

УДК 004.75

ОРГАНИЗАЦИЯ КВАНТОВО-ХИМИЧЕСКИХ РАСЧЕТОВ С ИСПОЛЬЗОВАНИЕМ ПАКЕТА FIREFLY В ГЕТЕРОГЕННОЙ ГРИД-СИСТЕМЕ НА БАЗЕ VOINC

Е. Е. Ивашко¹, Н. Н. Никитина¹

Рассмотрен опыт проведения большого числа вычислительных экспериментов на базе гетерогенной грид-системы под управлением программной платформы VOINC. Для проведения квантово-химических расчетов использован программный пакет Firefly, адаптированный для выполнения в распределенной среде с помощью программы-«обертки» GenWrapper. Постановка заданий производилась через специальный web-интерфейс доступа к грид-сегменту.

Ключевые слова: VOINC, Firefly, распределенные вычисления, грид.

1. Введение. Успешное выполнение современных научных и прикладных исследований требует, как правило, проведения большого числа ресурсоемких вычислительных экспериментов. Среди отраслей наук суперкомпьютерные вычисления являются особенно востребованными в квантовой химии, в особенности в следующих ее направлениях:

- изучение строения и структуры вещества;
- создание веществ и материалов с заранее заданными свойствами;
- механизмы протекания сложных химических реакций и др.

Одной из наиболее популярных форм организации вычислительных ресурсов для проведения большого числа расчетов является грид. Согласно определению, грид-вычисления — это форма организации распределенных вычислений, в которой участвуют территориально распределенные разнородные компьютеры, объединенные с помощью сети передачи данных и работающие совместно для решения задач, требующих значительных вычислительных ресурсов [1]. С точки зрения сетевой организации грид представляет собой согласованную, открытую и стандартизованную среду, которая обеспечивает гибкое, безопасное, скоординированное разделение вычислительных ресурсов и ресурсов хранения данных [2]. Наиболее эффективно использование грид при выполнении следующих задач:

- анализ и обработка независимых наборов данных;
- решение задач, обладающих хорошей степенью параллелизации по данным.

Существует большое количество систем промежуточного программного обеспечения (СППО), предназначенных для организации, сопровождения и управления грид. По назначению СППО можно условно разделить на две группы. Первая группа содержит системы, предназначенные для объединения относительно небольшого числа высокопроизводительных вычислителей (кластеров). К таким системам относятся, например, Condor [3], Globus Toolkit [4], Unicore [5] и др. Вторая группа СППО содержит системы, цель которых заключается в объединении в грид большого числа (до сотен тысяч) вычислителей, каждый из которых обладает относительно невысокой производительностью. Такие грид-сети также называют системами распределенных вычислений (distributed computing) и системами добровольных вычислений (volunteer computing); их специфика заключается, в частности, в высокой вероятности недоступности отдельных вычислительных узлов. Наиболее популярной СППО этой группы является платформа VOINC (Berkeley Open Infrastructure for Network Computing) [6]; другой пример — разработка НИВЦ МГУ им. М. В. Ломоносова система X-Com [7].

2. Грид-сегмент КарНЦ РАН. В августе 2010 г. на базе Центра коллективного пользования Карельского научного центра РАН «Центр высокопроизводительной обработки данных» (ЦКП КарНЦ РАН) [8] был создан и запущен в тестовую эксплуатацию гетерогенный грид-сегмент, объединяющий вычислительные ресурсы кластера ЦКП КарНЦ РАН, серверов институтов и рабочих персональных компьютеров сотрудников КарНЦ РАН.

В качестве системы промежуточного программного обеспечения грид-сегмента КарНЦ РАН используется разработанная в университете Беркли платформа организации распределенных вычислений VOINC — активно развивающаяся СППО с открытым исходным кодом, ставшая основой большого числа независимых проектов добровольных вычислений (наиболее популярные из них — Climateprediction.net [9],

¹ Институт прикладных математических исследований Карельского научного центра РАН, ул. Пушкинская, 11, 185910, г. Петрозаводск; Е. Е. Ивашко, науч. сотр., e-mail: ivashko@krc.karelia.ru; Н. Н. Никитина, стажер-исследователь, e-mail: nikitina@krc.karelia.ru

SETI@home [10] и Einstein@home [11]). Платформа BOINC отличается простотой в установке, настройке и администрировании, а также обладает хорошими возможностями по масштабируемости, простоте подключения новых вычислительных узлов, использованию дополнительного ПО, интеграции с другими грид-системами и т.д.

