

РЕЦЕНЗІЯ

на дисертаційну роботу

Кривого Ярослава Вячеславовича

на тему «Подієво-орієнтована мікросервісна архітектура масштабованої

платформи скінченно-елементного аналізу»,

представлену на здобуття наукового ступеня доктора філософії

з галузі знань 12 «Інформаційні технології»

за спеціальністю 122 «Комп'ютерні науки»

Актуальність теми.

Роботу присвячено розробленню й експериментальному дослідженню подієво-орієнтованої мікросервісної платформи для виконання задач скінченно-елементного аналізу.

Актуальність теми визначається необхідністю виконання великих пакетів незалежних обчислювальних задач, різними профілями споживання ресурсів на стадіях обчислювального конвеєра та потребою обґрунтованого визначення кількості обчислювальних виконавців.

Особливо важливим є те, що автор досліджує не лише технічну можливість збільшення кількості реплік, а фактичний ефект масштабування: пропускну здатність, розподіл наскрізних затримок з окремим контролем його верхньої частини, тривалість стадій конвеєра, час роботи чисельного розв'язувача та збереження чисельної коректності.

Обґрунтованість мети, завдань і методів.

Мета дослідження – підвищення масштабованості й керованості програмної платформи скінченно-елементного аналізу шляхом розроблення подієво-орієнтованої мікросервісної архітектури, реалізації прототипу повного FEA-конвеєра та експериментального визначення впливу кількості обчислювальних виконавців і складності задачі на продуктивність платформи.

У роботі сформульовано такі завдання дослідження:

1) провести аналіз монолітного та мікросервісного підходів до побудови програмних систем, дослідити існуючі комерційні, хмарні й відкриті FEA-рішення та визначити невирішені архітектурні проблеми створення масштабованих платформ скінченно-елементного аналізу;

2) сформулювати функціональні й нефункціональні вимоги та розробити концептуальну подієво-орієнтовану мікросервісну модель повного FEA-конвеєра з визначенням складу компонентів, меж їхньої відповідальності, контрактів взаємодії та способу розділення керівних подій, транзакційних метаданих і великих обчислювальних артефактів;

3) розробити й реалізувати експериментальний прототип FEA-платформи та її обчислювальний контур за моделлю «споживач подій – незалежно масштабовані виконавці», що забезпечує зміну кількості реплік processor-worker без масштабування інших компонентів і зміни зовнішніх контрактів;

4) реалізувати двовимірну та тривимірну тестові задачі Пуассона із застосуванням методу скінченних елементів і методу виготовлених розв'язків для контролю чисельної коректності платформи;

5) розробити методику експериментального оцінювання продуктивності та масштабованості платформи з урахуванням пропускну здатності, наскрізної затримки, тривалості окремих стадій, внутрішнього часу чисельного розв'язувача, прискорення, ефективності масштабування та контрольних чисельних характеристик;

6) провести серії обчислювальних експериментів при різній кількості обчислювальних виконавців і різній складності скінченно-елементних задач, кількісно оцінити ефект горизонтального масштабування, встановити його залежність від обчислювального навантаження та розробити рекомендації щодо вибору кількості виконавців;

7) перевірити збереження контрольних чисельних характеристик при зміні кількості виконавців, оцінити зменшення похибки при згущенні сітки та визначити обмеження і загрози валідності експериментального дослідження.

Мета роботи сформульована коректно. Її досягнення забезпечується послідовним переходом від аналізу предметної області до концептуального проєктування, програмної реалізації та експериментального оцінювання.

Методика дослідження ґрунтується на п'яти гіпотезах. Перші чотири стосуються продуктивності й масштабованості, п'ята – незалежності математичного результату від кількості виконавців.

Експериментальні дані отримано з декількох рівнів системи: часових міток стадій у базі даних, внутрішніх метрик обчислювального компонента, параметрів артефактів і результатів пакетного запуску. Такий багаторівневий збір даних дозволяє відокремити час корисного обчислення від очікування в чергах та інших накладних витрат.

Оцінка основних результатів.

У роботі виконано 36 серій обчислювальних експериментів та змодельовано 1080 варіантів задачі Пуассона. Досліджено двовимірну та тривимірну постановки, два рівні дискретизації та конфігурації з одним, двома й чотирма виконавцями.

Для двовимірних задач перехід до двох виконавців забезпечує приріст пропускної здатності та зменшення 95-го перцентиля наскрізної затримки, однак подальше масштабування до чотирьох виконавців не дає стабільного покращення.

Для тривимірної задачі з грубішою сіткою максимальний ефект також досягається за двох виконавців. Для дрібнішої тривимірної сітки чотири виконавці забезпечують найбільшу абсолютну пропускну здатність, проте ефективність використання додаткових ресурсів знижується.

