

$$v = \left(J \cdot e^{\frac{1}{8\omega^2}} \right)^{-1} * \left(\left(f_m(0.5 \cdot J) \log_2 \left(\frac{J}{f_m(0.5 \cdot J)} \right) + (J - f_m(0.5 \cdot J)) \log_2 \left(\frac{J}{J - f_m(0.5 \cdot J)} \right) \right) - \left(\log_2(J) + (J - 1) \log_2 \left(\frac{J}{J - 1} \right) \right) \right)$$

де $f_m(*)$ – функція виділення цілої частини числа, ω – параметр значення якого $\approx 0.15 - 0.2$ для кількості класів $J \leq 10$.

На рис. 2 зображений розкид значень показника $\eta_j = \frac{\delta(J, p(f_k | y_j), \omega, v) - \delta'(J, p(f_k | y_j), \omega, v)}{\sigma(p(f_k | y)) \cdot e^{\frac{\sigma(p(f_k | y))^2}{2\omega^2}}}$ при $J = 10$ зі застосуванням корегуючого коефіцієнта $v \cdot e^{\frac{\sigma(p(f_k | y))^2}{2\omega^2}}$ (рис. 2 В) та без його застосування (рис. 2 А).

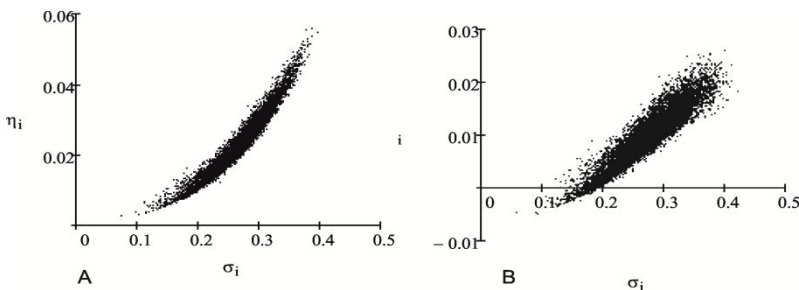


Рисунок 2 - Розкид η_j при різних значеннях середньоквадратичного відхилення ймовірностей появи класів

Література

УДК 004.75: 004.4'2

ФРЕЙМВОРК ДЛЯ РОЗРОБКИ GRID-ЗАСТОСУВАНЬ З ПІДТРИМКОЮ ЗАДАЧ РОЗПОДІЛЕНОГО МОДЕЛЮВАННЯ

В.В. Казимир, О.А. Пріла

Чернігівський державний технологічний університет, Україна

Вступ

Незважаючи на розвиток та впровадження grid-технологій у різні галузі науки, вирішення прикладних задач у grid-середовищі являє собою складну технічну задачу. Основними проблемами є відсутність високорівневих засобів, недостатність відповідної документації та

спеціалістів, відсутність інтеграції між існуючим проміжним програмним забезпеченням грид (ППЗ, middleware) [1, 2].

Існує велика кількість задач різних наукових галузей (економічне прогнозування, аналіз економічних ризиків, екологічний моніторинг, гена інженерія та ін.), які потребують високорівневих засобів організації імітаційного моделювання у grid-середовищі. Більшість існуючих на сьогодні віртуальних організацій (ВО) [3] не надають зручного інтерфейсу для використання спеціалізованих грид-сервісів. Навіть базові грид-операції виконуються в консольному режимі, що робить неможливим використання грид-середовища прикладними спеціалістами.

Метою роботи є формування вимог до високорівневих засобів створення grid-застосувань, окреслення недоліків існуючих засобів та проектування фреймворку для створення grid-застосувань згідно окреслених вимог.

1 Проблеми застосування grid-середовища для вирішення прикладних задач розподіленого моделювання

Розглянемо типи задач, що є найбільш адаптованими до вирішення у grid-середовищі.

