

Реалізація розподілених обчислень на основі грід-платформи з відкритим кодом BOINC

Шийка Ю.Я., Шувар Р.Я.

*Львівський національний університет імені Івана Франка,
факультет електроніки, e-mail: shiyka@mail.lviv.ua*

Розглянуто особливості реалізації розподілених обчислень на основі грід-інфраструктури. Представлено особливості вибраної грід-платформи BOINC. Описано проект розподіленої обробки растрових зображень.

Однією з галузей використання вільного програмного забезпечення є реалізація нових комп'ютерних технологій на основі програмного забезпечення з відкритим кодом. Зокрема останнім часом стрімко розвиваються технології розподілених обчислень (грід). Грід або грід-інфраструктура - це розподілена програмно-апаратна комп'ютерна мережа з принципово новою організацією обчислень і управління потоками завдань і даних.

Така комп'ютерна інфраструктура призначена для об'єднання обчислювальних потужностей окремих робочих станцій чи організацій. На основі грід-технології здійснюється інтегрування регіональних і навіть національних обчислювальних комп'ютерних інфраструктур для створення об'єднаних інтернаціональних ресурсів, для розв'язування науково-технічних задач, що потребують дуже великих обчислювальних ресурсів.

Одна з основних частин програмного забезпечення грід-інфраструктури – так зване “Middleware” (проміжне програмне забезпечення). Його завданням є організація та об'єднання ресурсів у мережі, забезпечення автоматичної взаємодії “машина-машина” (M2M) для об'єднання у єдину обчислювальну грід-мережу. Очевидно, що від вибору проміжного програмного забезпечення значною мірою залежить структура грід-мережі та вимоги до адаптації програм, що будуть виконуватися. В зв'язку з цим постає задача вибору проміжного програмного забезпечення, враховуючи його можливості та поставлені до нього вимоги і задачі.

Серед відомого Middleware програмного забезпечення є як платні так і безкоштовні платформи. Серед них найбільш поширені: NorduGrid, Unicore, MOBIDICK, NeuroGrid, CrossGrid, GridLab, SAGA API, BOINC та ін.

Однак найбільший інтерес викликають платформи з відкритим кодом, оскільки вони забезпечують найбільшу гнучкість і можливість відносно легкого впровадження власних проектів розподілених обчислень.

Проаналізувавши можливості різних грід-платформ, для реалізації розподілених обчислень у локальній мережі факультету електроніки було вибрано платформу з відкритим кодом BOINC (Berkeley Open Infrastructure for Network Computing). Серед особливостей цієї платформи варто відмітити:

1. Автономність проектів – різні BOINC проекти є незалежними, кожен з них керує власними серверами та базами даних. Немає центрального

каталогу або процесу затвердження проектів.

2. Гнучкість для учасників – учасники можуть приймати участь в декількох проектах, вони контролюють, в яких проектах приймають участь і як їх ресурси розподілені між цими проектами. Коли проект не працює або не має завдань, ресурси розподіляються між іншими проектами.
3. Гнучка структура програм – існуючі програми написані різними мовами (C, C + +, Fortran) можуть працювати як додатки VOINC практично без змін. Додаток може складатися з декількох файлів (наприклад, декілька програм і об'єднуючий скрипт). Нові версії програм можуть бути впроваджені без необхідності будь-яких дій учасників проекту.
4. Безпека – VOINC захищений проти атак. Наприклад, використання цифрового підпису при розповсюдженні виконавчих файлів забезпечує захист від розповсюдження зловмисниками вірусів під виглядом задач VOINC.
5. Продуктивність сервера і масштабованість – один сервер середнього класу може відправляти задачі і обробляти результати мільйонів робочих місць на день. Серверна архітектура також масштабована, що дозволяє легко збільшити потужність сервера за рахунок додавання нових машин.
6. Доступність вихідного коду – платформа VOINC розповсюджується вільно з відкритим кодом на умовах Lesser General Public License. Проте, VOINC додатки не обов'язково повинні бути з відкритим вихідним кодом.
7. Незалежність від платформи учасника – клієнт VOINC доступний для більшості поширених платформ (Mac OS X, Windows, Linux та інших Unix-подібних систем). Клієнт може використовувати кілька процесорів, в тому числі і GPU.
8. Спільнота учасників – VOINC надає веб-інструменти, такі як дошки оголошень, профілі користувачів, особисті повідомлення, тощо. Це дозволяє об'єднувати учасників проектів у формі інтернет-спільнот.

На основі платформи VOINC у міжкафедральній лабораторії комп'ютерних інформаційних технологій факультету електроніки реалізовано проект віртуального суперкомп'ютерного центру розподілених обчислень для виконання складних обчислювальних науково-практичних і навчальних задач. Здійснено встановлення та налаштування серверної частини грид-платформи VOINC під управлінням Debian Linux на сервері міжкафедральної лабораторії комп'ютерних інформаційних технологій факультету електроніки. Участь в розподілених обчисленнях приймають комп'ютери локальної мережі факультету, надаючи свої незадіяні ресурси.

В рамках цього проекту реалізовано задачу розподіленої обробки растрових зображень великого розміру, а саме їх сегментацію за кольорами. Для цього використано програму сегментації растрових зображень, яку було адаптовано для розподіленої обробки, за принципом поділу великого завдання (обробка великого зображення, чи багатьох зображень) на окремі незалежні підзадачі, які порівняно легко виконуються на окремих комп'ютерах учасників розподілених обчислень.

Таким чином реалізовано серверну та клієнтську частини розподіленої сегментації растрових зображень. Завданням серверної частини є формування підзадач, перевірка результатів виконання підзадач та формування остаточного результату. А завданням клієнтської частини є виконання переданих їй окремих підзадач і передача результатів серверній частині.

Отримана система розподілених обчислень використана для обробки даних дистанційного зондування Землі.