

УДК 004.75

СТВОРЕННЯ НАЦІОНАЛЬНОЇ GRID- ІНФРАСТРУКТУРИ ДЛЯ ЗАБЕЗПЕЧЕННЯ НАУКОВИХ ДОСЛІДЖЕНЬ

**Згуровський М.З., Петренко А.І., Кисельов Г.Д.
НТУУ „Київський політехнічний інститут”**

Надано огляд робіт зі створення освітянського сегменту національної Grid-інфраструктури для використання в освітніх і дослідницьких цілях, зокрема, для навчання користувачів Grid.

Ключові слова: Grid-інфраструктура, кластер, Сертифікаційний центр, портлети.

Вступ

Проект створення національної Grid – інфраструктури для розвитку інформаційного суспільства в Україні, її склад і задачі вперше були озвучені на саміті WSIS (World Summit on Information Society) у 2004 році [1,2]. Мова йшла про необхідність створення освітнього і дослідницького сегменту інформаційного суспільства України з двома головними напрямками: широке використання інформаційних і комунікаційних технологій на всіх стадіях наукових досліджень і освіти, а також інформаційне управління відповідними галузями. Головні задачі першого напрямку включали:

- 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 й експлуатації розподілених інформаційних систем, розробка алгоритмів і методів рішення прикладних задач в середовищі розподіленого комп'ютингу.

Другий напрям передбачав:

- Впровадження ефективного інформаційного управління освітою і наукою.
- Розробка засобів для зберігання, обробки і відкритого доступу до наукових та освітніх інформаційних ресурсів (бази даних, архіви, електронні бібліотеки і т.п.).
- Підключення українського Grid до Європейського Grid, що створюється згідно проекту EGEE, і до мережі Світових Центрів Даних.

Запропоновані цілі та задачі знайшли своє відображення вже у 2005 році в *Державній цільовій програмі «Інформаційні та комунікаційні технології в освіті й науці на 2006-2010 роки»*, прийнятій Кабінетом Міністрів України (Постанова КМ України №1153 від 7 грудня 2005 року). Національний технічний університет «Київський політехнічний інститут» виграв тендер проектів з реалізації завдання *«Створення національної Grid-інфраструктури для підтримки наукових досліджень»* цієї Державної цільової програми, оголошений Міністерством освіти і науки України (договір № ІТ /506 – 2007 від 22 серпня 2007 р.).

Проект Ugrid (www.grid.ntu-kpi.kiev.ua) присвячено побудові сегменту МОНУ 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 вагомим наукових застосувань, які використовуються в Grid – середовищі[3,6].

Результати виконання проекту

У створенні національної Grid – інфраструктури приймали участь шість університетів МОН України два інститута НАНУ два промислових. Головною організацією виступив Інститут Прикладного Системного Аналізу (ІПСА) Національного Технічного Університету

«Київський політехнічний Інститут». Науковим керівником проекту є Михайло Захарович Згуровський, ректор НТУУ «КПІ», академік НАНУ.

Серед виконавців проекту, крім НТУУ «КПІ», є Інститут проблем моделювання в енергетиці імені Г.Є. Пухова НАНУ (ІПМЕ), Харківський національний університет радіоелектроніки (ХНУРЕ), Львівський національний технічний університет «Львівська політехніка» (НУЛП), Запорізький національний технічний університет (ЗНТУ), Донецький національний технічний університет (ДонНТУ), Дніпропетровський національний гірничий університет (ДНГУ), державне підприємство «Львівський науково-дослідний радіотехнічний інститут» (ЛНДРІ) і підприємство ЮСТАР.

Створений Grid – сегмент МОНУ (на відміну від Grid – сегменту НАНУ [8], який є Grid обчислювального типу – Computing Grid) за своїм призначенням відноситься до Grid інформаційного типу (Data Grid). Завдання на виконання проекту Ugrid включало забезпечення обслуговування *Української філії Світового Центру Даних* (УФ СЦД), надаючи його клієнтам віддалений доступ до світових сховищ наукових даних, можливості ефективного сумісного використання комп'ютерів, унікальних експериментальних установок і приладів (www.wdc.org.ua).