1. Задача являє собою задачу великої розмірності, однак при виконанні задачі немає потреби у використанні всієї доступної grid-інфраструктури. Тобто необхідність використання grid-середовища зводиться лише до використання віддаленого потужного кластерного ресурсу.
 2. Задача являє собою єдиний обчислювальний блок та декомпозується на декілька підзадач шляхом розподілу вхідних даних на декілька порцій для паралельної обробки.
 3. Задача, яка декомпозується на декілька паралельно-последовних підзадач з певною схемою синхронізації обчислень. Беручи до уваги принципи конвеєрної обробки [4], третій тип задач будемо називати такими, що характеризуються конвеєрною схемою обчислень.
- Окремо слід виділити комбінований тип задач, коли задачі властиві характеристики декількох типів, та частковий випадок третього типу, коли задача декомпозується на повністю незалежні паралельні підзадачі.

Задача розподіленого моделювання може полягати у наступному:

- запуск загальної моделі на декількох кластерах одночасно з різними порціями вхідних даних (у випадку задачі аналізу економічних ризиків отримання прогнозів для різних діапазонів значень факторів);
- поділ загальної моделі на декілька часткових (наприклад, для задач економічного прогнозування поділ за регіональним критерієм) та одночасних запуск часткових моделей на декількох кластерах з наступним об'єднанням результатів.

Отже, задача імітаційного моделювання може являти собою другий або третій тип задач.

Основними проблемами застосування grid-середовища для вирішення прикладних задач є наступні:

- необхідність установки набору пакетів програмного забезпечення grid на клієнтській машині;
- складність використання grid-середовища у консольному режимі прикладними користувачами;
- складність формування низькорівневих специфікацій задачі (причому єдиного формату специфікації задачі для різного ППЗ на сьогодні не існує);
- складність адаптації прикладної задачі до grid-середовища. Із зазначених вище типів задач, лише перші два частково підтримуються проміжним програмним забезпеченням grid на сьогодні.

Розробка високорівневих grid-застосунків з графічним інтерфейсом користувача дозволить значно підвищити ефективність використання grid-середовища прикладними користувачами.

2 Огляд існуючих засобів розробки grid-застосунків та систем розподіленого моделювання

Було проведено дослідження існуючих відкритих фреймворків та бібліотек для створення високорівневих grid-застосунків: Ganga [5], Grid Portal Development Kit (GRDK) [6], GridSphere Portal Framework [7], SimpleGrid [8], Intel GPE API [9], Libarcclient [2]. За результатами досліджень було окреслено недоліки існуючих засобів розробки прикладних grid-застосунків:

- відсутність інтеграції із різним проміжним програмним забезпеченням grid. Із зазначених, лише бібліотека Libarcclient спрямована на підтримку різного middleware;
- недостатня реалізація базового функціоналу для роботи у grid-середовищі та підтримки роботи віртуальних організацій;
- відсутність автоматизованого запуску для вирішення у grid-середовищі задач третього типу;
- відсутність зручного інтерфейсу для розробки спеціалізованих grid-сервісів ВО;
- орієнтація на консольний режим роботи та використання скриптів. API для реалізації клієнтських застосунків з графічним інтерфейсом є не повнофункціональним;
- часта зміна API, для більшості проектів відсутня підтримка на сьогодні.

В [10] наведено результати дослідження існуючих систем розподіленого моделювання CSE, CUMULVS, VISIT, VIPER, CACTUS,

Discover, TENT. Зазначемо проблеми використання цих систем для вирішення задач імітаційного моделювання у grid-середовищі:

- більшість з систем характеризуються централізованою архітектурою, що не відповідає стандартам grid;
- відсутність підтримки паралельних бібліотек;
- відсутність підтримки сучасного middleware;
- відсутність підтримки різних підходів щодо декомпозиції задачі імітаційного моделювання (за даними або поділ загальної моделі на часткові);
- для деяких відсутність підтримки на сьогодні.

Проведене дослідження підтверджує актуальність розробки фреймворку для створення грід-застосувань з підтримкою задач імітаційного моделювання.

