

ВІДГУК

офіційного опонента про дисертаційну роботу
Філімончук Тетяни Володимирівни
«МЕТОДИ ТА ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ РОЗПОДІЛУ
ЗАВДАНЬ В ГЕТЕРОГЕННИХ GRID-СИСТЕМАХ»,
подану на здобуття наукового ступеня кандидата технічних наук
за спеціальністю 05.13.06 – інформаційні технології

Актуальність теми виконаних досліджень

Дисертація Філімончук Т.В. є науково-практичним дослідженням, метою якого є розробка методів та інформаційної технології розподілу завдань в гетерогенних GRID-системах, які дозволяють зменшити час простою обчислювальних ресурсів системи за рахунок скорочення часу виконання вхідного пулу завдань.

В даний час поява нових інформаційних технологій, таких як хмарні обчислення, GRID, гібридні кластерні системи, зумовило якісний розвиток технологій розв'язання задач великої обчислювальної розмірності. До найбільш перспективних напрямків сьогодні можна віднести технології паралельних та розподілених обчислень, які є основою GRID-обчислень.

Важливе місце в GRID-системах при розподілі завдань відводиться планувальникам (брокерам), які формують розклад використання обчислювальних ресурсів. Однак, на сьогодні не існує універсального по ефективності методу розподілу, який розподіляв би будь-які класи задач, враховуючи при цьому всі особливості наявних обчислювальних ресурсів.

Розподіл ресурсів між завданнями, які надходять до GRID-системи – це процес, що включає в себе ряд етапів: визначення множини ресурсів, які відповідають заданим вимогам для вирішення завдання; резервування ресурсів; планування виконання завдань; моніторинг системи.

Для досягнення цих цілей необхідна розробка ефективного методу розподілу ресурсів та планування їх використання, що тягне за собою необхідність багаторазового повторення ряду експериментів. Проведення таких експериментів часто є неможливим через розподіленість та гетерогенність GRID-систем. Тому, моделювання є оптимальним рішенням, яке дозволяє досліджувати GRID-систему та проводити аналіз її поведінки як системи в цілому, так і її окремих частин.

Однак, розподілення завдань на обчислювальні ресурси, особливо в умовах



використання гетерогенних програмних компонентів та ресурсів, потребує використання особливих методів розподілу. Ефективне використання ресурсів значною мірою впливає на вартість розподілених обчислень, тому питання його підвищення є актуальною, особливо у теперішній час бурного росту глобальних обчислювальних систем рівня GRID та хмарних сервісів.

Відзначаючи актуальність теми дисертації, треба також підкреслити її відповідність тематичним координаційним планам держбюджетних науково-дослідних робіт «Розробка структури Харківського ресурсно-операційного GRID-центру та його ресурсів», договір №9 (28.09.07-31.10.07, 01.12.07-31.12.07) між ХНУРЕ і «ІПСА» НТУУ «КПІ», що виконувалася на підставі договору «ІПСА» НТУУ «КПІ» з Міністерством освіти і науки України №ІТ/506-2007, Державної програми «Інформаційні та телекомунікаційні технології в освіті і науці» на 2006-2010 рр. (№ДР 0107U010616) та «Розробка та дослідження застосування GRID-порталу Харківського ресурсно-операційного GRID-центру», договору №08-22 (08.04.07-27.06.08) і №08-22/9 (01.07.08-30.09.08) між ХНУРЕ і «ІПСА» НТУУ «КПІ», що виконувалася на підставі Договору «ІПСА» НТУУ «КПІ» з Міністерством освіти і науки України №ІТ/506-2013, Державної програми «Інформаційні та телекомунікаційні технології в освіті і науці» на 2006-2013 рр. (№ДР 0108U008261).

Структура і склад дисертаційної роботи

Дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та двох додатків.

У *вступі* наведені основні загальні характеристики роботи, такі як актуальність, мета та задачі дослідження, визначені наукова новизна, відомості про апробацію результатів та публікації за темою роботи.

У *першому розділі* наведено аналіз предметної області та розглянуто питання розвитку розподілених систем у якості середовища для розв'язання задач великої обчислювальної складності. Наведено огляд технологій розподілених обчислень і Cloud-технологій, аналіз існуючих планувальників завдань, сформульована мета та задачі дослідження.

На підставі проведеного аналізу, автор виявив ряд недоліків для існуючих планувальників GRID-системи, головний з яких – це жорстка прив'язаність їх до вирішення конкретного класу задач.