Отриманий результат має практичне значення: доцільна кількість виконавців залежить не лише від кількості задач, а й від частки корисного чисельного обчислення в повній наскрізній затримці.

Позитивно слід оцінити сформульований автором принцип найменшої достатньої кількості виконавців. Він дозволяє уникати ситуації, коли збільшення

кількості реплік підвищує витрати ресурсів, але не забезпечує пропорційного приросту продуктивності.

Наукова новизна та практична цінність.

Вперше розроблено архітектурну модель подієво-орієнтованої мікросервісної FEA-платформи, у якій повний обчислювальний цикл логічно поділено на окремі мікросервіси, відокремлено канали передавання керівних подій, службових та великих обчислювальних даних, а також визначено правила й формати взаємодії між компонентами системи, що дає змогу незалежно змінювати обчислювальну потужність окремих стадій без перебудови решти платформи.

Вперше для FEA-платформ експериментально встановлено, що розмежування функцій координації подій і виконання ресурсоемних чисельних задач між незалежними компонентами забезпечує відтворюваність чисельних результатів при зміні кількості обчислювальних виконавців, що дає змогу розглядати горизонтальне масштабування як математично нейтральну операцію щодо точності розв'язку.

Отримали подальший розвиток методи оцінювання масштабованості розподілених обчислювальних платформ шляхом включення до складу показників пропускну здатності, наскрізної затримки та її статистичних параметрів, тривалостей окремих стадій конвеєра і часу роботи розв'язувача, що дає змогу комплексно характеризувати ефективність горизонтального масштабування платформи.

Уточнено закономірності впливу обчислювальної складності скінченно-елементної задачі на ефективність горизонтального масштабування розподіленої FEA-платформи: встановлено непропорційний характер масштабування та визначено частку часу чисельного розв'язання в повній наскрізній затримці як визначальний критерій вибору кількості обчислювальних виконавців, що дає змогу обґрунтовано планувати ресурси платформи та уникати надлишкового виділення обчислювальних потужностей.

Практична цінність полягає в можливості використання запропонованої методики для планування ресурсів розподілених платформ чисельного моделювання.

Публікації та апробація.

Результати, що становлять основний зміст дисертаційного дослідження, відображено у трьох наукових публікаціях у фахових виданнях України за спеціальністю 122 «Комп'ютерні науки». У них представлено результати розроблення архітектури платформи, її програмної реалізації та експериментального дослідження характеристик масштабування. Окремі положення роботи було представлено на науковій конференції та обговорено під час міжкафедральних семінарів. Тематика й зміст опублікованих праць відповідають основним положенням дисертації, а наведені відомості свідчать про достатній рівень апробації отриманих результатів.

Дискусійні питання.

1. Ефективність масштабування оцінюється переважно за кількістю виконавців. Доцільно було б також враховувати вартість масштабування, наприклад, співвідношення приросту пропускної здатності до додаткових витрат CPU та пам'яті.

2. Незалежне масштабування реалізовано лише для етапу чисельного розв'язання. Водночас для великих моделей побудова сітки також може бути ресурсоємною. Доцільно дослідити можливість масштабування препроцесингу та визначити умови, за яких саме він стає вузьким місцем системи.

3. Платформа виконує серії однотипних розрахунків і зберігає всі проміжні та кінцеві артефакти. Чи розглядався механізм кешування та повторного використання результатів попередніх етапів для зменшення обчислювальних витрат?

4. Експериментальна оцінка проведена на досить обмеженому наборі задач. Для підтвердження узагальнення отриманих результатів доцільно було б використати ширший спектр задач різної складності.

5. Для повнішої оцінки масштабованості доцільно дослідити роботу платформи у багатовузловому кластері із залученням GPU та оцінити ефективність гібридної CPU–GPU реалізації з розподілом обчислювальних задач між різними типами ресурсів.

Висловлені зауваження мають рекомендаційний характер і не впливають на загальну позитивну оцінку дослідження.

Висновок.

Дисертаційна робота Кривого Ярослава Вячеславовича є завершеним науковим дослідженням, у якому розроблено, реалізовано та експериментально перевірено подієво-орієнтовану мікросервісну платформу скінченно-елементного аналізу.

Результати мають наукову новизну, є достовірними в межах визначених умов експерименту та мають практичне значення для побудови й ресурсного планування розподілених обчислювальних платформ.

Дисертація відповідає спеціальності 122 «Комп’ютерні науки» та чинним вимогам до робіт на здобуття наукового ступеня доктора філософії. Кривий Ярослав Вячеславович заслуговує на присудження наукового ступеня доктора філософії з галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп’ютерні науки».

РЕЦЕНЗЕНТ:

кандидат фізико-математичних наук,
доцент, професор кафедри програмної інженерії
Запорізького національного університету

Олексій КУДІН