

УДК 681.3

О. Я. Матов, І. О. Храмова

Інститут проблем реєстрації інформації НАН України
вул. М. Шпака, 2, 03113 Київ, Україна

Перспективні інформаційні технології та розвиток GRID-систем у високопродуктивних глобально-розподілених обчислювальних інфраструктурах корпоративної співпраці

Розглянуто GRID-технології як важливий та перспективний для України напрямок в організації IT-інфраструктур для корпоративного співробітництва і спільного використання розподілених обчислювальних й інформаційних ресурсів в урядуванні, бізнесі, науці й освіті. Виконано порівняння та аналіз можливостей розподілених технологій, типів GRID-систем, як засобу спільного використання ресурсів; напрямків подальших досліджень, використання, розвитку та впровадження GRID-технологій. Підкреслено необхідність координації робіт вітчизняних учених і фахівців у даному напрямку, який розвивається дуже динамічно.

Ключові слова: інформаційні технології, метаобчислення, GRID-технологія, розподілені обчислення, обчислювальна інфраструктура, корпоративна співпраця.

Вступ

Унікаючи жодного заперечення можна стверджувати, що вимоги суспільства до інформаційних систем, обумовлені впровадженням у життя швидкісних обчислювальних мереж, щодалі зростатимуть. Цьому в великій мірі сприяє поширення глобальних програмних проектів на зразок електронних бібліотек (e-Library), електронної науки (e-Science) та додатків електронного бізнесу типу B2B, C2B, C2C, B2C, що дедалі набувають популярності. Поширення в світі концепції електронного урядування та темпи впровадження відповідних рішень (G2C, G2B, C2G, B2G) у більшості країн світу ще більше загострили проблему так званої «інформаційної повені».

Іншим джерелом вимог залишаються наукові дослідження, в яких застосування новітніх технічних засобів (синхротронів, радіотелескопів, електронних мікроскопів тощо) і технологій призвело до значного збільшення потоку результатів експериментів та обчислень, що потребують значних ресурсів для обробки та аналізу.

© О. Я. Матов, І. О. Храмова

Глобальні обчислення вимагають нових пріоритетів і принципів, єдиного й універсального представлення комп'ютерної інформації. Колишні способи представлення, виробництва і споживання інформації, що були перенесені в Інтернет майже без змін, більш не здатні забезпечувати її ефективне використання. Перед суспільством постало питання, як впоратись із наростаючим валом цифрових даних, беручи до уваги, що зростання їх обсягу, на жаль, веде до зменшення питомої ваги корисної інформації. Для вирішення глобальних задач замало тільки забезпечувати споживача інформацією. Виникає потреба у такому ступені автоматичної її обробки, який у багато разів перевищує можливості, що надаються сучасними ІТ-інфраструктурами (ІТ — інформаційні технології).

Технологічні зусилля, націлені на вирішення завдання значного збільшення потужності обчислень, зараз розглядаються у двох аспектах: нові рішення в архітектурі комп'ютерів (суперкомп'ютери) та нові рішення в організації обчислень з використанням обчислювальних мереж (метакомп'ютинг).

Сьогоднішні суперкомп'ютери найвищої продуктивності складаються з декількох тисяч процесорів, поєднаних спеціалізованою високошвидкісною з низькими затримками мережею. Рівень швидкодії досягає декількох Терафлопів на секунду. Більше того, розробляються експериментальні системи, що мають за мету досягнення продуктивності у Петафлопи на секунду [1, 2]. Однак, таких високоякісних систем, звичайно, мало і їхні ресурси розташовані у відносно невеликій кількості місць.

Незважаючи на вражаючі успіхи, досягнуті на першому напрямку, неможливо обминути увагою і другий, який, зробивши старт десь у середині 90-х років минулого сторіччя, набув на сьогодні значного поширення в світі. На сьогодні мережі вже довели свою практичну корисність як засіб глобальної доставки різних форм інформації. Проте потенціал застосування мереж значно ширший: вони можуть розглядатися ще й як засіб організації обчислень наступної генерації.

Що уявляє собою Internet? Це — величезна кількість вузлів із власними процесорами, оперативною і зовнішньою пам'яттю, пристроями введення/виводу. Вузли з'єднані один з одним комутаційним устаткуванням і лініями передачі даних. Така конструкція дуже схожа із багатопроцесорною системою, у якій за магістральну шину править мережа.

Подолавши ж бар'єри між різнорідними, просторово-розподіленими обчислювальними системами, можна отримати метакомп'ютер, який для користувача виступав би як єдине обчислювальне середовище, доступне безпосередньо з персонального комп'ютера.

Технології, що підтримують спільне і скоординоване використання різних ресурсів у динамічних розподілених віртуальних організаційних структурах отримали назву GRID і ставлять за мету створення з географічно й організаційно розподілених компонентів віртуальних обчислювальних систем, що достатньо інтегровані, щоб надати бажану якість обслуговування [4].

