

Петренко А.І.

ННК "ІПСА" НТУУ "КПІ"

GRID технології в обробці даних

Основні ресурсні елементи Grid-мереж – це суперкомп'ютери та суперкомп'ютерні центри, а найважливіша інфраструктурна складова – високошвидкісні мережі передачі даних. Суперкомп'ютери, не об'єднані в територіально-розподілену систему, мають як мінімум три істотних недоліки: вони швидко морально старіють (суперкомп'ютери з першої сотні рейтингу Top-500 уже через два-три роки, як правило, опиняються в самому хвості цього списку чи взагалі випадають із нього); вони майже не піддаються серйозній модернізації, що найчастіше не дозволяє використовувати їх для вирішення задач нового рівня складності; ККД їх використання незначний із-за нерівномірності завантаження процесорів. Цих недоліків можна позбутися, об'єднавши суперкомп'ютери у Grid-мережу. Однак для ефективної експлуатації таких Grid-мереж необхідно обладнати суперкомп'ютери ліцензійним прикладним програмним забезпеченням (ППЗ), яке базується на паралельних методах обчислень чи обробки інформації і не обмежено кількістю робочих місць для користувачів.

Нажаль, такого ППЗ не так ще багато в світі, взагалі, і в Україні, зокрема, і його коштовність надто висока. Тому не має сенсу дублювання ППЗ в різних сегментах (академічному і освітянському) єдиної національної Grid-мережі, що контролюються відповідно установами Національної Академії наук і Міністерства освіти і науки України. Можна рекомендувати для погодженого оснащення Grid-мережі України вибирати необхідне ППЗ з існуючого зараз в світі за такими напрямками:

1. *Інженерія* – ANSYS і ABAQUS (програми, реалізуючі метод кінцевого елемента для вирішення задач з моделями у вигляді диференціальних рівнянь з частинними похідними); LSDYNA (тривимірний аналізатор деформації нееластичних твердих тіл і структур); SAS (інструментарій для маніпулювання, аналізу і узагальнення даних);
2. *Динаміка рідин* – Fidap і Fluent (програми моделювання перебігу нестискованої рідини); Flow 3D (тривимірний аналізатор динаміки рідин); Gambit (пакет програм для побудови і аналізу моделей обчислювальної динаміки рідин); TGrid (програма генерації тривимірної сітки фактично безмежного розміру і складності для вирішення задач динаміки рідин);
3. *Хімія* – AMBER (програма молекулярного моделювання); CNS (аналізатор структури макромолекул для X-гау кристалографії і NMR спектроскопії); CSD (база даних про сотні тисяч структур молекул кристалів); GAMESS (програма аналізу структур кристалів); GAUSSIAN (комплекс програм визначення орбітальної густини молекул); GROMACS (аналізатор динаміки молекул); Jmol (програма візуалізації молекул); SnB (будівник графічної структури кристалів для X-гау дифракційних даних);
4. *Біоінформатика* – ClustalW (програма багаторазового вирівнювання послідовності ДНК або протеїнів); EMBOSS (пакет програм аналізу послідовностей в молекулярній біології); HMMER (програма аналізу послідовностей протеїну); Mascot Search Engine (пошукова програма для ідентифікації протеїнів за первинними даними мас-спектрометрії);
5. *Обчислення* – MPI Library (бібліотека для виконання паралельних операцій, використовуючи розподілену пам'ять); bcMPI (пакет програм, який виконує продовження MPI для MATLAB); FFTW (бібліотека підпрограм для обчислення дискретного перетворення Фур'є); Intel MK (бібліотека Intel, що містить LAPACK, BLAS і деякі програми обчислень); MATLAB (обчислювальне середовище для числових розрахунків і візуалізації); MINPACK (бібліотека програм Fortran для вирішення задач нелінійної багатовимірної мінімізації); NAG Libraries (набір підпрограм для вирішення задач чисельного аналізу і статистики); Octave (мова високого рівня для числових

обчислень); *ScaLAPACK* (бібліотеки для паралельних обчислень в лінійній алгебрі); *Mathematica* (базовий інструментарій для наукових досліджень і інженерного аналізу і моделювання, для технічної освіти у вузі);

6. *Візуалізація – AVS* (універсальний інструментарій візуалізації даних); *Gnuplot* (програма інтерактивного відображення даних і побудови графіків з управляючим рядком); *VolSuite* (багатофункціональний відкритий інструментарій для побудови додатків).

Слід відмітити, що оснащення української Grid-мережі засобами ППЗ вже розпочалося: Інститут теоретичної фізики (головна організація академічного сектору Grid-мережі) придбала програми *GAUSSIAN* і *Mathematica*, а НТУУ "КПІ" (головна організація освітнянського сектору Grid-мережі) замовила програму *ANSYS*. Але здається, що подальше збагачення арсеналу ППЗ – це не тільки справа головних організацій, а скоріше Віртуальних організацій (ВО), що об'єднують за професійними інтересами організації і фізичні особи, які домовляються спільно використовувати свої ресурси через Grid-мережу. Умови для створення ВО в Україні з'явилися з початку 2008 року завдяки відкриттю Центра сертифікації, або Центра відкритих ключів.

Бібліографія

1. Zgurovsky M., Petrenko A. National Ukrainian Grid infrastructure (UGRID) for science and education // Proc. Intern. Conference on Computer Science and Information Technologies, 2006, October, Lviv, Ukraine, pp.6–12.
2. Петренко А.І. Національна Grid – інфраструктура для забезпечення наукових досліджень і освіти // Системний аналіз і інформаційні технології. – Київ, ПСА НТУУ "КПІ", №1, 2008.
3. Академічний сегмент української Grid-мережі: www.acadgrid.org.ua
4. Освітнянський сегмент української Grid-мережі: www.grid.ntu-kpi.kiev.ua