assume that each node can accept read and write requests. Now we focus on the mechanism of database request accepting and processing. If one of nodes accepted write request, we claim that the datastore system will reach fast enough of consistency value to maintain consistent response.

What if to try to get round this problem by handling request to database by the list of consistent nodes for. appropriate to the dataunit incoming in the request. We want to estimate how it may impact availability and partition tolerance if we want to get round consistency issue this way.

There are several solutions. But they all have general architecture: (Picture of architecture)