

Grid-технології: розвиток, моделювання та перспективи постбінарного комп'ютингу

Анопрієнко О.Я., Дзьоба В.В., Конопльова Г.П., Аль-Абабнех Х.
Донецький національний технічний університет

Abstract

Anoprienko A., Dzjoba V., Konopljova A., Al-Ababneh H. Grid-technologies: development, simulation and perspectives of postbinary computing. Development of Grid-technologies in the world and in Ukraine is described. The importance of using appropriate simulation tools, new code-logical basis and experience of using GridSim Toolkit are analyzed.

Розвиток Grid-технологій

Grid на сьогодні визнається однією з найбільш вагомих концепцій як з точки зору сучасного розвитку комп'ютерних технологій, так і їх майбутнього. Авторами цієї концепції вважаються Ян Фостер з Арагонської національної лабораторії Чиказького університету і Карл Кессельман з Інституту інформатики Університету Південної Каліфорнії. Саме вони в 1998 році вперше запропонували термін Grid-комп'ютинг (Grid-computing) для позначення універсальної програмно-апаратної інфраструктури, що об'єднує комп'ютери та суперкомп'ютери в територіально-розподілену інформаційно-обчислювальну систему [1]. Grid, відповідно до їх визначення, що стало вже класичним, це - «погоджене, відкрите і стандартизоване середовище, яке забезпечує гнучке, безпечне, скоординоване розподілення ресурсів у рамках віртуальної організації» [2].

Концепція Grid орієнтована на створення комп'ютерної інфраструктури нового типу, що забезпечує глобальну інтеграцію інформаційних і обчислювальних ресурсів на основі керівного і оптимізуючого програмного забезпечення (middleware) нового покоління. Для досягнення цієї мети створюється набір стандартизованих служб для забезпечення надійного, сумісного, дешевого і універсального доступу до географічно розподілених високотехнологічних інформаційних і обчислювальних ресурсів. Основними напрямками використання Grid на даний момент є наступні:

- розподілені суперобчислення, рішення дуже великих задач, що вимагають величезних процесорних ресурсів, пам'яті і таке інше;
- «обчислення з високою пропускну здатністю» (High-Throughput Computing), що дозволяють організувати ефективне використання ресурсів для невеликих задач, утилізуючи тимчасово вільні комп'ютерні ресурси;

- обчислення «на вимогу» (On-Demand Computing) - великі разові розрахунки;
- обчислення із залученням великих об'ємів розподілених даних (Data-Intensive Computing), наприклад, в метеорології, астрономії, фізиці високих енергій;
- колективні обчислення (Collaborative Computing).

Як прогнозується, еволюційні зміни в напівпровідникових технологіях і архітектурі мікропроцесорів приведуть в найближчі п'ять років до десятиразового збільшення обчислювальних потужностей. Вже сьогодні можливості рядових користувачів, підключених до цифрових каналів зв'язку з наданням комплексних послуг майже рівні з тими можливостями, якими володіли суперкомп'ютерні центри ще 10-15 років назад.

Технологічними підставами для створення Grid-інфраструктури є вже існуючі волоконно-оптичні мережі, високопродуктивні процесори, паралельні архітектури, протоколи зв'язку, математичне забезпечення розподілених структур, механізми забезпечення безпеки та веб-сервіси.

Фактично, з початком XXI століття почався перехід на нові Grid-технології, коли на зміну звичайному Інтернету з його web-послугами йде веселиття Grid-мережа як засіб сумісного використання обчислювальних потужностей і сховищ даних. Grid дозволяє вийти за рамки простого обміну даними між комп'ютерами і врешті-решт перетворити глобальну мережу в свого роду гігантський віртуальний комп'ютер, доступний в режимі видаленого доступу з будь-якої комп'ютера незалежно від місця розташування користувача.

На сьогодні в світі реалізується декілька масштабних проектів, орієнтованих на розвиток Grid, серед яких в першу чергу треба назвати EGEE (Enabling Grids for E-sciencE) – проект, направлений на побудову Grid-інфраструктури, яка зможе використовуватися в численних наукових дослідженнях в Європі і світі. Проект

фінансується по 6-й рамковій програмі Європейської комісії IST FP6. Консорціум учасників проекту включає більше 70 інститутів з 27 європейських країн. Для досягнення заявленої мети EGEE використовує унікальний досвід, одержаний в попередніх проектах Європейського співтовариства (DataGrid, Crossgrid, DataTAG) і національних ініціатив (UK e-Science, INFN Grid, NorduGrid, US Trillium).