На першому етапі проекту проведено аналіз концепції, архітектури та ресурсів сучасних Grid, систематизовано відомості про методи побудови архітектури розподілених систем з використанням технології Grid. Побудовано Grid – інфраструктуру з 6-тю регіональними центрами (в Києві, Харкові, Донецьку, Дніпропетровську, Запоріжжі і Львові) і розпочато віддалене обслуговування майбутніх користувачів-науковців з університетів та наукових установ України. Проведено порівняльний аналіз відомих пакетів організації географічно розподілених обчислювальних мереж (*проміжного Grid програмного забезпечення*), на основі якого зроблено пропозицію використати у створенні Grid-середовища програмне забезпечення gLite, яке має широкий набір сервісів, підтримує сервісно-орієнтовану архітектуру, широко застосоване у європейських проектах, має достатні можливості з модернізації та створенню додаткової функціональності.

На другому етапі проекту проведено підключення кластерів ІТФ НАНУ і НТУУ «КПІ» на основі ПГЗ NorduGrid для вивчення технологій віртуалізації обчислювальних ресурсів країни з одночасним підключенням до кластеру НТУУ «КПІ» 5-ти регіональних центрів. У січні 2008 року розпочав діяльність *Сертифікаційний центр* відкритих ключів для надання українським користувачам доступу до європейських ресурсів і сховищ даних. (www.ca.ugrid.org). Цей центр обслуговує зараз користувачів обох сегментів Grid-інфраструктури України. Розроблено *портал доступу SDGrid* (www.sdgrid.org.ua) до Grid – інфраструктури як єдиної точки доступу користувачів до різномірних інформаційних ресурсів і програм. Grid портал являє собою програмно-апаратний комплекс створений з використанням бібліотек GridSphere 3 і Vine Toolkit, основними функціями якого є постановка завдань на обчислення у Grid-інфраструктурі, моніторинг стану обчислювальних ресурсів та ресурсів збереження даних, керування ресурсами та політиками безпеки Grid – інфраструктури, навчання користувачів, надання доступу до інших Grid систем, побудова віртуальних організацій та інше. Безпека доступу до Grid-інфраструктурі з порталу забезпечується за допомогою використання GSI (Grid Security Infrastructure), підтримці HTTPS/SSL (Secure Sockets Layer), шифрування даних на всіх рівнях.

Позитивною рисою порталу можна відзначити можливість інтеграції з різними популярними пакетами проміжного програмного забезпечення. Для цього в порталі SDGrid використовуються портлети Grid Portlets, Interactive Job Submission Portlet, Batch Job Submission Portlet, CONDOR-G Job Submission Portlet. Основним портлетом є GridPortlets. Він доступний для скачування з репозиторія на офіційному сайті GridSphere.

Портал надає можливість роботи з файлами (за допомогою GRIDFTP) на видалених ресурсах. Так, для цього в порталі SDGrid використовуються наступні портлети: File Transfer Portlet, Comprehensive File Management Portlet, Comprehensive File Management Portlet, Srb-portlet. Ці портлети також зв'язуються з портлетами Grid Portlets і Proxy Manager Portlet і

перевіряють наявність активних сертифікатів. За наявності таких, портлети надають можливість користувачам проглядати списки файлів на видалених ресурсах, завантажувати файли, а також пересилати файли між видаленими ресурсами. Так, наприклад, Srb-portlet (Storage Resource Broke) є системою послуг, які об'єднують різноманітні ресурси зберігання (наприклад, баз даних, робочих станцій, архівних матеріалів і систем NFS) з віртуальним інтерфейсом.

