ЦЕНТР КОЛЛЕКТИВНОГО ПОЛЬЗОВАНИЯ «ЦЕНТР ВЫСОКОПРОИЗВОДИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ» КАРНЦ РАН

Е.Е. Ивашко

Введение

В феврале 2009 г. в Учреждении РАН Карельский научный центр РАН (КарНЦ РАН) был запущен вычислительный кластер. Пусконаладочные работы проводились совместно с компанией-поставщиком Т-Платформы.

Для обеспечения доступа пользователей к вычислительным ресурсам и организации бесперебойной работы кластера был создан Центр коллективного пользования «Центр высокопроизводительной обработки данных», базовой организацией которого установлен Институт прикладных математических исследований (ИПМИ) КарНЦ РАН.

В статье рассказывается о ЦКП КарНЦ РАН, используемом вычислительном кластере и задачах, решаемых с его помощью.

Основные характеристики вычислительной системы

Аппаратное обеспечение

Кластер КарНЦ РАН состоит из 1 управляющего и 10 вычислительных узлов следующей конфигурации:

Вычислительные узлы	CPU: 2*Quad-Core Intel Xeon 5430 2,66 GHz, cache 12 Mb RAM: 4Gb FB-DIMM DDR2-667 REG ECC HDD: HDD WD SATA-II 250Gb
Управляющий узел	CPU: 2*Quad-Core Intel Xeon 5430 2,66 GHz, cache 12 Mb RAM: 4Gb FB-DIMM DDR2-667 REG ECC HDD: 6*HDD WD SATA-II 146Gb

Связь между узлами осуществляется посредством высокоскоростного интерфейса Infiniband, что позволяет минимизировать накладные расходы при взаимодействии узлов кластера.

Пиковая производительность кластера составляет 850 GFlops. На тестах Linpack была показана производительность в 637.2 GFlops, т.е. примерно 75% от пиковой.

Для хранения данных, используемых при проведении вычислений, установлена дисковая система хранения данных ReadyStroage SAN 3994, предоставляющая доступ к массиву дисков суммарным объемом 2.33 Тб, объединенных для обеспечения надежности по технологии RAID.

Программное обеспечение

Управляющий и вычислительные узлы кластера работают под управлением ОС SUSE Linux Enterprise Server 10. Для разработки программ используется Intel Cluster Toolkit Compiler Edition for Linux [1], в состав которого входят:

- Intel C++ and Fortran 10.1
- Intel Debugger 10.1
- Intel MKL Library 10.0
- Intel MPI Library 3.1
- Intel Trace Analyzer and Collector 7.1
- Intel Cluster OpenMP for Intel C++ and Fortran

Кроме того, в системе дополнительно установлена библиотека Boost C++ Library 1.43.0 [2]. Для отладки параллельных программ предустановлена программа Total View Debugger [3].

Доступ к кластеру для компиляции и запуска программ осуществляется по протоколу ssh по внутреннему (доступному только из локальной сети КарНЦ РАН) и внешнему IP-адресам. Использование внешнего IP-адреса позволяет получить доступ к кластеру через Интернет, а доступ по внутреннему IP-адресу осуществляется со значительно более высокой скоростью доступа, что важно при копировании больших объемов данных.

Помимо средств разработки, важное значение также имеют средства мониторинга и управления совместным использованием кластера.

Для предоставления пользователям необходимой информации о работе вычислительного кластера, был создан сайт ЦКП «Центр высокопроизводительной обработки данных» КарНЦ РАН [4] (см. рис. 1).