Термін GRID (англ.: сітка, мережа), як назва нової технології корпоративних обчислень, був запозичений американськими науковцями у енергетиків, за аналогією з енергетичними мережами. Зважаючи на велику термінологічну навантаженість терміну «мережа», ми залишаємо в якості позначення згаданої нової техно-

логії запозичений з англійської термін, до того ж він є досить розповсюдженим і застосовується в згаданій формі у багатьох мовах світу.

GRID як технологія створення розподіленої обчислювальної інфраструктури

Напрямок, про який ідеться, можна позначити як використання комп'ютерних мереж для створення обчислювальної інфраструктури, розширеної до національного і, навіть, світового масштабу. На відміну від існуючих технологій віддаленого доступу глобальними мережами, ідеологія GRID спирається на концепцію надання цілком прозорого доступу користувача до GRID-мережі, надаючи йому при підключенні ресурси адекватної для його потреб потужності.

Існуючи спочатку поряд з поняттям метакомп'ютингу, останніми роками GRID, як технологія розподілених обчислень, зайняла власне місце і, розробивши достатньо розвинутий власний стандарт OGSA (Open Grid Services Architecture) [5], зараз вже розглядається, як останній і самий могутній еволюційний стрибок більшості з відомих нині розподілених технологій, таких як Web, однорангові мережі, кластери і розподілені обчислення, технології віртуалізації. Однак, окрім загальних з іншими технологіями розподілених обчислень властивостей, GRID-технології мають відмінності (табл. 1), що роблять цю технологію найпривабливішою для багатьох секторів ринку корпоративних обчислень. Серед комерційних компаній, що висловили підтримку стандарту OGSA, такі як IBM, Microsoft, Platform, Sun, Avaki, Entropia та United Devices.

Таблиця 1. Порівняння розподілених технологій

Технологія	Визначення	Подібність	Розбіжності
Web	Служба в мережі Інтернет, що дозволяє одержувати доступ до інформації на будь-яких серверах, підключених до мережі.	Приховує складність доступу: користувачі різного ступеня підготовки користуються єдиним стандартизованим підходом для одержання послуг	Відкриті протоколи загального призначення Web підтримують доступ до розподілених ресурсів, але не скоординоване застосування цих ресурсів, достатнє для надання потрібного рівня якості послуг.
Однорангові мережі (Peer-to-peer)	З'єднання рівноправних вузлів ЛОМ (що відрізняється відсутністю виділеного файл-сервера)	Дозволяє користувачам розділяти файли	Дозволяє великій кількості користувачів крім файлів розділяти й інші IT-ресурси
Кластери та розподілені обчислення	Група зв'язаних серверів, що функціонують як єдина стосовно задач і/чи користувачів система	Поєднує обчислювальні ресурси	На відміну від кластерних і розподілених обчислень, для яких необхідна фізична близькість і операційна однорідність ресурсів, GRID може бути географічно розподіленим і гетерогенним середовищем

Технологія	Визначення	Подібність	Розбіжності
Технології віртуалізації	Створення об'єднаного пула ресурсів з різноманітного устаткування, приміром, пристроїв збереження (SAN, NAS)	Надає можливість віртуалізації ІТ-ресурсів системи	Технології віртуалізації мають справу з однією системою, у той час, як GRID уможливає віртуалізацію значної кількості різноманітних ІТ-ресурсів

Виходячи з необхідності позиціонувати GRID-технологію, серед інших був запропонований перелік критеріїв, за якими можна визначати клас розподіленої системи, як такої, що належить до GRID-систем [6]:

- координація ресурсів, що не підлягають централізованому керуванню;
- використання стандартних відкритих протоколів й інтерфейсів загального призначення;
- надання різних рівнів якості послуг.

GRID-системи як засіб групового використання ресурсів

У своїй спробі подолати обмеження поодинокі обчислювальної системи, GRID набув великої популярності і скоріше за все стане важливою ключовою технологією для високопродуктивних обчислень у найближчому майбутньому. Необхідно також звернути увагу на те, що переважно всі економічно розвинуті країни вже декілька років мають національні програми досліджень та пілотні проекти застосування GRID, розглядаючи цей напрям як одну зі складових забезпечення національної безпеки.

Ідея GRID обчислень спершу походила з академічних й університетських кіл як базисна складова концепції електронної науки (e-science) і була підтримана в державних структурах. Програмні інструменти для спільного використання ресурсів для наукових досліджень передбачали об'єднання досвіду за рахунок спільного використання та візуалізації великих наборів наукових даних, а також об'єднання обчислювальних ресурсів і ресурсів збереження для аналізу таких даних, що вимагають великих обсягів розрахунків і серйозної функціональності [7, 8]. Тобто класифікація GRID-систем була започаткована, виходячи з того типу ресурсу, масштабування якого засобами мережі висувалось методом розв'язання тієї чи іншої наукової проблеми.

