

Использование VOINC-грид в вычислительноемких научных исследованиях

Наталия Никитина

Институт прикладных математических исследований
Карельского научного центра РАН, г. Петрозаводск

2012

Грид-вычисления

Грид-вычисления — это форма организации распределенных вычислений, в которой территориально распределенные вычислительные кластеры объединяются высокоскоростными каналами связи для совместного решения задач, требующих значительных вычислительных ресурсов.

Грид-вычисления

Грид-вычисления — это форма организации распределенных вычислений, в которой территориально распределенные вычислительные кластеры объединяются высокоскоростными каналами связи для совместного решения задач, требующих значительных вычислительных ресурсов.

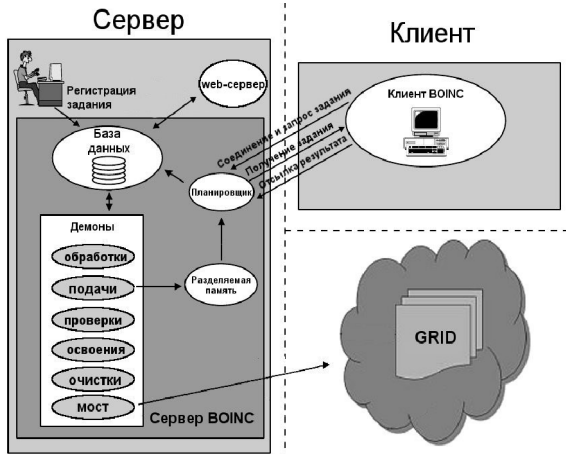
Desktop Grid

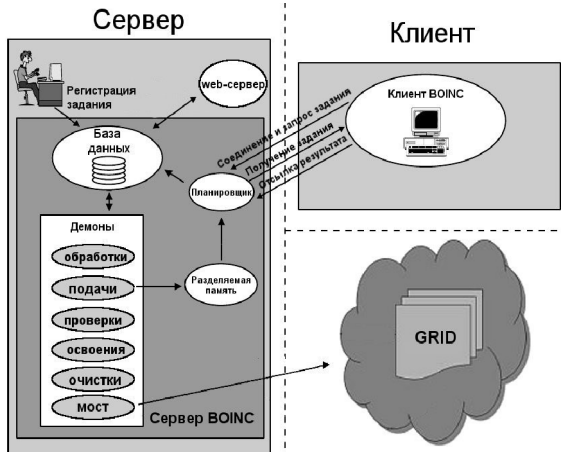
Desktop Grid — грид-система, использующая в качестве вычислительных узлов неспециализированные вычислители (как правило, персональные компьютеры), объединенные сетью Интернет и/или локальной сетью организации.

BOINC (Berkeley Open Infrastructure for Network Computing) — система промежуточного программного обеспечения с открытым исходным кодом (LGPL) для организации Desktop Grid в сети Интернет или локальной сети.

BOINC (Berkeley Open Infrastructure for Network Computing) — система промежуточного программного обеспечения с открытым исходным кодом (LGPL) для организации Desktop Grid в сети Интернет или локальной сети.

Платформа BOINC имеет архитектуру «клиент-сервер». Клиентская часть может работать на произвольном количестве компьютеров с различными аппаратными и программными характеристиками. BOINC предоставляет возможность гибкой настройки клиентской части, регулируя максимальный размер загружаемых файлов, время выполнения рабочих заданий, загрузку CPU или GPU, выделяемый объем оперативной памяти и дискового пространства.





Для переноса на VOINC неадаптированных приложений используются приложения-«обертки», которые берут на себя взаимодействие с ядром программы-клиента, запуская исходное приложение как свой дочерний процесс.

