

Использование VOINC-грид в вычислительном научных исследованиях*

Е.Е. ИВАШКО, Н.Н. НИКИТИНА
*Учреждение Российской академии наук
Институт прикладных математических исследований
Карельского научного центра РАН*
e-mail: ivashko@krc.karelia.ru, nikitina@krc.karelia.ru

Работа посвящена использованию VOINC-грид для проведения вычислительномемких научных исследований. Дано описание платформы VOINC, представлены примеры экспериментов, проведенных на грид-сегменте ЦКП КарНЦ РАН в рамках научно-исследовательских проектов.

1. Актуальность работы

Высокопроизводительные вычисления играют все большую роль при проведении научных исследований, разработке новых видов промышленной продукции и в социальной сфере. Они высоко востребованы как инструмент для решения задач квантовой химии, молекулярной биологии, гидродинамики и прочих отраслей наук, требующих проведения большого количества ресурсоемких вычислительных экспериментов и обработки больших объемов данных. С учетом возникновения новых вычислительномемких задач и необходимостью обработки возрастающих объемов информации, актуальной задачей является разработка и развитие технологий для проведения ресурсоемких вычислений в рамках различных научных и прикладных проектов.

2. Системы Desktop Grid

Для выполнения высокопроизводительных расчетов, а также обработки больших массивов данных, как правило, используют суперкомпьютеры или вычислительные кластеры. Для увеличения производительности вычислительные кластеры объединяют высокоскоростными каналами связи в специализированные грид-системы [1, 2]. Однако с развитием каналов связи (и соответствующим увеличением скорости и пропускной способности) сети Интернет и ростом производительности персональных компьютеров, становится все более популярным направление, связанное с организацией Desktop Grid — грид-сетей, объединяющих неспециализированные вычислители (как правило, персональные компьютеры). Desktop Grid позволяют достаточно быстро и легко объединить значительное число источников сравнительно небольших вычислительных ресурсов для решения вычислительномемких задач. В большинстве случаев такие системы построены на основе добровольно предоставляемых вычислительных ресурсов компьютеров

*Работа выполнена при поддержке Программы стратегического развития Петрозаводского государственного университета на 2012–2016 годы.

частных лиц и организаций (volunteer computing). Однако существуют и примеры построения подобных частных — в масштабах организации или группы организаций — распределенных систем.

В последнее время интерес к Desktop Grid значительно вырос, вычислительные расчеты с использованием этих технологий составляют важную часть отрасли высокопроизводительных вычислений. В рамках Седьмой рамочной программы ЕС (FP7) получил финансирование проект DEGISCO по развитию технологий Desktop Grid, была создана International Desktop Grid Federation, крупнейшие исследовательские проекты на базе Desktop Grid выполняют ведущие мировые университеты (SETI@home, Rosetta@home, ClimatePrediction@home и другие). Суммарный потенциал вычислительных ресурсов Desktop Grid превышает 120 Экзафлопс, что обуславливает повышенный интерес к данным технологиям. Наибольшей популярностью вычисления в рамках Desktop Grid пользуются в сфере научных исследований (астрономия, биология, математика и др.), однако эти технологии применяются и промышленными компаниями. Desktop Grid обладают целым рядом достоинств — высокой пиковой производительностью, низкой стоимостью, простотой развертывания и сопровождения. Системы Desktop Grid хорошо подходят для выполнения научных высокопроизводительных расчетов.

3. Платформа BOINC

Одной из наиболее популярных платформ организации распределенных вычислений в Desktop Grid является BOINC (Berkeley Open Infrastructure for Network Computing) [3]. Из 94 действующих проектов добровольных вычислений 77 основаны на BOINC, т.е. данная платформа практически является стандартом для разработки добровольных вычислений [4]. Ярким примером BOINC-проекта в сети Интернет, использующим свободные вычислительные ресурсы компьютеров добровольцев, является проект SETI@home, цель которого — анализ космических радиосигналов. Средняя производительность вычислений этого проекта по состоянию на 20 мая 2012 года составляет 1,5 петафлопс. Аналогичные показатели пиковой производительности имеет суперкомпьютер, находящийся на 8-м месте в списке TOP-500 (по состоянию на июнь 2012 года). При этом в проекте SETI@home участвует более 1 млн. пользователей, а вычислительная система состоит из 3 млн. узлов.

Платформа BOINC является активно развивающимся открытым (Open Source) программным обеспечением и предоставляет богатую функциональность. При этом она 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 произвольном количестве компьютеров с различными аппаратными и программными характеристиками. Ключевым понятием при организации BOINC-грид является проект — автономная сущность, в рамках которой производятся распределенные вычисления. Сервер поддерживает одновременную работу большого числа независимых проектов; каждый клиент системы может одновременно производить вычисления для нескольких BOINC-проектов. Проект однозначно идентифицируется своим URL-адресом. BOINC предоставляет возможность гибкой настройки клиентской ча-

сти, 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. Вычислительные узлы, имеющие свободные ресурсы, обращаются к серверу для получения новых рабочих заданий в рамках проекта. 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.

