

ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

доктора технічних наук Глибовця Андрія Миколайовича на дисертаційну роботу Жолткевич Галини Григоріївни “Математичне моделювання процесів реплікації даних у розподілених сховищах”, що представлена на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.02 — математичне моделювання та обчислювальні методи (технічні науки)

Актуальність теми дисертації. Процеси глобалізації, що є трендами сучасного етапу розвитку людства, не оминули, звичайно і галузі інформаційно-комунікаційних технологій (ІКТ), більш того, сучасні ІКТ створили технологічні умови для розвитку цих процесів. З точки зору інформатики, глобалізація пов’язана з переходом до масового використання глобально розподілених систем таких, як хмарні сервіси, соціальні мережі, глобальні пошукові системи тощо. Таким чином, проблема побудови науково обґрунтованих методів проектування розподілених інформаційних систем стала сьогодні одним з основних напрямків дослідження у галузі комп’ютерних наук і інформаційно-комунікаційних технологій. Безперечно, ядром розподіленої інформаційної системи є її сховище (або база) даних, яка також має будуватися за розподіленою архітектурою. Зважаючи на те, що при розподілені даних неможливо виконати принцип “кожна одиниця даних зберігається у сховищі тільки в одному примірнику, виникає ймовірність отримання різними користувачами не однакових реплік (примірників) однієї і тої самої одиниці даних. Тому архітектурне рішення для розподіленого сховища даних (РСД) мало б забезпечити узгодженість (consistency) реплік, їх доступність (availability) з будь-якого вузла мережі та стійкість до сховища до розділення мережі (partition tolerance). На ці вимоги зазвичай посилаються як на САР-гарантії. Однак пionерські роботи Е. Брюера, Н. Лінч та С. Гілберта початку цього сторіччя показали, що забезпечити безумовне виконання всіх САР-гарантій є неможливим. Ці результати спричинили інтенсивні дослідження цієї проблеми, зважаючи не тільки на її теоретичну важливість, але й принципову практичну значущість. Не зважаючи на те, що цій темі за даними Semantic

Scholar за останні п'ять років присвячено понад сорок тисяч публікацій, проблема й досі залишається відкритою.

Таким чином, дисертаційна робота Жолткевич Г.Г. "Математичне моделювання процесів реплікації даних у розподілених сховищах", що присвячена розв'язанню науково-технічної задачі забезпечення контролюваного рівня узгодженості реплік даних в розподіленому сховищі даних при необхідних рівнях доступності даних і стійкості сховища даних до розділення на етапі проектування розподіленої системи задля управління ризиками некоректного функціонування системи шляхом побудови комплексу математичних моделей об'єкту дослідження та розробки обчислювальних методів аналізу цих моделей, є безперечно актуальною.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалася на кафедрі теоретичної та прикладної інформатики факультету математики і інформатики Харківського національного університету імені В.Н. Каразіна відповідно до завдань прикладної НДР «Моделі інформаційних процесів та методи їх обробки» ДР № 0116U003141.

Ступінь обґрунтованості і достовірності наукових положень, висновків і рекомендацій, сформульованих у дисертації. Обґрунтованість викладених у дисертаційній роботі положень обґрунтовується коректністю постановки як загальної мети, так специфічних завдань дослідження, що підтверджується: використанням методів теорії графів та теорії ймовірностей, розробленням програмних засобів імітаційного моделювання РСД з використанням мови UML, проведением і аналізом результатів імітаційних експериментів, достатньою кількістю публікацій у наукових фахових виданнях України, міжнародних виданнях, апробацією результатів дослідження на міжнародних наукових конференціях.

Оцінка змісту дисертації та її завершеність. Матеріал дисертації викладено у традиційному для таких робіт стилі, викладення є логічним і завершеним. Дисертація включає анотацію українською та англійською мовами, вступ, чотири розділи, висновки, список використаних джерел з 102 найменувань на 11

сторінках та 7 додатків на 13 сторінках, а також містить 54 рисунка. Загальний обсяг роботи складає 138 сторінок, включаючи 110 сторінок основного тексту. Структура дисертації є вдалою та логічно скомпонованою, оформлення роботи відповідає діючим вимогам.

Зміст і результати дисертації відповідають паспорту спеціальності 01.05.02 — математичне моделювання та обчислювальні методи (технічні науки).

Матеріали дисертації достатньо повно викладено у 8 наукових працях, три з яких — у наукових фахових виданнях України, 1 — в іноземному спеціалізованому виданні (наукометрична база «Scopus») і 4 тези доповідей на фахових міжнародних наукових конференціях.

Автореферат точно відповідає змістові дисертації, причому його зміст дає повне уявлення про наукову цінність і практичну значущість дослідження.

У **вступі** наведена загальна характеристика роботи, а саме: актуальність теми та аналіз пов'язаних з темою робіт, предмет, мету і задачі дослідження, об'єкт дослідження, наукову новизна та практичне значення отриманих результатів, а також наведений список публікацій за темою дисертаційної роботи.

У **першому розділі** подана детальна характеристика стану досліджень у предметній області дисертації, обґрунтовано мету та задачі дослідження, актуальність дослідження в цілому.

Другий розділ присвячений побудові математичних моделей для опису мережі розподіленої системи, централізованої моделі зберігання даних та її узагальнення для випадку розподіленого сховища даних. Всі ці моделі узгоджені між собою, а їх сукупність забезпечує повноту представлення об'єкту в контексті мети дослідження. В розділі описано також метод реплікації даних у РСД у вигляді відповідного алгоритму.