Так, за типом поділюваного ресурсу визначились:

- GRID обчислень (Compute Grids);
- GRID даних (Data Grids);
- GRID інструментальних засобів (Instrumentation Grids);
- GRID додатків (Application Grids).

GRID обчислень

До останнього часу потужні обчислення виконувалися переважно корпоративними обчислювальними центрами, де інтегрувалися високопродуктивні обчислювальні засоби, пристрої реєстрації (збереження), копіювання та друку тощо.

Хоча й існують розподілені обчислювальні структури (системи керування, резервування, служба доменних імен Інтернет), це, власне кажучи, системи, що займають свої спеціалізовані ніші.

Змужніння Інтернет і зростання потреб електронного бізнесу привели до розуміння того, що зовнішні мережі, ресурси і сервіси також відносяться до корпоративної IT-інфраструктури, що й привело до започаткування створення інтелектуальних корпоративних мереж, які з'єднали інформаційні центри традиційного тлумачення з граничними серверами (edge server). Рішення будувались на зразок корпоративного Web-сайту або сервера віртуальної приватної мережі, яка поєднує мережу підприємства із ресурсами постачальника послуг, що надають окремі чи групу Web-сервісів (пошта, кешування, буферизація інформації, балансування навантаження). Така спроба декомпозиції IT-сервісу визначалася внутрішніми можливостями корпоративної IT-інфраструктури.

Розроблене із застосуванням компонентних моделей програмування (Enterprise Java Beans, Net-компоненти) програмне забезпечення, що підтримує можливості міжплатформного перенесення, надало змогу перенести несуттєві функції (на зразок обробки запитів) на сервери проміжного рівня. Вимоги до швидкості обробки запиту для Web-доступу до корпоративних ресурсів зростають, і це змушує розподіляти і кешувати контент ближче до границі мережі. Кінцевим результатом такого процесу є декомпозиція високоінтегрованої IT-інфраструктури в гетерогенний набір фрагментів системи.

GRID даних

GRID орієнтовані дані — тип GRID-системи, призначений для рішення однієї з розповсюджених, але й дотепер складно розв'язуваних проблем: як забезпечити надійний і ефективний доступ до даних, що розподілені глобально й існують у різноманітних не порівнюваних програмних системах, таких як файл-сервери і бази даних. Ця неможливість простого доступу до даних веде до дарма витраченого часу і додаткових фінансових витрат на діяльність, на взірць розробки чи надання послуг. У промисловості, у фінансах і на виробництві вартість може бути вищою за просто додаткові витрати — це ще й неможливість конкурувати на ринку. У даному випадку GRID даних можуть у відомій мірі спростити доступ до даних, збільшуючи ефективність і знижуючи витрати.

GRID інструментальних засобів

Практичний досвід з багатьох прикладних областей показує, що швидких обчислень буває замало: часто необхідно в реальному часі збирати великі обсяги даних, що надходять з датчиків, аналізувати поточну ситуацію, виробляти рішення і генерувати керуючі дії. Все це вимагає тісної інтеграції керування, обробки даних різного виду, моделювання процесів, візуалізації даних в реальному часі. Обчислювальні комплекси такого роду отримали назву інтелектуальних інструментальних засобів.

Основна мета GRID інструментальних засобів полягає в тому, щоб істотно збільшити можливості реєстрації, організації, пошуку і забезпечення високоефек-

тивного і незалежного від географічного розміщення доступу до даних, що генеруються подібними інструментальними засобами. Такі дані зазвичай відносять до великих об'єктів даних (large data object — LDO). Екземпляри великих об'єктів даних найчастіше є результатом окремого операційного циклу роботи наукового приладу, цільової системи спостереження чи відображення, або суперкомп'ютера і характеризуються розміром від десятків Мб до десятків Гб. У багатьох областях людської діяльності вони є основним елементом сучасного середовища аналітичних досліджень.

Великі об'єкти даних — це найчастіше структурована колекція метаданих, що включає посилання на первинні дані. Метадані звичайно описують характеристики екземплярів об'єктів, забезпечують інформацію про методи доступу, керують доступом до інформації і т.д. Вони часто також вміщують похідні дані, що забезпечують, наприклад, візуальну інтерпретацію первинних даних.

Управління засобами масового збереження первинних даних здійснюється за рахунок виконання віддалених програм на Web-серверах. Для подальшого використання екземпляри об'єктів даних можуть бути перенесені до локальних дисків для доступу користувачів, за звичай, за допомогою Web-оглядачів, або вміщені до високопродуктивного кешу для доступу спеціалізованих додатків.