4. Использование BOINC-грид в научных исследованиях

В научных исследованиях часто возникают задачи, требующие реализации алгоритмов высокой вычислительной сложности или обработки значительных объемов данных в течение ограниченного периода времени. В частности, эти требования характерны для задач моделирования сложных природных систем, имитационного моделирования, обработки данных наблюдений и др.

Наиболее эффективно использование грид-систем на базе BOINC при выполнении ресурсоемких приложений, которые могут быть разделены на большое количество независимых подзадач, например, для анализа и обработки независимых наборов данных. Характерными примерами подобных приложений являются:

- проведение большого количества независимых вычислительных экспериментов,
- моделирование случайных событий методом Монте-Карло,
- задачи, решаемые перебором множества вариантов.

Платформа BOINC позволяет выполнять в грид-системе специализированные научные приложения, в которых, в частности, может быть реализован механизм сохранения промежуточных результатов вычислений, графическое отображение прогресса выполнения приложения, автоматическое распараллеливание вычислительного процесса в зависимости от количества доступных на клиенте вычислительных ядер и т.п. Однако возможен перенос на платформу BOINC и приложений, не адаптированных для работы в грид, без необходимости изменения их исходного кода и перекомпиляции. Для этого используются специальные приложения-«обертки» [5], которые берут на себя взаимодействие с ядром программы-клиента, запуская исходное приложение как свой дочерний процесс.

5. Грид-сегмент КарНЦ РАН

В августе 2010 г. на базе ЦКП КарНЦ РАН «Центр высокопроизводительной обработки данных» [6] был создан и запущен в тестовую эксплуатацию гетерогенный грид-сегмент, объединяющий вычислительные ресурсы кластера ЦКП КарНЦ РАН, серверов институтов и отдельных рабочих персональных компьютеров сотрудников. В качестве системы промежуточного программного обеспечения грид-сегмента используется платформа BOINC. К настоящему времени на базе грид-сегмента успешно реализован ряд

научно-исследовательских проектов пользователей ЦКП КарНЦ РАН. Далее представлены описания некоторых из этих проектов.

5.1. Расчеты энергий многоатомных систем

На базе грид-сегмента КарНЦ РАН была проведена серия расчетов энергий многоатомных систем с использованием пакета Firefly [7]. 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 квантовохимических расчетов модельных систем [8, 9]. Объектом исследования в проекте являются расплавы фторидов и хлоридов щелочных металлов, содержащих небольшие (~ 1 моль%) добавки фторидных и хлоридных комплексов переходных металлов.

Расчеты проводились квантовохимическими методами с помощью программного пакета Firefly. Пакет Firefly [10], частично основанный на исходном коде системы GAMESS (US) [11], является одним из наиболее популярных программных средств проведения квантовохимических 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.

Исходные данные для проведения вычислительных экспериментов представляли собой наборы параметров, регулирующих работу программы, в том числе координаты атомов, метод расчета и базис волновых функций. Выходные файлы содержали информацию о полной энергии рассчитываемой системы, ее энергетических уровнях, параметрах молекулярных орбиталей и т. п. Для проведения вычислительных экспериментов в грид-системе было сгенерировано 47 независимых вычислительных подзадач, общее время выполнения которых составило 66,5 часов.

5.2. Алгоритмы поиска ассоциативных правил в больших наборах исходных данных

Также на базе грид-сегмента КарНЦ РАН были проведены исследования, связанные с реализацией на BOINC-грид алгоритмов поиска ассоциативных правил в больших наборах исходных данных [12]. Поиск ассоциативных правил 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 [13].

В ходе исследований был разработан комплекс программ на языке C++ для выполнения в гетерогенной грид на базе BOINC, реализующий алгоритм Partition, предназначенный для поиска ассоциативных правил. Выполнение алгоритма состоит из трех этапов, два из которых выполняются параллельно на вычислительных узлах грид-сети. На завершающем этапе происходит объединение промежуточных результатов.

Исходные данные представляли собой фрагменты транзакционной базы данных и ряд параметров алгоритма. Выходные данные на завершающем этапе работы алгоритма содержали обнаруженные ассоциативные правила. На основе результатов проведенных экспериментов были сделаны выводы о производительности разработанного ПО на

крупных (до 100 Гб) тестовых базах данных. По результатам экспериментов, ускорение анализа больших наборов данных на 32 узлах грид-системы составило порядка 6 раз.