Проект EGEE використовує middleware типу gLite, яке створюється колективно – в його розробці беруть участь більше 80 фахівців з 11 дослідницьких центрів. Основна відмінність комплексу gLite від класичного GlobusToolkit у тому, що крім інструментальних засобів, в нього входить широкий набір різноманітних служб. Наприклад, «обчислювальний елемент» (Computing Element – CE) – це служба, що представляє ресурсний вузол Grid і виконує на ньому функції управління завданнями (запуск, видалення і т.д.). Ініціативи з використання CE можуть виходити або від інтерфейсу користувача, або від «менеджера завантаження» (Workload Manager – WM), який розподіляє завдання по множині CE. Однією з найбільш важливих є «підсистема управління даними» (Data Management Subsystem – DM), яка включає три служби, підтримуючі доступ до файлів: «елемент пам'яті» (Storage Element – SE), «служби каталогу» (Catalog Services – CS) і «диспетчер даних» (Data Scheduling – DS). Всі служби працюють з даними на файловому рівні, у протилежність, наприклад, системам баз даних, які оперують такими елементами як записи і поля.

В Україні перший Grid-вузол з'явився в 2002 році. Він був створений групою фізиків з Національного наукового центру «Харківський фізико-технічний інститут» (ННЦ ХФТІ) і був побудований в рамках співпраці з Об'єднаним Інститутом Ядерних Досліджень (ОІЯІ, Дубна, Росія) і вченими ЦЕРН (CERN, Женева, Швейцарія).

А вже у 2004 році ідеї проекту створення української національної Grid-інфраструктури були проголошені академіком НАНУ Згуровським М.З на саміті WSIS (World Summit on Information Society) [3]. Пізніше ця доповідь була передрукована в вітчизняній літературі [4]. Мова йшла про необхідність створення освітнього і дослідницького сегменту інформаційного суспільства України з вирішенням наступних головних задач:

- Створення Grid-інфраструктури для підтримки освіти і наукових досліджень.
- Інфраструктурна розробка української дослідницької і академічної мережі УРАН і інтеграція її до Європейської мережі GEANT-2.
- Надбання досвіду роботи з Grid-інфраструктурою і експлуатації розподілених інформаційних систем, розробка алгоритмів і методів рішення прикладних задач в середовищі розподіленого комп'ютерингу.

Запропоновані цілі і задачі знайшли своє відображення вже у 2005 році у Державній цільовій програмі «Інформаційні та комунікаційні технології в освіті й науці на 2006-2010 роки», прийнятій Кабінетом Міністрів України. А з 2007 року Інститут Прикладного Системного Аналізу (ІПСА) Національного Технічного Університету «Київський політехнічний Інститут», розпочав, як головна організація, проект Ugrid «Створення національної Grid-інфраструктури для підтримки наукових досліджень» Цей проект [5] розпочав побудову національної Grid-інфраструктури з відповідними сервісами для надання можливості вітчизняним науковцям плідно співпрацювати в Європейському науковому просторі (European Research Area, ERA) і сприяти створенню економіки інформаційного суспільства, заснованої на знаннях, шляхом впровадження наукових концепцій Grid і найбільш вагомих наукових додатків, які використовуються в Grid середовищі [6, 7]. Виконавцями проекту стали провідні технічні університети України, серед яких Донецький національний технічний університет [8].

Паралельно відповідні розробки і дослідження проводились в НАН України [9], де вже у 2006 році була розроблена Концепція розвитку Grid-інфраструктури НАН України.

У цьому ж році Національна академія наук України на засіданні керівного комітету організації WLCG (Worldwide LHC Computing Grid) була офіційно прийнята в члени цієї організації. Розпорядженням Президії НАН України була також затверджена Програма «Упровадження Grid-технологій і побудова кластерів в НАН України» і була створена Координаційна Рада Програми. Базовим інститутом, відповідальним за виконання Програми, був призначений Інститут теоретичної фізики.

У вересні 2006 року на конференції EGEE-06 (Enabling Grid for E-science), найбільшого європейського Grid-проекту, який об'єднує 90 учасників з 32 країн і фінансується Європейським Союзом, Український Академічний Грід був зареєстрований кандидатом в члени цього проекту. В кінці 2006 року кластери в провідних інститутах НАН України були побудовані, а в березні 2007 року в тестовому режимі почав працювати перший в Україні Grid-сегмент, який в даний час об'єднує обчислювальні ресурси (більш півтисячі процесорів і близько 60 Терабайт дискової пам'яті) 7-ми провідних наукових організацій НАН України і Київського національного університету імені Тараса Шевченка. Офіційна презентація першого в Україні Grid-сегменту відбулася 4-го квітня 2007 року в Інституті теоретичної фізики на виїзному засіданні бюро Президії НАН України. На цьому засіданні відбулося широке обговорення задач розвитку Grid-технологій не тільки в Академії наук. Були обговорені і намічені перспективи співпраці НАН України і Міністерства освіти і

науки України, зокрема, з Національним Технічним Університетом України Київський Політехнічний Інститут (НТУУ КПІ). Досягнута домовленість про створення Національного Гріда України. В даний час, відповідно до рішення бюро Президії НАН України, активно налагоджується співпраця Українського Академічного Гріда і відповідного проекту МОН України, результатом якого буде створення Українського Національного Гріда і його інтеграція з міжнародною Grid-інфраструктурою.