Характерним є те, що багато які із систем, що генерують такі об'єкти даних, використовуються різноманітним і географічно розподіленим співтовариством: у науці, приміром, це прискорювачі і системи різноманітних детекторів — приладів, що «фотографують» зіткнення; великі телескопічні системи; великі електронні мікроскопи; могутні джерела рентгенівського випромінювання тощо. Подібні комплекси спостереження і візуалізації, так само як і системи, що реєструють, існують також у галузі охорони здоров'я, входять до складу різноманітних систем моніторингу й т. ін. Важливість забезпечення і керування розподіленим доступом до систем збереження подібних об'єктів даних складається також у тому, що інфраструктури контрольно-вимірвальних лабораторій, лікарень, служб моніторингу тощо, часто є не кращим місцем для обслуговування великомасштабних цифрових систем збереження. Такі системи можуть одержати від GRID-технологій значну економію в масштабі у функціональному значенні, а легко доступна високопродуктивна мережа може забезпечити незалежність таких систем від місця розташування.

У будь-якій прикладній області, де дані генеруються у великих обсягах і з високою продуктивністю, і, особливо в розподіленому середовищі, де системи, що генерують дані, просторово відділені від систем, що здійснюють каталогізацію чи використання даних, є кілька важливих вимог для керування даними, що генеруються інструментальними системами:

- автоматичне породження хоча би мінімальних метаданих;
- автоматична каталогізація даних і метаданих негайно після отримання даних (чи наскільки це є можливо наближеним до реального часу);
- прозоре керування зовнішніми системами збереження, де заархівовані оригінальні (первинні) дані;
- допомога в проведенні кооперативних досліджень шляхом надання зазначеним локальним і віддаленим користувачам як негайного, так і довгострокового доступу до даних;

— включення даних в інші бази даних чи документи.

GRID додатків

GRID-додатки відрізняються від звичайних мережових додатків більшим спрямуванням на бізнес-цілі та потужнішою науковістю рішень. В якості класичного прикладу GRID-додатку можна розглядати Semantic Grid, що мав би відтворювати бізнес-процес отримання нових знань «Дані → Інформація → Знання» у будь-якому проблемно-орієнтованому середовищі.

GRID-технології у даний час розглядаються чи вже розгорнуті для численних додатків, включаючи:

- додатки для фармацевтичних компаній, що аналізують складні послідовності ДНК для розробки високоефективних і більш безпечних ліків;
- додатки для виробників автомобілів, що моделюють автомобільні аварії, оптимізують час виконання проекту й удосконалюють безпеку автомобілів;
- додатки для проектування літаків, що оптимізують аеродинамічні характеристики дизайну, в особливості, форму крила;
- додатки для фінансового, страхового ринку і ринку нерухомості, що виконують «що якщо» сценарії для портфелів клієнта при кожному ринковому коливанні, вирішують задачі керування ризиками;
- додатки для енергетичних компаній, що оптимізують рішення для планувальників параметрів передачі й розподілу електроенергії в електричних мережах;
- додатки для нафтових компаній, що аналізують величезну кількість геологічних даних, щоб визначити найперспективніші місця буровлення і т. ін.

Особливу цінність здобувають такі додатки за необхідності аналізу критичних даних у реальному часі (EnFuzion компанії Axceleon). Повідомляється про збільшення швидкодії окремих аналітичних додатків у результаті додавання до них GRID-шару в 10000 разів.

Напрямки використання GRID

Серед основних напрямків використання GRID у науковій сфері на даний момент виділяють:

- розподілені суперобчислення, рішення дуже великих задач, що вимагають величезних процесорних ресурсів, пам'яті тощо;
- потужні обчислення (High-Throughput Computing), що дозволяють організувати ефективне використання ресурсів для невеликих задач, використовуючи комп'ютерні ресурси, що звільнюються та тимчасово простоюють;
- обчислення «за вимогою» (On-Demand Computing), великі разові розрахунки;
- обчислення із залученням великих обсягів розподілених даних (Data-Intensive Computing), наприклад, у метеорології, астрономії, фізиці високих енергій;
- колективні обчислення (Collaborative Computing).

Перший напрямок застосування GRID — запуск традиційних, але ресурсно-містких програм і додатків на адекватних обчислювальних потужностях. Уже пе-

рші спроби використання для цих цілей традиційних мережевих технологій показали високій ступінь впливу інфраструктури на можливості реалізації концепції високопродуктивних розподілених обчислень [17].

Другий напрямок застосування — виконання паралельних програм на процесорах, встановлених в окремі комп'ютери. Обмін даними між процесорами забезпечується не спеціально розробленими каналами зв'язку, як у суперкомп'ютерних архітектурах, а локальними і глобальними мережами серійного і, завдяки цьому, не такого вартісного устаткування. Підтримка розподілених паралельних процесів є головною рисою, що відрізняє технології метакомп'ютингу від звичайних сьогодні клієнт-серверних обчислень, де виклики процедур відбуваються за синхронним принципом. Такий напрямок застосування здається найцікавішим тому, що дозволяє отримати еквівалентні обчислювальні потужності без мільйонних вкладень у суперкомп'ютери.