5.3. Численная идентификация модели дегидрирования

Еще одним примером использования грид-сегмента КарНЦ РАН являются вычислительные эксперименты, проведенные в рамках обратной задачи определения по экспериментальным данным параметров математической модели выделения водорода из порошка гидрида алюминия. Методика и выбранный подход к математическому моделированию формирования и разложения гидридов подробно описаны в работе [14]. В ходе экспериментов решения задачи были найдены путем минимизации среднеквадратичного функционала методом слепого поиска на сетке в пространстве параметров. В результате экспериментов были получены оценки параметров и качественные выводы, показано, что задача имеет большое количество решений, соответствующих ограничениям, налагаемым физическим смыслом.

Вычисления были организованы с учетом специфики гетерогенной грид-системы (15000 и 250 малосвязанных вычислительноемких подзадач в первом и втором экспериментах соответственно). Роль рабочего задания выполняло однопроцессорное приложение, написанное на языке Fortran-90/95, не адаптированное для выполнения в среде BOINC. Общее время вычислений в грид-системе составило 6 часов 13 минут в ходе первого эксперимента и 66 минут в ходе второго эксперимента. По сравнению с расчетами на персональном компьютере общее время снижено в 13 и 24 раза для каждого эксперимента соответственно.

6. Заключение

В работе рассматривается использование BOINC-грид в качестве распределенной среды для проведения вычислительноемких научных исследований. Системы BOINC-грид эффективны для решения отдельных классов задач, часто возникающих в научных исследованиях. Приводится описание платформы BOINC. Представлены примеры экспериментов, проведенных на грид-сегменте ЦКП КарНЦ РАН в рамках выполнения научно-исследовательских проектов. Данные примеры включают в себя использование как специально разработанного для платформы BOINC приложения, так и запуск неадаптированного приложения и внешней бинарной программы с помощью специальной программы-«обертки».

Список литературы

- [1] Foster I., Kesselman C. The Grid: Blueprint for a New Computing Infrastructure // Morgan Kaufmann, San Francisco. 1999.
- [2] Foster I., Kesselman C., Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations // International J. Supercomputer Applications, 15(3). 2001.
- [3] BOINC: Программное обеспечение с открытым исходным кодом для организации добровольных распределённых вычислений и распределённых вычислений в сети. URL: <http://boinc.berkeley.edu/index.php>
- [4] Distributed Computing Platforms. URL: <http://distributedcomputing.info/platforms.html>

- [5] 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), 2009 May, Rome, Italy.
- [6] Центр высокопроизводительной обработки данных ЦКП КарНЦ РАН / Институт прикладных математических исследований Карельского научного центра РАН. URL: <http://cluster.krc.karelia.ru>
- [7] Ивашко Е.Е., Никитина Н.Н. Организация квантовохимических расчетов с использованием пакета Firefly в гетерогенной грид-системе на базе VOINC // Вычислительные методы и программирование, Изд-во МГУ, Москва, 2012. Т. 13. С. 216–220.
- [8] Кременецкий В.Г., Кременецкая О.В., Кузнецов С.А., академик Калининков В.Т. Квантово-химический подход к оценке состава устойчивых комплексных частиц в расплавах галогенидов щелочных металлов // Доклады Академии наук. 2011. Т. 437, №6. С. 782–784.
- [9] Stulov Yu.V., Kremenetsky V.G., Kremenetskaya O.V., Fofanov A.D., Kuznetsov S.A. Standart rate constants of charge transfer for the redox couple Cr(III)/Cr(II) in chloride melts: experiment and calculation // 9th International Frumkin Symposium. Electrochemical technologies and materials for XXI century. Abstracts, Moscow 2010. P. 31.
- [10] Granovsky A.A. Firefly version 7.1.G. URL: <http://classic.chem.msu.su/gran/firefly/index.html>
- [11] 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. Pp. 1347–1363.
- [12] Ивашко Е.Е., Головин А.С. Методы Data Mining для анализа больших массивов данных в гетерогенной грид на базе VOINC. Труды международной суперкомпьютерной конференции «Научный сервис в сети Интернет: поиск новых решений». 2012. С. 196–199.
- [13] Дюк В., Самойленко А. Data Mining: учебный курс (+CD). СПб: Изд. Питер, 2001. 368 с.
- [14] Chernov I.A., Gabis I.E. Mathematical Modelling of Hydride Formation // Mathematical Modeling, Clustering Algorithms and Applications / Caroline L. Wilson. Mathematics Research Developments series. Nova publishers. 2011. Pp. 203–246.

Employment of BOINC-grid in computationally intensive scientific research

The work is devoted to employment of a BOINC-grid to conduct computationally intensive scientific research. The description of the BOINC platform is given, examples of experiments conducted on the grid segment of CCU KarRC RAS within scientific research projects are presented.

Ключевые слова: BOINC, Desktop Grid, грид.