Платформа BOINC имеет архитектуру “клиент–сервер”, при этом клиентская часть может работать на компьютерах с различными аппаратными и программными характеристиками. Ключевым объектом системы является проект — автономная сущность, в рамках которой производятся распределенные вычисления. BOINC-сервер поддерживает одновременную работу большого числа независимых проектов; каждый вычислительный узел может одновременно производить вычисления для нескольких BOINC-проектов. Проект однозначно идентифицируется своим URL-адресом. 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.

Рабочий процесс в грид-системе, основанной на платформе BOINC, организован следующим образом (рис. 1). Вычислительные узлы, имеющие свободные ресурсы, обращаются к серверу для получения новых рабочих заданий. Сервер BOINC рассылает клиентским приложениям экземпляры рабочих заданий, клиенты выполняют вычисления и отсылают обратно результаты. После получения результатов сервер проверяет и обрабатывает их, например заносит в базу данных или автоматически создавая на их основе новые рабочие задания.

Платформа BOINC позволяет выполнять в грид-системе специализированные приложения, в которых, в частности, может быть реализован механизм сохранения промежуточных результатов вычислений, графическое отображение процесса выполнения приложения, автоматическое распараллеливание вычислительного процесса в зависимости от количества доступных на клиенте вычислительных ядер и т.п. Однако унаследованные (неадаптированные для работы в грид) приложения с помощью специальных приложений-“обертки” также могут быть перенесены на платформу BOINC без необходимости изменения их исходного кода и перекомпиляции [12]. Приложения-“обертки” берут на себя взаимодействие с ядром программы-клиента, запуская исходное приложение как свой дочерний процесс.

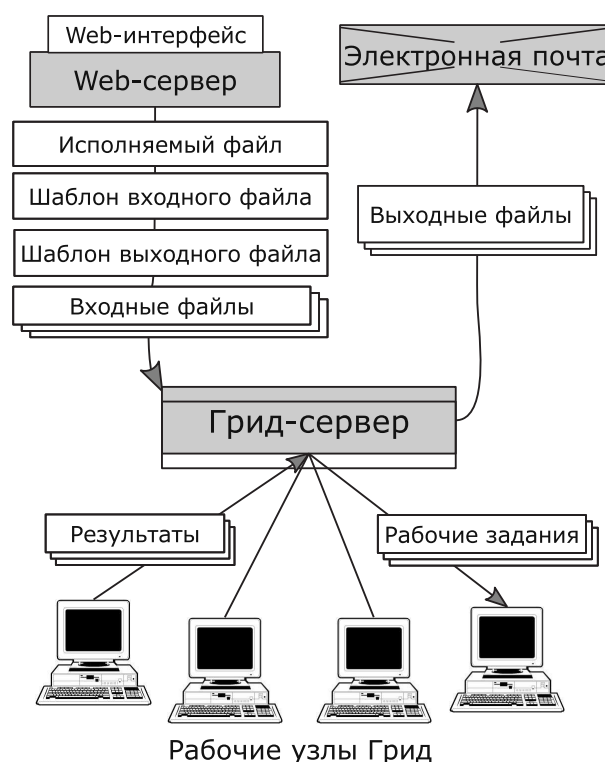


Рис. 1. Архитектура системы

3. Web-интерфейс доступа к грид-сегменту. Для обеспечения удобного доступа пользователей ЦКП КарНЦ РАН к вычислительным ресурсам грид был разработан специальный web-интерфейс. На рис. 2 представлен вид web-интерфейса доступа к грид-сегменту. HTML-форма содержит:

- поля для загрузки на web-сервер исполняемого файла приложения, дополнительных файлов приложения (при необходимости) и zip-архива входных файлов;
- поля для ввода имен входных и выходных файлов, которые требует исполняемый файл приложения;
- поля для ввода оценки ресурсов, требуемых для выполнения заданий;
- поле для ввода адреса электронной почты, на который требуется отправить результат.