Зусилля багатьох дослідників в області розподілених обчислень спрямовані на те, щоб довести можливість ефективної роботи паралельних додатків в умовах ненадійних мереж групового використання.

Проблема, однак, полягає не тільки в мережах. Основною проблемою залишається створення паралельних програм. З появою таких інструментальних засобів як PVM [9], адаптованої для GRID та повної версії MPI для різномірних обчислювальних платформ [10, 11], паралельних версій мов С [12] і Fortran [13], ситуація покращилася, проте розробка і налагодження програм з паралельними процесами все ще надзвичайно ускладнена. Це й є ще однією з перешкод розширення сфери застосування метакомп'ютерних мережевих обчислень.

Іншим фактором, що стримує розвиток метаобчислень, вважають відсутність технологій програмування, позбавлених властивостей відтворення інформаційного шуму. Один з підходів вирішення цієї проблеми пропонує принципи і способи формування в ресурсах мережі єдиного, математично однорідного поля комп'ютерної інформації. Він спрямований на випереджуваче створення засобів програмування глобально-розподілених обчислень й інформаційно-обчислювальних систем масового використання [14, 15].

Зараз у багатьох великих університетських центрах Америки, Європи й Азії ведуться роботи зі створення інфраструктур із застосуванням GRID-технологій.

Так, нещодавно було повідомлено про оголошення Міністерством освіти Китаю про відкриття спільного з IBM проекту на підтримку створення інфраструктури метаобчислень, що поєднає понад 200 тис. персональних комп'ютерів у 100 Університетах країни [16]. Стрімкий розвиток досліджень у галузі GRID-обчислень вже дозволяє вести мову про третю генерацію подібних систем.

До третього покоління відноситься нинішня версія проекту Globus, що є розвитком одного з перших проектів у цій області I-WAY [17]. Це — географічно розподілена обчислювальна мережа. Центральним елементом проекту є інструментарій класу middleware Globus Toolkit™ [18]. Його остання версія — GT3 — є першою повномасштабною реалізацією нової специфікації Open Grid Services Architecture (OGSA), у визначенні якої розробник інструментарію Globus Alliance зіграв чималу роль.

Ще один приклад — проект Legion [19], побудований як об'єктно-орієнтована метасистема, що дозволяє користувачам зі своїх робочих станцій одержу-

вати доступ до інтегрованої інфраструктури незалежно від масштабу, фізичного розташування, мови й операційних систем. Проект Legion був представлений широкому науковому загалу ще у 1997 р. У 1998 р. була створена компанія Applied Metacomputing (з 2001 р. — Avaki), що почала комерційну експлуатацію ідей GRID-обчислень, зараз вона займає одне з перших місць у цій ніші.

На відміну від наукових GRID-обчислень, для бізнес-структур потужність комп'ютеру або систем збереження не є головною вимогою. У світі бізнесу скоріше матимуть місце такі вимоги до IT-інфраструктури, що ведуть до застосування GRID-технологій:

— ефективне використання існуючих ресурсів. Так, за даними META Group у 2002 р. ефективність використання UNIX та Windows-серверів не перевищувала 25 % протягом 24 годин;

— оптимальне та автоматизоване управління обчислювальним центром. Масштабування ресурсів шляхом збільшення кількості виділених серверів, підключення нових пристроїв і т. ін. збільшує навантаження на персонал, що веде до збільшення накладних витрат і, можливо, зниження якості обслуговування;

— управління відповідністю угодам про рівень послуг (service level agreement — SLA) та якість послуг, що надаються (quality of service — QoS);

— прискорення відновлення даних у разі скоєння лиха;

— поліпшення безперервної діяльності.

Найпершими кандидатами до застосування GRID-технологій у бізнес-сфері можна вважати наступні.

Індустріальний сектор наук про життя (геноміка, біоінформатика, протеоміка, обчислювальна хімія) — прискорення аналізу та поліпшення доступу до експериментальних даних шляхом забезпечення уніфікованого доступу до колекцій даних нестандартних форматів для винаходу нових та дослідження існуючих структур.

Фінансовий сектор — бізнес-аналітика й оптимізація використання IT-інфраструктури. Перше рішення допоможе розширити конкурентноздатність і збільшити активність компанії на фінансовому вторинному ринку, прискорюючи аналіз торгової діяльності і збільшуючи продуктивність обчислень. Друге — допоможе клієнтам експлуатувати доступні, недозавантажені обчислювальні ресурси і пристрої збереження. Іншими застосуваннями можуть бути рішення в сфері управління ризиками та розподілу активів.

Автомобільна та аерокосмічна промисловість — прискорення розробки нових видів продукції за рахунок управління витратами в реальному часі та оптимізації інвестицій у розробку в існуючій інфраструктурі, підтримка співпраці під час розробки.