3 Проектування фреймворку для створення grid-застосувань

Фреймворк розробляється на базі бібліотеки Libarcclient згідно наведених нижче функціональних та нефункціональних вимог.

1. Надання API для виконання базових грід-операцій: створення проксі-сертифікату, перегляд доступних обчислювальних ресурсів та сховищ даних, запуск задачі, моніторинг виконання в динамічному режимі, отримання результатів та ін.
2. Автоматизована генерація файлу опису задачі згідно визначених користувачем вимог до ресурсів.
3. Підтримка наведених вище трьох типів задач з наданням графічних компонентів для визначення обчислювальних блоків задачі та схеми синхронізації між ними.
4. Надання API для підтримки роботи віртуальних організацій:
5. Надання API для розгортання нових грід-сервісів VO.
6. Орієнтація на підтримку різного ППЗ (гнучкість реалізації для забезпечення можливості розширення). Підтримка ППЗ, що використовується в межах української національної грід-інфраструктури, а саме ARC Nordugrid та gLite.
7. Кросплатформеність, розширюваність, масштабованість.
8. Забезпечення безпеки грід-порталу та ресурсів VO.

Фреймворк надає засоби визначення workflow задачі (як у графічному, та і у текстовому режимі) з метою виділення послідовних та паралельних блоків обчислень, потоків даних та схеми синхронізації обчислень між блоками. При запуску задачі з конвеєрною схемою обчислень система надає механізм оптимального вибору ресурсів з урахуванням параметрів мережі з метою мінімізації витрат на передачу даних між обчислювальними блоками з одного грід-вузла на інший. Підсистема вирішення задач імітаційного моделювання розробляється на базі стандартів High Level Architecture (HLA) [10].

Фреймворк надає розробнику grid-застосувань набір API у вигляді java-бібліотек. На даний момент фреймворк реалізує підтримку лише ARC Nordugrid 11.05, в подальшому планується підтримка gLite.

Висновок

Фреймворк для розробки грид-порталів розробляється у межах відкритого проекту. Результат буде представлений у вигляді java-бібліотеки та прикладу застосування фреймворку для створення грид-порталу віртуальної організації економічного прогнозування. Застосування фреймворку дозволить зменшити витрати при розробці грид-порталів та вирішити проблему складності використання грид-серведовища прикладними користувачами.

Література

1. gLite - Lightweight Middleware for Grid Computing [Електронний ресурс]. – Режим доступу: <http://glite.cern.ch>.
2. Nordugrid. Advanced Resource Connector [Електронний ресурс]. – Режим доступу: <http://www.nordugrid.org/arc>.
3. ARC Грид-монитор [Електронний ресурс]. – Режим доступу: <http://www.nordugrid.org/monitor/loadmon.php?display=vo=Ukraine>.
4. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. – СПб.: БХВ-Петербург, 2004. – 608 с.
5. Ganga. Simplifying use of the Grid [Електронний ресурс]. - Режим доступу: <http://ganga.web.cern.ch/ganga>.
6. DOE Science Grid: Grid Portal Development Kit (GPDK) [Електронний ресурс]. - Режим доступу: <http://doesciencegrid.org/projects/GPDK>.
7. Install GridSphere Portal Framework [Електронний ресурс]. - Режим доступу: http://technical.bestgrid.org/index.php/Install_GridSphere_Portal_Framework.
8. Projects:simplegrid:index [CyberInfrastructure and Geospatial Information Laboratory] [Електронний ресурс]. - Режим доступу: <https://www.cigi.uiuc.edu/doku.php/projects/simplegrid/index>.
9. GPE4GTK project [Електронний ресурс]. – Режим доступу: <http://gpe4gtk.sourceforge.net/>.
10. Prof. mr. P. F. van der Heijden, Katarzyna Rycerz. Grid-based HLA Simulation Support // University of Amsterdam and AGH Krakow, 2006.