Обязательными для заполнения являются поле для загрузки исполняемого файла приложения и поле для ввода адреса электронной почты.

Пользователь загружает исполняемый и дополнительные файлы приложения, указывает имена входных и выходных файлов и загружает zip-архив входных файлов, каждый из которых послужит основой для создания отдельного рабочего задания. Данные, введенные пользователем, обрабатываются на стороне web-сервера и передаются BOINC-серверу. Программа обработки введенных данных на стороне web-сервера создает новое BOINC-приложение и шаблоны входных и выходных файлов. В данном случае

каждое приложение включает в себя программу-“обертку” в дополнение к исполняемому файлу приложения, загруженному пользователем. Рабочие задания VOINC генерируются на основе загруженных входных файлов и созданных шаблонов.

Рабочие узлы грид периодически отсылают на грид-сервер запросы новых заданий. Программа-планировщик, входящая в состав VOINC, отвечает на запросы, распределяя незасланные рабочие задания между узлами, которые запросили задания и которые соответствуют определенным критериям, например имеют достаточный объем оперативной памяти и свободного дискового пространства для выполнения заданий.

Ход выполнения рабочих заданий отображается на web-сайте. Пользователю доступна информация об общем количестве рабочих заданий на сервере и их статусе (например, “в процессе выполнения” или “завершено”). По мере того как завершенные результаты поступают с рабочих узлов грид на сервер, программа-валидатор, входящая в состав VOINC, проверяет выходные файлы и отмечает результат как правильный или ошибочный. Правильные результаты обрабатываются программой-ассимилятором, которая упаковывает их в zip-архив и отсылает пользователю по электронной почте после того, как все результаты рабочих заданий успешно обработаны.

4. Квантово-химические расчеты с использованием системы Firefly. Программный пакет Firefly [13], частично основанный на исходном коде системы GAMESS (US) [14], является одним из наиболее популярных программных средств проведения квантово-химических расчетов. Firefly предназначен для выполнения расчетов *ab initio* и DFT на Intel-совместимых процессорах x86, AMD64 и EM64T. Вычислительные методы *ab initio*, или неэмпирические, предназначены для расчета с максимально возможной точностью физических и химических свойств заданного химического соединения (как многоатомной системы) на основе представлений и методов квантовой механики.

В феврале 2011 г. на базе грид-сегмента КарНЦ РАН была проведена серия расчетов энергий многоатомных систем с использованием пакета Firefly. Вычислительные эксперименты проводились магистранткой Петрозаводского государственного университета Кременецкой О.В. в рамках проекта по прогнозированию состава устойчивых комплексных частиц в расплавах галогенидов щелочных металлов на основе квантово-химических расчетов модельных систем. Проект ведется ПетрГУ совместно с лабораторией высокотемпературной электрохимии Института химии Кольского научного центра РАН с использованием вычислительных ресурсов ЦКП КарНЦ РАН “Центр высокопроизводительной обработки данных” и при финансовой поддержке РФФИ (грант № 08-03-00397-а). Основные результаты, полученные в рамках проекта, опубликованы в работах [15, 16]. Объектом исследования в проекте являются расплавы фторидов и хлоридов щелочных металлов Na, K, Cs, содержащих небольшие (~ 1 моль) добавки фторидных и хлоридных комплексов переходных металлов. Интерес к данным объектам обусловлен тем, что они широко используются в промышленности для проведения таких технологических процессов, как:

- получение чистых и высокочистых металлов;
- получение защитных и каталитически активных покрытий;
- синтез соединений, которые невозможно получить из водных и неводных сред при низких температурах.

Знание состава комплексов в расплаве позволяет гораздо точнее моделировать химические реакции и перенос заряда в расплаве. При этом знание механизма реакции и влияния на него разных факторов — состава, температуры, поверхности и др. — дает возможность управления реакцией.