Урядування — прискорення доступу до інформаційних ресурсів, максимальне використання існуючих інформаційних ресурсів, добування даних та зіставлення даних, що походять з різних джерел з метою поліпшення прогнозування та упереджуючого реагування на негативні обставини.

Розвиток і впровадження GRID-технологій

Сьогодні активність у напрямку розвитку й впровадження GRID-технологій вражає.

Вже розроблені і застосовуються принципово нові інструменти:

- наукові портали, що дозволяють викликати по мережі пакети, спеціалізовані на рішення визначених задач;
- засоби для організації розподілених обчислень, що збирають для рішення задач десятки тисяч комп'ютерів;
- засоби для аналізу надвеликих обсягів даних, що можливий завдяки об'єднанню засобами GRID-систем збереження;
- засоби об'єднання великих обчислювальних ресурсів для обробки експериментів у режимі он-лайн.

Про активність у галузі розвитку і впровадження нової технології свідчить кількість проектів, що відкриті і відкриваються в світі. Звичайно, поштовхом є вільно розповсюджуваний інструментарій Globus Toolkit™, що став фактичним стандартом конструювання GRID-систем. На його основі створені численні проекти: від систем забезпечення роботи наукових груп, яким потрібно віддалений доступ до спеціалізованих експериментальних лабораторій, наприклад, Network for Earthquake Engineering Simulation, до систем розподіленої аналітичної обробки великих обсягів інформації, таких як Grid Physics Network, EU Data Grid Project або Particle Physics Data Grid. Прикладом комерційного застосування Globus Toolkit™ стали продукти фірми ІВМ. У табл. 2 наведений перелік відомих наукових проектів США та країн Європи, в основу яких було покладено Globus Toolkit™.

Таблиця 2. Проекти, що застосовують інструментарій Globus Toolkit™

Найменування проекту	Мета	URL, джерело фінансування
Access Grid	Створення і поширення систем співробітництва з використанням новітніх технологій	http://www.accessgrid.org ; US Department of Energy, National Nuclear Security Administration, далі — U.S. DOE), NSF
BlueGrid	Створення єдиної GRID для потреб тестування для лабораторій ІВМ	ІВМ
Distance and Distributed Computing and Communication (DisCom ²)	Створення реальної GRID для доступу до ресурсів для 3-х національних лабораторій, що працюють в галузі оборони	http://www.cs.sandia.gov/discom ; U.S.DOE Defense Programs
DOE Science Grid	Створення реальної GRID для доступу до ресурсів для U.S. DOE, наукових лабораторій і університетів	http://www.sciencegrid.org ; U.S.DOE Office of Science
Earth System Grid (ESG)	Створення інфраструктури для віддаленого доступу й аналізу даних моделювання клімату	http://www.earthsystemgrid.org/ ; U.S.DOE Office of Science
EuroGrid, Grid Interoperability (GRIP)	Демонстрація переваг використання GRID для окремих наукових та індустріальних співтовариств з урахуванням їх специфічних потреб	http://www.eurogrid.org ; European Union

Найменування проекту	Мета	URL, джерело фінансування
European Union (EU) DataGrid	Створення реальної GRID для забезпечення високошвидкісної обчислювальної інфраструктури та аналізу розподілених баз даних дуже великого масштабу (сотні Терабайтів до Петабайтів).	http://www.eu-datagrid.org ; European Union
Fusion Collaboratory	Створення національного співробітництва для термоядерних досліджень	http://www.fusiongrid.org ; U.S.DOE Office of Science
Globus Project™	Дослідження в області GRID-технологій; створення і технічна підтримка Globus Toolkit™, додатків.	http://www.globus.org ; U.S. DARPA, DOE, NSF, NASA, Msoft
Grid Application Development Software	Дослідження і створення технологій для GRID-додатків	http://www.hipersoft.rice.edu/grads ; U.S.NSF
Grid Physics Network	Створення технологій для аналізу даних у фізиці: ATLAS, CMS, LIGO, SDSS	http://www.griphyn.org/ ; U.S.NSF
Grid Research Integration Dev. & Support Center	Створення і підтримка проміжної GRID-інфраструктури	http://www.grids-center.org ; U.S.NSF
GridLab	GRID-технології і додатки	http://www.gridlab.org ; European Union
GridPP	Створення реальної GRID в Англії для досліджень в області фізики елементарних часток .	http://www.gridpp.ac.uk ; U.K. e-Science
Information Power Grid	Створення реальної GRID для аеродосліджень NASA	http://www.ipg.nasa.gov ; U.S.NASA
International Virtual Data Grid Laboratory	Створення реальної міжнародної GRID для експериментів над GRID-технологіями і додатками	http://www.ivdgl.org ; U.S.NSF
Network for Earthquake Engineering Simulation Grid (NEESgrid)	Створення реальної GRID для досліджень у сейсмології	http://www.neesgrid.org ; U.S.NSF
Particle Physics Data Grid	Створення реальної GRID для аналізу даних у фізиці високих енергій і ядерній фізиці	http://www.ppdg.net ; U.S.DOE Office of Science