Активне використання Grid-технологій потребуватиме наявності відповідних фахівців, тому необхідно забезпечити підготовку та перепідготовку у вищих навчальних закладах таких фахівців, розробити відповідні плани і програми. Відповідно до цієї мети вже у 2008 році в Україні був виданий перший навчальний посібник по використанню Grid-технологій в науці і освіті [10] і з'явилися відповідні навчальні курси в провідних університетах країни. Треба також констатувати, що важливою передумовою широкого розповсюдження Grid-технологій, у тому числі у навчальному процесі, є можливість використання відповідних засобів моделювання. Особливо важливо це на початкових етапах впровадження нових технологій, коли експерименти на реальних системах є або ще неможливими, або пов'язані з чисельними труднощами.

Моделювання Grid-інфраструктури

Участь ДонНТУ у проекті створення національної Grid-інфраструктури на протязі 2007-2008 рр. передбачала низку конкретних завдань, серед яких одним з головних був аналіз наявних методів та засобів моделювання Grid навчального та наукового призначення. В результаті такого аналізу був вибраний GridSim Toolkit – відкритий фреймворк для моделювання Grid-систем. Розробляється цей продукт у Мельбурні (Австралія). Головними спонсорами цього проекту є Університет Мельбурну і фірма Microsoft. Проект активно розвивається та бази технологій Java та розповсюджується за відкритою ліцензією GPL (GNU General Public License) [11].

GridSim – це платформа, яка надає можливість споживачам моделювати і симулювати характеристики ресурсів Grid і мереж з різними конфігураціями. GridSim надає великі можливості як студентам, так і досвідченим дослідникам, які хочуть вивчати Grid, або перевіряють нові алгоритми і стратегії в оточенні, яким управляють. Використовуючи GridSim, вони можуть виконувати повторювані експерименти, які не можливі в справжньому динамічному оточенні Grid. GridSim заснований на пакеті моделювання SimJava2, реалізованому на базі технологій Java. Усі компоненти в GridSim

спілкуються один з одним через передачу повідомлень, визначених у SimJava2. Другий рівень моделює основні елементи інфраструктури, а саме ресурси Grid, як, наприклад, кластери, сховища інформації та зв'язки мережі. Типова модель Grid-системи характеризується при цьому наступними параметрами: кількість вузлів, параметри обчислювальних ресурсів, політика розподілу задач, кількість задач, складність задач, тип мережі, завантаженість мережі та інше.

В процесі моделювання було встановлено, що одним з найбільш вагомих параметрів, які впливають на ефективність функціонування Grid-систем, є політика розподілу задач серед наявних вузлів. При цьому розглядалися наступні політики:

- проста (BASIC_ADAPTIVE_JOB_SCHEDULING) – для кожної задачі окремий вузол;
- економії ресурсів (ECONOMIC_JOB_SCHEDULING) – декілька задач на вузол;
- координації віртуальних організацій (VO – Virtual Organization) (VO_POLICY_COORDINATION) – обмін задачами між VO задля максимізації завантаження ресурсів;
- «соціальної мережі» (SOCIAL_NETWORK_GRID) – обмін задачами між VO за рекомендаціями агентів вузлів.

В процесі моделювання був використаний Agent Grid Repast Simulator – проект з відкритим кодом, який також розповсюджується за ліцензією GPL і базується на GridSim Toolkit.

В результаті моделювання було запропоновано удосконалену політику економії ресурсів ECONOMIC_RESOURCE_ALLOCATION, яка дозволяє мінімізувати загальну ціну рішення задач. При цьому була реалізовано низка спеціальних методів сценарію економії ресурсів, найважливішими з яких є наступні:

- void bootstrapAgents() – до кожного з вузлів GRID-системи приєднується агент «покупця» та «продавця» які будуть проводити операції з «купівлі/продажу» ресурсу;
- EconomicAgent getBestScoringAgent(ArrayList lisOfAgents) – з усіх агентів обирається той агент, у якого найбільший показник загальної користі передачі задачі;
- void doScenarioAction() – один з головних методів політики, який перш за все відновлює ресурсами; далі повідомляє «продавців» про те, які ресурси для виконання задач потрібні; проводить обробку отриманих даних; запускає «аукціон продажу/купівлі» ресурсів; повідомляє покупців о результатах «аукціону»; «покупці» «купають» ці ресурси,

- а ресурси виділяються для «покупців»; запускається механізм моделювання результатів вищенаведених дій;
- void sendCFPs() – повідомлення «продавців» про те, які ресурси необхідні для виконання задач;
- void selectProposals() – серед усіх «покупців» обирається найкраща пропозиція ціни;
- void evolveInterGroups() – відбувається міграція між групами, агенти з максимальної користю угод розподіляються посеред усіх груп (VO).