У *другому розділі* досліджена існуюча модель розподілу завдань у GRID-

системі. У ході її аналізу було виявлено що, в процесі розподілу завдань на обчислювальні ресурси не враховується ряд параметрів, які можуть підвищити ефективність використання ресурсів GRID-системи (пропускна здатність каналу, затримка часу передачі пакету по каналу). Також під час розподілу завдань на обчислювальні ресурси планувальники не враховують характер пакета завдання (зв'язність задач у завданні. Проведений аналіз показав, що метод Backfill, також не враховує час, витрачений на передачу вхідних і вихідних даних завдання, і під час його використання немає важелів впливу на хід розподілу конкретного завдання, а розподіл здійснюється за принципом «перший відповідний».

Тому обґрунтована необхідність розробки нових методів розподілу завдань у GRID-системах. Для цього у роботі вирішується науково-технічна задача розробки нових методів розподілу завдань, що орієнтовані на використання нової інформаційної технології розподілу завдань, яка дозволяє враховувати й усувати недоліки існуючих рішень задачі розподілу завдань, що були виявлені в ході порівняльного аналізу.

В *третьому розділі* запропонована модифікована модель розподілу завдань в GRID-системі за рахунок розширення кортежів подання завдань та ресурсів. Ці дії дозволяють усунути недоліки, які були виявлені при аналізі, та підвищити ефективність використання обчислювальних ресурсів GRID-системи.

Однак підбір обчислювальних ресурсів не за одним параметром накладає труднощі для процесу розподілу, тому у роботі пропонується згортка параметрів у єдиний критерій оцінки завдання, за рахунок розподілу параметрів на два класи: якісні та кількісні. Використання цього параметру допомагає відібрати множину обчислювальних ресурсів, на яких може бути запущено завдання, а також показує, яку частину ресурсу займає завдання в процесі виконання.

На даний час в системах розподілу завдань на обчислювальні ресурси поширено метод зворотного заповнення Backfill, перевагою якого є те, що він гарантує отримання ресурсів для високопріоритетних завдань за мінімально можливий час та допускає порушення порядку черги, що дозволяє завданням з більш низьким пріоритетом використовувати обчислювальні ресурси, які на даний час не використовуються жодним із завдань. Також метод Backfill запобігає дефрагментацію обчислювальних ресурсів, однак жодним чином не враховує час, який витрачається на передачу вхідних та вихідних даних завдання (даний час безпосередньо залежить від пропускної здатності каналу зв'язку і часу затримки). Тому в роботі запропонована модифікація методу Backfill, яка враховує трафік обміну даними

між задачами в завданні, що дозволяє скоротити час на пересилання результатів між окремо взятими ресурсами і зменшити трафік у мережі за рахунок розвантаження каналів зв'язку.

Актуальною задачею у GRID-системі є задача ефективного розподілу завдань за наявними обчислювальними ресурсами. У роботі подано метод пошуку розподілу з мінімальним часом виконання пулу завдань та мінімальним простоем обчислювальних ресурсів. За допомогою даного методу можливо обрати план розподілу, який дозволяє мінімізувати ряд критеріїв: час виконання пулу завдань, час простою обчислювальних ресурсів і час перебуванням заявок у черзі.

На основі модифікованої математичної моделі розподілу завдань в GRID-системі, модернізації методу Backfill з консервативним резервуванням, розробці методу пошуку найкращого розподілу для пулу завдань здобувачем запропоновано інформаційну технологію розподілу завдань. Ця інформаційна технологія була впроваджена в середовище моделювання GRASS.

У *четвертому розділі* наведено практичну реалізацію запропонованої інформаційної технології та показано, що після її впровадження в імітаційне середовище моделювання GRASS стало можливим:

- паралельне моделювання реального процесу розподілу завдань на обчислювальні ресурси в GRID-системі за різними методами розподілу;
- вибір плану розподілу з мінімальним часом виконання пулу завдань і мінімальним простоем обчислювальних ресурсів;
- врахування зв'язності задач у завданні і класу завдань в процесі розподілу їх на обчислювальні ресурси.

У *додатках* наведено акти про впровадження результатів дисертаційної роботи у виробництво та навчальний процес.

Ступінь обґрунтованості та достовірності нових наукових положень, отриманих в дисертаційній роботі

Достовірність роботи забезпечується тим, що вона ґрунтується на відомих практичних і теоретичних викладках, які в роботі набувають потрібного розвитку згідно зі специфікою мети та задач досліджень, а також узгодженням із експериментальними даними, отриманими автором.

Результати роботи достатньо апробовано на міжнародних наукових та науково-практичних конференціях.

Оформлення дисертаційної роботи

Структура та обсяг дисертаційної роботи відповідають вимогам щодо тексту дисертацій на здобуття наукового ступеня кандидата технічних наук. Дисертацію написано грамотною мовою, з використанням сучасної бібліографії та наукової термінології. Зміст та результати досліджень викладено лаконічно та аргументовано. Задачі дисертації в логічній послідовності вирішуються на усіх етапах дослідження – від аналізу проблеми до побудови моделей та методів розв'язання задач, з наступним проведенням експериментів з конкретними чисельними даними та впровадженню результатів дослідження. Автореферат відповідає змісту дисертаційної роботи.