У **третьому розділі** введені ймовірнісні оцінки рівня (метрики) виконання САР-гарантій. Певним чином, цей розділ є ключовим, оскільки в ньому запропонований підхід до розуміння розв'язку проблеми забезпечення САР-гарантій, оскільки традиційний підхід у відповідності до САР-теореми до розв'язку призвести не може. Таким чином, поставлена задача вирішується через

забезпечення контролюваного відхилення рівнів виконання гарантій від ситуації повного задоволення вимог за рахунок балансування цих рівнів за допомогою розробленого і описаного в цьому розділі методу.

Четвертий розділ містить опис програмних реалізацій імітаційних моделей, що використовуються для експериментальних досліджень, а також в ньому наведений аналіз результатів цих досліджень.

Нові наукові результати дисертаційної роботи. У дисертаційній роботі отримані наступні нові наукові результати:

- зазнала подальшого розвитку стохастична модель розподіленої системи за рахунок забезпечення формалізації виконання САР-гарантій, яка, на відміну від існуючих моделей, об'єднує формалізовані визначення для САР-гарантій в одну цілісну структуру і дає підставу для оцінки гарантій на іншому рівні;
- вперше запропоновано систему статистичних метрик для оцінювання виконання САР-гарантій у РСД, що базується на формалізації визначень САР-гарантій, і замість бінарної оцінки гарантій РСД пропонується оцінювати систему гарантій за допомогою ймовірнісних величин, що дає підставу для розвитку метрик для оцінки рівня САР-гарантій;
- удосконалено методи обчислення ймовірностей втрати даних на основі запропонованої моделі, що дає можливість оцінки САР-гарантій і формування метрик для оцінки з більшою точністю, з врахуванням реальних умов таких систем, і, таким чином, це дає підставу для формування методу регулювання САР-характеристик на рівні, потрібному конкретній розподіленій системі;
- зазнали подальшого розвитку методи забезпечення САР-вимог шляхом розробки нової стратегії контролюваного балансування узгодженості у формі спеціального методу контролюваного балансування узгодженості, що може бути інтегрований з методами реплікації у РСД.

Практична цінність отриманих результатів. На основі запропонованих методів та моделей реалізовано метод регулювання САР-гарантій у формі алгоритму для покриття такого рівня САР-гарантій, які будуть необхідні

конкретному РСД. Цей метод може бути застосований на етапі проектування системи РСД для запобігання помилок і неоптимальних рішень побудови розподіленої мережі, а також невірного вибору або застосування методів забезпечення САР-гарантій, тобто зменшення ризиків, які коштуватимуть дорожче на етапі застосування такої системи. Також такий метод містить елементи, які мають працювати на етапі застосування системи, а саме регулювання рівня узгодженості з метою утримування її на необхідному рівні.

Результати дисертаційної роботи були використані при підготовці відповідних спеціальних освітніх компонентів за двома програмами у Запорізькому національному і Херсонському державному університетах.

Зауваження щодо дисертаційної роботи.

1. Доречно було б надати більш детальний аналіз існуючих РСД, особливо від лідерів галузі Google, Amazon, MicroSoft, IBM та Oracle.
2. Доречно було б у другому розділі більш детально спинитися на обґрунтуванні запропонованої моделі сховища даних у вигляді словника, спираючись на так званні документарні сховища.
3. В розділі 3.5 необхідно було звернути увагу на HDFS та NameNode, що дуже схожа на модель, що запропонована автором
4. Цікавим було б порівняння запропонованого методу балансування метрик рівня виконання САР-гарантій з методами, що використовуються при багатокритеріальній оптимізації, або в задачі задоволення обмежень.
5. На стор 90 зазначено "З іншого боку, важко уявити систему, у якій запити на запис на одну й ту саму одиницю надходять настільки часто". В системах IoT, прилад може передавати декілька разів на секунду свій стан. Тобто оновлювати одну й ту саму одиницю (зберігати всі попередні стани дуже затратно). А таких приладів в системі налічуються сотні тисяч. Ситуація стає гіршою коли з'являються зв'язками між приладами, та залежність стану підсистеми приладів від стану всіх своїх складових.
6. У четвертому розділі не зважим було б запропонувати мову специфікації для автоматизованої побудови імітаційних моделей сховищ даних, що аналізуються.

7. Доречним було б також навести приклад імплементації моделей, методів, і прототипів програмного забезпечення для конкретної системи управління розподіленими базами даних (технічні науки).

Робота також містить стилістично невдалі вирази, вузько професійний жаргон та друкарські помилки.

Однак ці зауваження не впливають на загальну позитивну оцінку дисертаційної роботи.

Загальний висновок. Наведений вище аналіз дозволяє стверджувати, що дисертація Жолткевич Г.Г. “Математичне моделювання процесів реплікації даних у розподілених сховищах” є завершеною науковою роботою, в якій вирішується актуальна науково-технічна задача забезпечення контролюваного рівня узгодженості реплік даних в розподіленому сховищі даних при необхідних рівнях доступності даних і стійкості сховища даних до розділення.

Дисертація відповідає паспорту спеціальності 01.05.02 — математичне моделювання та обчислювальні методи (технічні науки) та всім вимогам до кандидатських дисертацій згідно з пп. 9, 11, 12 Порядку присудження наукових ступенів, затвердженого Постановою Кабінету міністрів України 24.07.2013 року № 567, а її авторка Жолткевич Галина Григоріївна заслуговує на присудження наукового ступеня кандидата технічних наук за спеціальністю 01.05.02 — математичне моделювання та обчислювальні методи.

Офіційний опонент:

доктор технічних наук, доцент
Глибовець Андрій Миколайович,
декан факультету інформатики
Національного університету
«Києво-Могилянська академія»

А. М. Глибовець
“15” березня 2021 року