При цьому вважається, що кожний вузол Grid-системи має своїх агентів «продажу» та «купівлі». «Ціна» ресурсу змінюється випадковим чином при кожній ітерації процесу на ± 1 умовного значення, або може не змінюватися. Після зміни ціни кожний агент «купівлі» шукає серед агентів «продажу» найвигіднішу пропозицію. VO, у яких знаходяться однотипові задачі, можуть виступати у ролі агента. Таким чином процес пошуку вигідних ресурсів зменшується завдяки об'єднанню однотипових задач. Дослідження довели, що використання запропонованої політики економії ресурсів у порівнянні з простою політикою дає підвищення ефективності використання ресурсів майже у 2 рази, а в порівнянні з іншими політиками – в більшості випадків майже в 1,5 рази.

Перспективи постбінарного комп'ютингу на базі Grid

Подальший розвиток Grid-технологій разом з розвитком швидкісних комп'ютерних мереж в найближчі 5-9 років має привести до ситуації, коли використання глобально розподілених обчислювальних ресурсів буде майже таким же простим і ефективним, як використання локальних ресурсів. При цьому найбільш важливим буде те, що подібна ситуація буде мати місце як для великих задач, так і для відносно простих, але потребуючих для свого ефективного вирішення спеціальних видів ресурсів. Це сформує передумови для ефективної реалізації різноманітних варіантів постбінарного комп'ютингу, у тому числі – на основі використання постбінарного кодо-логічного базису, однією з природних властивостей якого є орієнтація на паралельні обчислювальні ресурси і процеси [12].

Література

1. Foster I., Kesselman C. The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Publishers, 1998. – 701 p.
2. Foster I. What is the Grid? A Three Point Checklist // Department of Computer Science, University of Chicago, Chicago, IL 60637, July 20, 2002, <http://www-fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf>.
3. Zgurovsky M. Z. Development of Educational and Research Segment of Information Society in Ukraine //Proc. WSIS . – Tunis. 2004 . P. 103-107.
4. Zgurovsky M. Z. Development of Educational and Research Segment of Information Society in Ukraine. - //Системні дослідження та інформаційні технології. – Київ. 2006. №1. С.7-17.
5. Національна Grid-інфраструктура України, <http://grid.ntu-kpi.kiev.ua>
6. Petrenko A. I. Development of Grid - infrastructure for Educational and Research segment of Information Society in Ukraine with focus on Ecological monitoring and Telemedicine // Data Science Journal.- Volume 6.- Supplement, 14 April 2007. – P.583-590.
7. Петренко А.І. Національна Grid - інфраструктура для забезпечення наукових досліджень і освіти.- // Системні дослідження і інформаційні технології, -Київ.- №1- 2008.- С.79-92.
8. Донецкий узел украинской национальной GRID-инфраструктуры, <http://grid.donntu.edu.ua/>
9. Грід – нова інформаційно-обчислювальна технологія для науки / А. Загородній, Г. Зінов'єв, Є. Мартинов, С. Свистунов, В. Шадура // Вісн. НАН України. — 2005. — N 6. — С. 17-25.
10. Петренко А.І. Вступ до Grid-технологій в науці і освіті (навчальний посібник).-Київ: Політехніка. – 2008. – 124 С.
11. **GridSim Toolkit**. The Grid Computing and Distributed Systems (GRIDS) Laboratory, University of Melbourne, <http://www.gridbus.org/gridsim>.
12. Аноприенко А.Я. Обобщенный кодо-логический базис в вычислительном моделировании и представлении знаний: эволюция идеи и перспективы развития // Научные труды Донецкого национального технического университета. Серия: Информатика, кибернетика и вычислительная техника (ИКВТ-2005) выпуск 93: - Донецк: ДонНТУ, 2005. С. 289-316.

Как правильно ссылаться на эту статью:

Аноприенко О.Я., Дзьоба В.В., Конопльова Г.П., Аль-Абабнех Х. Grid-технології: розвиток, моделювання та перспективи постбінарного комп'ютингу // Наукові праці Донецького національного технічного університету. Серія “Інформатика, кибернетика і обчислювальна техніка”. Випуск 10 (153) – Донецьк, ДонНТУ, 2009, с. 324-327.