Расчеты проводились квантово-химическими методами HF, MP2, DFT с помощью программного па-

Рис. 2. Web-интерфейс доступа к грид-сегменту

кета Firefly. Производилась оптимизация геометрической структуры систем вида $3\text{MCr(III)F}_6 + 18\text{MCl}$ и $3\text{MCr(II)F}_6 + 18\text{MCl}$ ($\text{M} = \text{Na}, \text{K}, \text{Cs}$) и расчет энергии неравновесной структуры $3\text{MCr(III)F}_6 + 18\text{MCl}$ в геометрии $3\text{MCr(II)F}_6 + 18\text{MCl}$. Для каждой из систем были рассчитаны энергии “псевдочастиц” вида $*(n\text{MCrF}_6)$, $n = 1, \dots, 7$, где верхний индекс звездочка означает, что геометрическая структура и базис волновых функций для таких частиц соответствуют всей системе.

Таким образом, требовалось рассчитать по 7 энергий частиц для трех систем и трех видов катионов, т.е. 63 энергии, не включая еще некоторые дополнительные менее затратные по времени расчеты. При квантово-химических расчетах время счета пропорционально четвертой степени числа волновых функций (атомных орбиталей), составляющих молекулярные орбитали. В данном случае расчет энергий частиц $n\text{MCrF}_6$ производился в базисе волновых функций всей системы ($3\text{MCr(III)F}_6 + 18\text{MCl}$, например); следовательно, число атомных орбиталей очень велико. Кроме того, все исследуемые параметры составляют сравнительно небольшую разность двух больших по величине значений энергии, что предъявляет дополнительные требования к точности вычислительного эксперимента. Для этого необходимо было использовать широкие базисные наборы, т.е. раскладывать каждую атомную орбиталь на большое количество математических функций (например, Гаусса) и при расчете подбирать большое количество коэффициентов разложения, что также увеличивает время счета.

Указанные обстоятельства приводят к необходимости задействования больших вычислительных ресурсов. Однако требование организации большого числа однотипных расчетов при вычислении энергий частиц дало возможность использовать грид-сегмент КарНЦ РАН для проведения вычислительных экспериментов в рамках данной задачи.

Грид-сегмент использовался для проведения расчетов энергий частиц, состоящих из комплекса CrF_6 и от одного до семи катионов Na^+ или Cs^+ , содержащих трехвалентный хром, но имеющих расположение атомов, соответствующее системе с двухвалентным хромом ($3\text{MCr(II)F}_6 + 18\text{MCl}$, $\text{M} = \text{Na}, \text{Cs}$). Значения энергии частиц использовались для оценки энергии активации переноса электрона и сравнения с экспериментальными данными. Кроме того, рассчитывались энергии комплексов Cr(III)F_6 и частиц, состоящих только из катионов Na^+ или Cs^+ (от одного до семи) в геометрии систем $3\text{MCr(II)F}_6 + 18\text{MCl}$, $\text{M} = \text{Na}, \text{Cs}$. Такие расчеты позволяют оценить вклад энергии, требующейся для перестройки электронных оболочек атомов, изменения геометрической структуры комплекса и сближения катионов между собой, в общую энергию образования частицы, состоящей из комплекса и катионов.

Исходные данные для проведения вычислительных экспериментов представляют собой наборы параметров, регулирующих работу программы, в том числе координаты атомов, метод расчета и базис волновых функций. Выходные файлы должны содержать информацию о полной энергии рассчитываемой системы, ее энергетических уровнях, параметрах молекулярных орбиталей и т.п.

5. Реализация вычислительных экспериментов в грид-сегменте КарНЦ РАН. На момент проведения вычислительных экспериментов в состав грид-сегмента КарНЦ РАН входили 14 вычислительных узлов с различными аппаратными и программными характеристиками, а также разными настройками, связанными с организацией вычислений. В частности, два узла обслуживали проекты BOINC в монопольном режиме, а на десяти вычислительных узлах кластера расчеты в рамках грид регулярно приостанавливались для выполнения расчетов пользовательских задач, запускаемых на кластере с помощью системы управления заданиями. Суммарная пиковая производительность грид составила 1.04 TFLOPS.

Наборы исходных данных, исполняемый файл пакета Firefly, а также необходимые для его работы вспомогательные библиотеки были загружены через web-интерфейс доступа к грид-сегменту. Технология организации вычислений в рамках BOINC-проекта с использованием web-интерфейса была представлена выше в разделе 2. По окончании расчетов все выходные файлы были автоматически отосланы на электронную почту пользователя, выполнявшего расчеты.