Ще однією важливою складовою портлалу є наявність засобів моніторингу виконуваних робіт на видалених ресурсах. На поточний день наймогутнішим засобом для цього є портлет GPIR TGUP Browser Portlet. GPIR надає простір для зберігання даних про Grid, які легко доступні для додатку порталу. Він об'єднує як динамічні дані, так і статичні дані, наприклад, контактну інформацію і місце розташування ресурсу декількома способами:

- динамічні дані оновлюються за допомогою веб-сервіса GPIRIngester, а отримання даних відбувається за допомогою сервісу GPIRQuery;
- статичними даними можна управляти через клієнт GPIR Administration, який є звичайним веб-приложением з доступом через браузер.

Щоб надати користувачам порталу можливість спілкування і обговорення проектів в порталі створено форум, чат, блог і так далі. Зручним засобом для цього є BlueSquid Collaboration Group Portlet. Він вільно розповсюджується і доступний для скачування з репозиторія на офіційному сайті GridSphere. Портлет надає можливість для створення робочого середовища засобами порталу і сприяє обговоренню ідей, пов'язаних з проектом.

На базі розробленого порталу SDGrid розроблена *віртуальна лабораторія* для демонстрації можливостей технологій Grid, в склад якої входять п'ять територіально-рознесених серверів: Web-сервер, сервер MyProху, сервер СА (Сертифікаційний центр), НРС-сервер обчислювального ресурсу, сервер моніторингу і резервного копіювання, призначений для збору статистики про стан серверів учбової Grid-системи. За допомогою цієї віртуальної лабораторії можна отримати перший досвід роботи в Grid-середовищі. Цей досвід включає реєстрацію і отримання сертифікату MyProху, переглядання файлів у каталозі, навігацію в межах каталогу, формування і запуск завдань, відстеження стану їх виконання тощо.

Виконавцями проекту досліджена методологія використання Grid – технологій в вищій школі і створена експериментальна *навчальна програма* з Grid курсу; розпочата магістерська підготовка з напрямку «Grid технології в науці і освіті». Підготовлений і вийшов з друку навчальний посібник «Вступ до Grid технології в науці і освіті» (автор – проф. А.І. Петренко), розпочата підготовка відповідного підручника спільним авторським колективом фахівців МОНУ і НАНУ.

Вклад виконавців проекту

Всі організації – учасники проекту (крім обов'язків з організації регіональних центрів, забезпечення їх функціонування, підготовки і підтримки користувачів) проводять наукові дослідження в галузі сумісності (interoperability) проміжного програмного слою різних Grid-інфраструктур, забезпечення наскрізної інформаційної безпеки при об'єднанні національних Grid-інфраструктур чи їх сегментів, розробки додаткових сервісів для розширення кола можливих користувачів (окрім науки) на представників різних прошарків суспільства: інженерії, бізнесу і соціальної сфери тощо.

Докладніше дослідження, що проводяться, виглядають так:

1. Інститут прикладного системного аналізу НТУУ"КПІ"

- Створення Grid-порталу та порталу знань.
- Дослідження сумісності (interoperability) проміжного програмного слою Grid-інфраструктури.
- Створення Grid-дodatка для моделювання сучасних мікро-електромеханічних систем (MEMS).
- Створення тримовного інформаційного сайту проекту (www.grid.ntu-kpi.kiev.ua), інформаційний ресурс якого вже перевищив 90 Мб, на якому регулярно

розміщується і обновляється інформація, яка висвітлює хід виконання проекту і містить також навчальні матеріали.

Розпочате реальне об'єднання існуючих сегментів наукової і освітньої обчислювальної і комунікаційної інфраструктури НАНУ і МОНУ в єдину **Українську Національну Grid – ініціативу (UNGI)**, проект якої був підготовлений в серпні 2007 року і **прийнятий як** складова європейських провідних проектів **EGEE** (Enabling Grids for E-science) і **EGI** (European Grid Initiatives) [7].

2. НТУ Львівська політехніка.