BOINC-грид в научных исследованиях

BOINC-грид эффективны для ресурсоемких приложений, которые можно разделить на большое количество независимых подзадач:

BOINC-грид в научных исследованиях

BOINC-грид эффективны для ресурсоемких приложений, которые можно разделить на большое количество независимых подзадач:

- Проведение большого количества независимых вычислительных экспериментов

BOINC-грид в научных исследованиях

BOINC-грид эффективны для ресурсоемких приложений, которые можно разделить на большое количество независимых подзадач:

- Проведение большого количества независимых вычислительных экспериментов
- Моделирование случайных событий методом Монте-Карло

BOINC-грид в научных исследованиях

BOINC-грид эффективны для ресурсоемких приложений, которые можно разделить на большое количество независимых подзадач:

- Проведение большого количества независимых вычислительных экспериментов
- Моделирование случайных событий методом Монте-Карло
- Задачи, решаемые перебором множества вариантов

Грид-сегмент КарНЦ РАН

Грид-сегмент КарНЦ РАН

- Кластер ЦКП КарНЦ РАН (10 вычислительных узлов, один управляющий узел, пиковая производительность 850 GFlops, объем доступной дисковой памяти 2,33 Тб)

Грид-сегмент КарНЦ РАН

- Кластер ЦКП КарНЦ РАН (10 вычислительных узлов, один управляющий узел, пиковая производительность 850 GFlops, объем доступной дисковой памяти 2,33 Тб)
- Серверы институтов

Грид-сегмент КарНЦ РАН

- Кластер ЦКП КарНЦ РАН (10 вычислительных узлов, один управляющий узел, пиковая производительность 850 GFlops, объем доступной дисковой памяти 2,33 Тб)
- Серверы институтов
- Персональные компьютеры сотрудников

Грид-сегмент КарНЦ РАН

- Кластер ЦКП КарНЦ РАН (10 вычислительных узлов, один управляющий узел, пиковая производительность 850 GFlops, объем доступной дисковой памяти 2,33 Тб)
- Серверы институтов
- Персональные компьютеры сотрудников
- Система промежуточного ПО — BOINC

Грид-сегмент КарНЦ РАН

- Кластер ЦКП КарНЦ РАН (10 вычислительных узлов, один управляющий узел, пиковая производительность 850 GFlops, объем доступной дисковой памяти 2,33 Тб)
- Серверы институтов
- Персональные компьютеры сотрудников
- Система промежуточного ПО — BOINC
- Пиковая производительность грид-сегмента — 1,04 TFlops

Пример 1

Серия расчетов энергий многоатомных систем квантовохимическими методами: вычислительные эксперименты в рамках проекта по прогнозированию состава устойчивых комплексных частиц в расплавах галогенидов щелочных металлов на основе квантовохимических расчетов модельных систем.

Пример 1

Серия расчетов энергий многоатомных систем квантовохимическими методами: вычислительные эксперименты в рамках проекта по прогнозированию состава устойчивых комплексных частиц в расплавах галогенидов щелочных металлов на основе квантовохимических расчетов модельных систем.

- Расчеты с использованием программного пакета Firefly
- 47 независимых вычислительных подзадач
- Общее время выполнения — 66,5 часов
- Программа-«обертка» GenWrapper

Пример 2

Обратная задача определения по экспериментальным данным параметров математической модели выделения водорода из порошка гидрида алюминия: поиск решений путем минимизации среднеквадратичного функционала методом слепого поиска на сетке в пространстве параметров.

Пример 2

Обратная задача определения по экспериментальным данным параметров математической модели выделения водорода из порошка гидрида алюминия: поиск решений путем минимизации среднеквадратичного функционала методом слепого поиска на сетке в пространстве параметров.

- 15000 (250) малосвязанных вычислительноемких подзадач
- Язык реализации приложения — Fortran-90/95
- Общее время выполнения — 6 часов 13 минут (66 минут)
- Программа-«обертка» BOINC Wrapper

Пример 3

Поиск ассоциативных правил в больших наборах исходных данных. Задача является одним из направлений Data Mining — процесса обнаружения в «сырых» данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Пример 3

Поиск ассоциативных правил в больших наборах исходных данных. Задача является одним из направлений Data Mining — процесса обнаружения в «сырых» данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

- Язык реализации приложения — C++
- Комплекс программ: генератор рабочих заданий, адаптированное клиентское приложение, программа освоения результатов
- Ускорение анализа больших наборов данных в 6 раз на 32 узлах BOINC-грид

Спасибо за внимание!