Інструментарій Globus Toolkit™ включає такі основні компоненти:

- побудований на основі HTTP протокол резервування обчислювальних ресурсів GRAM (Globus Toolkit Resource Allocation Management);
- розширену версію протоколу для передачі файлів GridFTP;
- службу аутентифікації і безпеки GSI (Grid Security Infrastructure);

- розподілений доступ до інформації на основі протоколу LDAP;
- віддалений доступ до даних через послідовний і паралельний інтерфейс GASS (Globus Access to Secondary Storage);
- резервування ресурсів GARA (Globus Reservation and Allocation).

Впровадження технологій метаобчислень в Україні

Зусилля вітчизняної наукової спільноти в галузі застосування розподілених обчислень спрямовані в тому чи іншому ступені в усіх згаданих вище напрямках. Популярним методом в Україні, наприклад, стало створення потужних кластерів шляхом поєднання звичайних одно- та двопроцесорних ПК мережею під керуванням операційної системи Linux [20, 21]. В Інституті проблем реєстрації інформації НАН України, вивченню цих питань приділяється значна увага, починаючи з 2000 р. Певний інтерес науковців викликає застосування GRID-технологій у справі створення інфраструктури корпоративної співпраці для вирішення широкого спектру міждисциплінарних проблем. Можливо, деє вже виконуються проекти із застосування чи створення елементів подібних інфраструктур. Проте у вітчизняному інформаційному просторі існує якщо не роздвібненість відомостей про такі роботи, то повна їх відсутність.

Світовий досвід довів, що для реалізації подібних проектів за відсутності підтримки відповідної інфраструктури потрібні героїчні зусилля розроблювачів [17].

Одним з найбільш важливих ресурсів технології є лінії зв'язку. За оцінками російських фахівців, що включилися в міжнародний GRID-проект по лінії CERN, оптимальним рівнем пропускної здатності ліній зв'язку для участі в проекті за станом на 2003 р. можна вважати 100–155 Мбіт/с на внутрішніх і міжнародних каналах відповідно [22].

Побудова національної GRID-інфраструктури — не тільки питання забезпечення організації мережі й обчислювальних апаратних засобів. Проблема об'єднання цих ресурсів належним чином — все ще відкрите питання для досліджень. Існуюче на сьогодні програмне забезпечення доступне, але знаходиться в експериментальній стадії, неповне і частиною було розроблене до прийняття яких-небудь стандартів у цій області. Для розвитку наступних поколінь GRID-компонентів необхідно виконати ще багато роботи. Серед основних напрямків перспективних досліджень можна навести наступні:

- наскрізне керування ресурсами і методи адаптації, що здатні забезпечувати гарантоване виконання на рівні додатку, незважаючи на динамічний характер використання ресурсів;
- автоматизовані методи узгодження використання ресурсів, політик і розрахунків у великомасштабних GRID-середовищах;
- високоефективні методи зв'язку і протоколи;
- інфраструктура і підтримка інструментарію інформаційно насичених додатків, передових концепцій телеприсутності і нових методів проблемно-орієнтованих середовищ.

GRID-технологія надає реальну основу й якісно новий рівень інтеграції для рішення за допомогою обчислювальних мереж глобальних інтерактивних задач інформаційного забезпечення, управління і самоорганізації. І Україні на сьогодні

така технологія конче потрібна. Вона би надала реальний інструментарій для вирішення проблем, що постали перед державою: єдині інформаційні служби міст, регіонів з питань охорони здоров'я й екології, платежів, супроводу бюджету, казначейського обліку, загальнодержавного технагляду, митного контролю, електронні бібліотеки, електронна наука.

Це — нагальні завдання, проте володіння технологіями на зразок GRID допоможе винайти оптимальні рішення. Тут можуть бути знайдені ефективні рішення і для завдань, що постали перед вітчизняним бізнесом: інтеграція мобільної телефонії з Інтернет, постачання відео за замовленням, діловодства, бізнесу, торгівлі, логістики. Не обійде GRID-технологія й освітянську ниву, де буде надане підґрунтя для організації дистанційного навчання, надання доступу до електронних книг, юридичних і довідкових систем.

Висновок

GRID-технології уже заповнили світ. Зараз склалися найкращі умови для участі в загальному русі: тексти розробленого ПО зазвичай є відкритими, в Інтернеті доступні публікації розроблювачів і документація із застосування наукових розробок, можна брати участь в обговоренні проблем у чатах та електронною поштою. За таких сприятливих умов Україна повинна обов'язково приєднатися до світової GRID-спільноти.