- Створення потужної системи збереження даних на базі з використанням системи IBM BladeCenter QS21, що складається з 14 обчислювальних блейд-серверів на базі процесорів Cell та одного координуючого вузла.
- Впроваджуються заходи з навчання користувачів Grid-мережі в навчальному процесі НТУ «Львівська політехніка».

3. ЗНТУ, Запорізький національний технічний університет.

- Відпрацьована технологія підключення ресурсно-операційних центрів до ресурсів кластера НТУУ "КПІ", виконано навчання місцевих користувачів засобам віддаленої компіляції та запуску паралельних додатків.
- Розроблений паралельний додаток для моделювання нейронних мереж, який було використано для системи прогнозування часових рядів. Ефективність цієї системи перевірена на прикладі моделювання динаміки ВВП, зокрема експериментально визначена найкраща конфігурація паралельної системи для рішення цієї задачі.
- Розроблена і досліджена паралельна реалізація метода молекулярної динаміки для моделювання процесу іонного бомбардування поверхні кристалу міді. Ефективність реалізації даного методу перевірена для декількох сотень і декількох тисяч атомів. Показано, що зі зростом складності моделі (кількості атомів, що моделюються) ефективність реалізації значно залежить від характеристик комунікатора, який використовується для обміну даними між процесорами кластеру.
- Запропоновані підходи до оптимізації резервування розподілених обчислювальних ресурсів в мережах GRID, засновані на застосуванні модифікації точного алгоритму пошуку найбільшого підграфа.

Розроблені заходи дозволяють забезпечити підтримку користувачів і віртуальних організацій в Запорізькому ресурсно-операційному GRID-центрі.

4. ХНУРЕ, Харківський національний університет радіоелектроніки.

- Програмне забезпечення для тестування продуктивності бібліотеки PVM.
- Дослідження схеми обробки завдань у GRID шляхом використання платформи NorduGrid. Для вирішення поставленого завдання були застосовані операційна система Scientific Linux, програмні пакети Globus Toolkit та NorduGrid.
- Дослідження GRID-задач в комп'ютерних науках та інженерії. У процесі виконання роботи був проведений аналіз існуючих проектів та систем в області розподілених Грід обчислень, проведений огляд актуальних задач, які вирішуються у різних галузях науки і техніки з використанням технологій метакомп'ютингу, а також виконана реалізація тестової задачі на основі Грід-технологій.
- Дослідження схеми обробки завдань в GRID шляхом використання платформи X-Com. Для вирішення поставленого завдання застосовувалися обчислювальний кластер з встановленою на нього системою метакомп'ютингу X-Com. Програмне забезпечення розроблене з використанням середовища розробки Microsoft Visual Studio 2008.

5. Приватне науково-виробниче підприємство «Юстар».

- У дослідженні розглянуто нову парадигму обробки наукових даних – Data-Intensive Computing, дано опис світових тенденцій розвитку апаратного

забезпечення для високопродуктивних обчислень. Також наведено опис національних програм розвитку наукової обчислювальної інфраструктури у світі та представлено міжнародні проекти в галузях GRID та eScience;

- Викладено висновки та рекомендації щодо розвитку інфраструктури наукових обчислень в Україні відповідно до світових тенденцій.

6. ІПМЕ ім. Г.Є. Пухова НАНУ.

- Розроблена архітектура та проведено вибір технічних засобів для обчислювального кластера Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України.
- Розроблено концепцію подальшої модернізації обчислювального кластеру, що створює умови для забезпечення наукових досліджень в інституті новітніми інформаційними технологіями.
- Проведені роботи з установки, настроювання і підключення до Grid-сегменту України обчислювального кластеру Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України; сформульовані вимоги до системного програмного забезпечення та проведена його інсталяція.
- В рамках інфраструктури безпеки грід отримано цифровий сертифікат для обчислювального ресурсу ІПМЕ НАНУ та користувачів. Розроблені інструкції та методичні рекомендації з підтримки користувачів грід-вузлів, зокрема, ресурсно-операційного центру ІПМЕ.

7. ДНГУ, Дніпропетровський Національний гірничий університет.