Публікації за темою дисертації

Основні результати дисертації достатньо повно викладені в 7 статтях, опублікованих у наукових фахових видання. Зміст наведених публікацій не містить повтору. Вони адекватно відображають зміст та результати дисертації.

Практичне та наукове значення результатів дисертаційної роботи

Практична значимість отриманих теоретичних результатів дисертаційної роботи Філімончук Т.В. підтверджено поліпшенням продуктивності GRID-системи, за рахунок скорочення простою обчислювальних ресурсів.

Автором вперше запропоновано метод пошуку розподілу завдань з мінімальним часом виконання пулу завдань та мінімальним простим обчислювальних ресурсів, удосконалено метод розподілу завдань на обчислювальні ресурси Backfill з урахуванням трафіку всередині завдання, а також розширені моделі подання завдань і ресурсів у GRID-системі.

Запропоновано інформаційну технологію розподілу завдань на обчислювальні ресурси в GRID-системі, яка на підставі використання множини методів розподілу дозволяє здійснювати вибір розподілу завдань з мінімальним часом виконання пулу завдань і мінімальним простим обчислювальних.

Запропоновано блок, який відповідає за формування додаткових параметрів для ефективного розподілу. Він включає 2 модуля: модуль згортки кортежу і модуль аналізу зв'язності. Модуль згортки кортежу здійснює обчислення узагальненого критерію оцінки для кожного завдання, що дозволяє більш продуктивно управляти процесом розподілу завдань на обчислювальні ресурси. Модуль аналізу зв'язності дозволяє здійснювати підбір обчислювальних ресурсів з урахуванням скорочення часу, що витрачається на обмін між задачами у завданні.

Результати, отримані автором, слід вважати важливими для подальшого розвитку систем розподілу завдань на обчислювальні ресурси в GRID-системі. Крім того, деякі з цих результатів можуть бути використані у навчальному процесі для підвищення практичної підготовки фахівців в галузі створення та використання глобальних розподілених інформаційних систем.

Зауваження

1. В підрозділі 3.2 автором запропоновано розподіл параметрів кортежу завдань (3.1) на якісні та кількісні, але не надано пояснень щодо їх визначення.

2. Запропонована модель GRID-системи не враховує деякі архітектурні рішення, зокрема віртуальну організацію.

3. У розділі 4 немає детального опису експериментів, зокрема не вказано яким чином оцінювались час виконання завдань та час використання обчислювальних ресурсів GRID-системи.

4. У запропонованій модифікації методу Backfill на другому кроці розподілу завдання використовується критерій зв'язності задач у завданні. Не зрозуміло яким чином формується даний критерій.

5. З роботи не зрозуміло, яким чином запропонована модель GRID-системи враховує локальні та глобальні завдання.

6. Загальна цінність отриманих у дисертаційній роботі результатів була б більш переконливою, якби автор навів у додатках розгорнуті кількісні оцінки результатів тестового виконання розподілу завдань на обчислювальні ресурси за допомогою середовища GRASS.

Однак ці зауваження суттєво не впливають на загальну позитивну характеристику дисертації, що має визначені вище актуальність, наукову новизну і практичну значущість.

Висновок

Дисертаційна робота Філімончук Тетяни Володимирівни є завершеним науково-практичним дослідженням, що містить нові науково-обґрунтовані результати щодо удосконалення розподілу завдань на обчислювальні ресурси GRID-систем.

Впроваджені результати дають можливість більш продуктивно управляти процесом розподілу завдань на обчислювальні ресурси та здійснювати підбір обчислювальних ресурсів з урахуванням скорочення часу, що витрачається на обмін між задачами у завданні.

Вважаю, що дисертаційна робота відповідає вимогам пунктів 9, 11, 12-14 «Порядку присудження наукових ступенів», затвердженого постановою Кабінету Міністрів України від 24.07.2013 №567 щодо кандидатських дисертацій, а її автор Філімончук Тетяна Володимирівна заслуговує на присудження наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології.

Декан факультету комп'ютерних
наук і програмної інженерії Національного
технічного університету «Харківський
політехнічний інститут»,
доктор технічних наук, професор



І.П. Гамаюн

Підпис Гамаюна Ігоря Петровича засвідчую.

Вчений секретар



Підпис Галеюка О.П.
ЗАСВІДЧУЮ:
ВЧЕНИЙ СЕКРЕТАР
НАЦІОНАЛЬНОГО-ТЕХНІЧНОГО УНІВЕРСИТЕТУ
"ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"
Заковоротний О.Ю.
" 04 " 2017 р.