У сучасних умовах, на нашу думку, потрібні негайні зусилля з організації та координації досліджень та започаткування проектів у напрямку застосування GRID-технологій в Україні як базису для створення ефективних інфраструктур для корпоративних обчислень найближчого майбутнього. Відсутність координації робіт учених і фахівців у даному напрямку, що дуже динамічно розвивається в інших країнах світу, недостатність і розрізненість публікацій у вітчизняних інформаційних виданнях може привести в найближчому майбутньому до величезного технологічного запізнювання і нераціонального використання інвестицій, що зараз вкладаються у розвиток вітчизняних IT-інфраструктур.

1. *Allen F. et al.* BlueGene: A Vision for Protein Science Using a Petaflop Computer // IBM Systems J. — 2001. — Vol. 40, N 2. www.research.ibm.com/journal/sj/402/allen.html

2. *Sterling T.* The Gilgamesh MIND Processor-in-Memory Architecture for Petaflops-Scale Computing // ISHPC Conference. — Kansai (Japan). — 2002, May. www.sc-2002.org/paperpdfs/pap.pap105.pdf

3. *Орлов С.* Искусство объединения // LAN. — 2003. — № 9.

4. *Foster I., Kesselman C. and Tuecke S.* The Anatomy of the Grid: Enabling Scalable Virtual Organizations // International. J. High Performance Computing Applications. — 2001. — 15(3). P. 200–222. www.globus.org/research/papers/anatomy.pdf.

5. *Christensen E. et al.* Web Services Description Language 1.1. — W3C Note. — 2001, 15 Mar. <http://www.w3.org/TR/wsdl>.

6. *Foster I.* What is the Grid? A Three Point Checklist. — 2002, 20 Jul. <http://www.globus.org/research/papers/WhatIsTheGrid.pdf>

7. *Catlett C., Smarr L.* Metacomputing // Comm. ACM. — 1992, Jun.

8. Foster I., Kesselman C., eds. The Grid: Blueprint for a New Computing Infrastructure. — San Francisco: Morgan Kaufmann, 1999.
9. PVM. http://www.csm.ornl.gov/pvm/pvm_home.html.
10. Foster I. and Karonis N. A Grid-Enabled MPI: Message Passing in Heterogeneous Distributed Computing Systems // Proceedings of SC'98. — ACM Press. — 1998.
11. MPICH-G2. <http://www3.niu.edu/mpi/>
12. Параллельные расширения и диалекты языков C и C++. http://parallel.ru/tech/tech_dev/cc++.html
13. Параллельные расширения и диалекты языка Fortran. http://parallel.ru/tech/tech_dev/fortran.html
14. Затуливетер Ю.С. Информация и эволюционное моделирование // Труды Междунар. конф. «Идентификация систем и задачи управления» (SICPRO'2000). — М: Ин-т проблем управления РАН. — 2000, 26-28 сентября. — С. 1529–1573. http://zvt.hotbox.ru/1529_htm; http://zvt.by.ru/1529_htm
15. Затуливетер Ю.С. О компьютерных проблемах формирования единого информационного пространства виртуальных предприятий. Системы проектирования, технологии подготовки производства и управления этапами жизненного цикла промышленного продукта (CAD/CAM/PDM-2002) // Материалы 2-й Междунар. конф. (CAD/CAM/PDM-2002) В 2 т.— Том 1. — М.: Ин-т проблем управления РАН. — 2002. — С. 165–176. [Z3okl_1.htm](http://zvt.by.ru/Zatuliveter_dokl_1.htm); http://zvt.by.ru/Zatuliveter_dokl_1.htm
16. IBM. China's Ministry of Education Launch «CHINA GRID» // Grid Today. — 2003. — Vol. 2, N 42. <http://www.gridtoday.com/03/1020/102130.html>
17. Foster I., Geisler J., Nickless W., Smith W., Tuecke S. Software Infrastructure for the I-WAY High Performance Distributed Computing Experiment / Proc. 5-th IEEE Symposium on High Performance Distributed Computing. — 1997. — P. 562–571. <ftp://ftp.globus.org/pub/globus/papers/isoft.pdf>
18. The Globus Toolkit. <http://www-unix.globus.org/toolkit/>
19. Legion: A Worldwide Virtual Computer. <http://legion.virginia.edu/index.html>
20. ИВЦ Киевского национального университета им. Т. Шевченко. <http://www.cluster.kiev.ua/>
21. Институт физики конденсированных систем НАН Украины. <http://www.icmp.lviv.ua/cluster/>
22. Ильин В., Кореньков В., Солдатов А. Российский сегмент глобальной инфраструктуры LCG // Открытые системы. — 2003. — № 1.
23. Українська науково-освітня мережа УРАН. <http://www.cei.uran.net.ua:8101/ukr/uran.htm>
24. Факультет ВТ НТУ України «КПІ». <http://www.comsys.ntu-kpiu.kiev.ua/>

Надійшла до редакції 18.03.2004