Все рабочие задания выполнялись в однопроцессорном режиме. С учетом настроек ограничения времени работы BOINC на вычислительных узлах, а также накладных расходов на передачу данных между сервером и клиентами максимальное время выполнения одного задания составило 66.0 часов, среднее — 17.7 часов. Суммарное процессорное время выполнения всего набора рабочих заданий составило 383.7 часов, а время между отправкой первого рабочего задания и получением последнего результата составило 66.5 часов.

Адаптация пакета Firefly для выполнения на платформе BOINC была проведена при помощи программы “обертки” GenWrapper, которая позволяет в дополнение к основному приложению выполнять сценарии командной оболочки, соответствующие стандарту POSIX. Сценарии использовались для установки значений переменных окружения и составления параметров командной строки, необходимых для запуска

Firefly, а также для определения статуса завершившегося рабочего задания по содержимому выходного файла.

6. Заключение. В рамках грид-сегмента на базе ЦКП КарНЦ РАН было организовано проведение квантово-химических расчетов с использованием программного пакета Firefly. Пакет Firefly был адаптирован для выполнения в грид на платформе BOINC при помощи программы-“обертки” GenWrapper. Использование грид позволило снизить нагрузку и повысить эффективность использования вычислительного кластера ЦКП КарНЦ РАН “Центр высокопроизводительной обработки данных”.

СПИСОК ЛИТЕРАТУРЫ

1. Kamra V., Chugh A. TCP/IP security protocol suite for grid computing architecture // Communications in Computer and Information Science. 2011. Part 1. **169**. 30–35.
2. Foster I., Kesselman C. The grid: blueprint for a new computing infrastructure. San Francisco: Morgan Kaufmann Publishers, 2004.
3. Condor. High Throughput Computing (<http://www.cs.wisc.edu/condor>).
4. The Globus Toolkit (<http://www.globus.org/toolkit>).
5. UNICORE — Distributed Computing and Data Resources (<http://www.unicore.eu>).
6. BOINC: Программное обеспечение с открытым исходным кодом для организации добровольных распределенных вычислений и распределенных вычислений в сети (<http://boinc.berkeley.edu/index.php>).
7. Семейство программ X-Com (<http://x-com.parallel.ru/>).
8. Центр высокопроизводительной обработки данных ЦКП КарНЦ РАН. Институт прикладных математических исследований Карельского научного центра РАН (<http://cluster.krc.karelia.ru>).
9. Проект добровольных вычислений Climateprediction.net (<http://climateprediction.net>).
10. Проект добровольных вычислений SETI@home (<http://setiathome.berkeley.edu>).
11. Проект добровольных вычислений Einstein@home (<http://einstein.phys.uwm.edu>).
12. Marosi A.C., Balaton Z., Kacsuk P. GenWrapper: a generic wrapper for running legacy applications on desktop grids // 3rd Workshop on Desktop Grids and Volunteer Computing Systems (PCGrid 2009). Rome, 2009.
13. Granovsky A.A. Firefly version 7.1.G (<http://classic.chem.msu.su/gran/firefly/index.html>).
14. Schmidt M.W., Baldrige K.K., Boatz J.A., Elbert S.T., Gordon M.S., Jensen J.H., Koseki S., Matsunaga N., Nguyen K.A., Su S., Windus T.L., Dupuis M., Montgomery J.A. General atomic and molecular electronic structure system // J. Comput. Chem. 1993. **14**. 1347–1363.
15. Кременецкий В.Г., Кременецкая О.В., Кузнецов С.А., Калинин В.Т. Квантово-химический подход к оценке состава устойчивых комплексных частиц в расплавах галогенидов щелочных металлов // Докл. РАН. 2011. **437**, № 6. 782–784.
16. Stulov Yu.V., Kremenetsky V.G., Kremenetskaya O.V., Fofanov A.D., Kuznetsov S.A. Standard rate constants of charge transfer for the redox couple Cr(III)/Cr(II) in chloride melts: experiment and calculation // Proc. 9th Int. Frumkin Symp. Electrochemical Technologies and Materials for XXI Century. Moscow, 2010.

Поступила в редакцию
19.01.2012