- Досліджені обчислювальні задачі підприємств гірничо-металургійної галузі. Їх технологічні процеси характеризуються як багатовимірні системи, що знаходяться під впливом збурень які в переважній більшості випадків змінюються на стохастичні, в яких ступінь впливу є не постійним. До того є досить складним процес вимірювання параметрів, що є ґрунтуючими для технологічного процесу видобутку, збагачення та промислового перетворення корисних копалин.

8. ДонНТУ, Донецький Національний технічний університет.

- Розроблено структуру донецького регіонального ресурсного Центру Grid і створена постійно діюча група з питань підтримки, експлуатації і управління Центру. Сформовано коло користувачів Grid.

Особлива увага в дослідженнях приділяється учасниками проекту забезпеченню підтримки віддаленого спілкування вчених в реальному часі для реалізації засад е-науки; технологіям інтелектуального аналізу даних з метою виявлення прихованих закономірностей у вигляді значущих особливостей, кореляцій, тенденцій і шаблонів (DataMining); використанню метаданих і онтологій, у яких інформація розуміється не тільки як дані, що мають значення, але і знання, які здобуваються (Семантичний Grid).

Заключення

23 вересня 2009 року Кабінет Міністрів України прийняв постанову про затвердження державної цільової науково-технічної програми "Впровадження і застосування грід-технологій на 2009-2013 роки". Програмою передбачені завдання з розроблення і впровадження Grid технологій, розвитку матеріально-технічної бази Grid-інфраструктури, забезпечення інтеграції до європейської і світової Grid-інфраструктур, організаційного та методичного забезпечення підготовки фахівців з питань впровадження і застосування Grid-технологій. Питання наповнення національної Grid – інфраструктури засобами формування даних, їх збереження та використання для створення баз даних та знань потребують розвитку методів ефективного використання існуючих Грід-ресурсів. Тому, в середовищі створеної Grid-інфраструктури почалася розробка Семантичного Web-порталу знань, який призначений для побудови національних реєстраторів схем метаданих і онтологій, заснованих на міжнародних стандартах, алгоритмів вилучення знань з даних та їх подальшого ефективного використання в національній Grid-інфраструктурі і мережі Світових Центрів Даних (СЦД).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. M. Z. Zgurovsky. Development of Educational and Research Segment of Information Society in Ukraine. – //Proc. WSIS .-Tunis.- 2004.-P.103-107.
2. M. Z. Zgurovsky. Development of Educational and Research Segment of Information Society in Ukraine. – //Системні дослідження та інформаційні технології.-Київ.-2006.-№1.-С.7-17.
3. Петренко А.І. Національна Grid – інфраструктура для забезпечення наукових досліджень і освіти.- // Системні дослідження і інформаційні технології, – Київ. – №1. – 2008. – С.79-92.
4. Zgurovsky M. Impact of the Information Societ5y on Sustainable Development: Global and Regional Aspects.- Data Science Journal, Volume 6, Supplement, 11 March 2007
5. 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
6. Петренко А.І. Вступ до Grid технологій в науці і освіті (навчальний посібник).-Київ.- Політехніка.-2008.-124
7. Zagorodny A., Zgurovsky M. , Zinovjev G. , Petrenko A. , Martynov E. Integrating Ukraine into European Grid Infrastructure.- // Proc. of the 21-ST codata Conference.-Kiev.-3-8 Octomber 2008
8. Zinovjev G., Martynov E., Svistunov S. Academic segment of Ukrainian Grid infrastructure,- // Proc. of the 21-ST codata Conference.-Kiev.-3-8 Octomber 2008
9. Згуровський М.З., Петренко А.І. Grid технології для е-науки і освіти.-// “Наукові вісті НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ» , №2, 2009.- с.10-17.
10. Petrenko A., Velichkevich S., Alenin O. Educational segment of Ukrainian Grid infrastructure.- // Proc. of the 21-ST codata Conference.-Kiev.-3-8 Octomber 2008