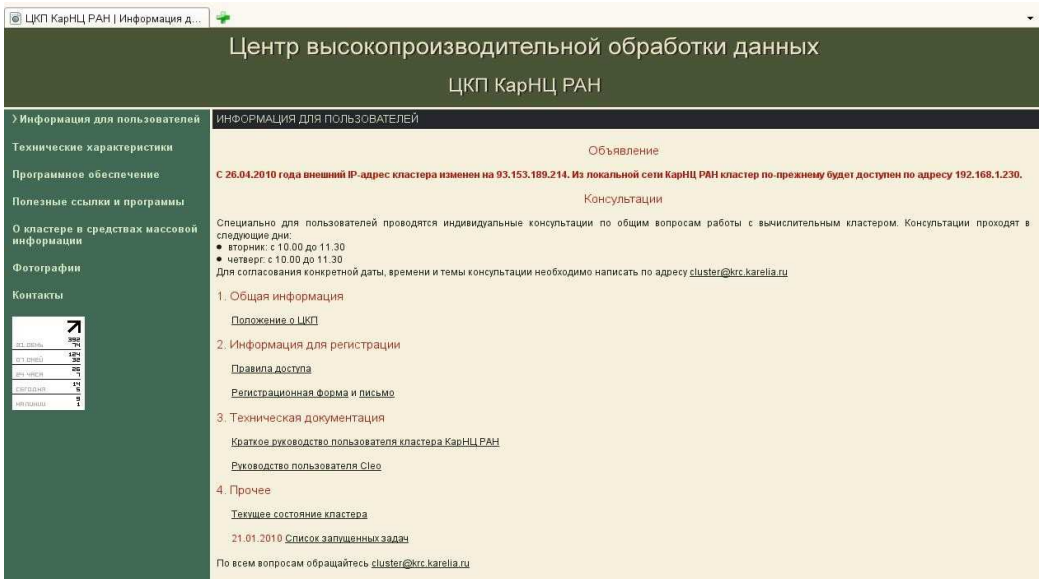

Рис. 1. Сайт ЦКП «Центр высокопроизводительной обработки данных» КарНЦ РАН

На сайте представлена как общая информация по использованию кластера (положение о ЦКП, руководства по работе с кластером, правила регистрации и доступа, технические характеристики и др.), так и сведения о текущей загруженности кластера.

Мониторинг загруженности вычислительных узлов кластера осуществляется с помощью системы Ganglia [5]. Эта система позволяет в реальном времени получать информацию об использовании ресурсов кластера (процессоров, оперативной памяти, сети и др.), а также создавать сводные отчеты за определенные промежутки времени. Пример отчета, генерируемого Ganglia, представлен на рис. 2.

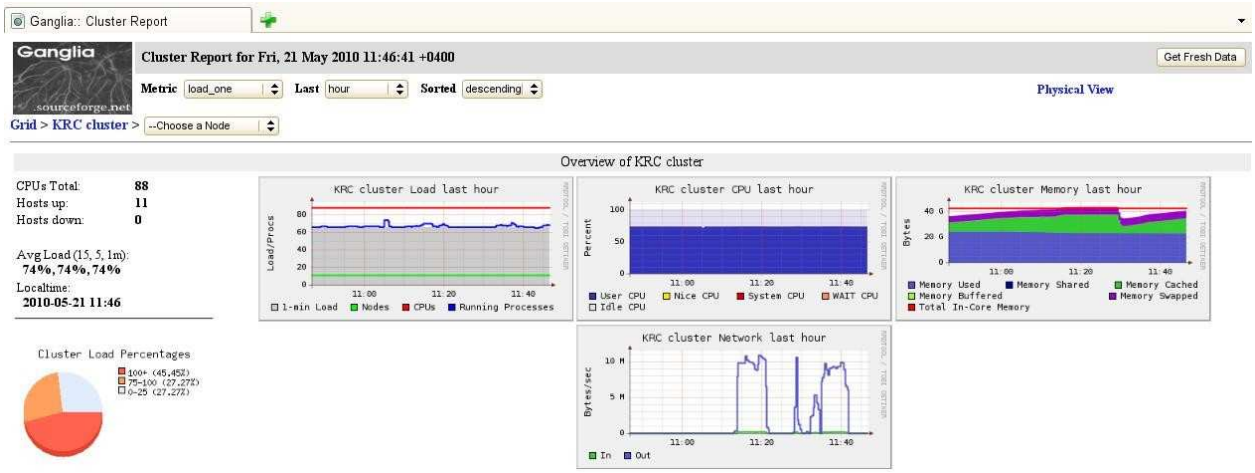

Рис. 2. Отчет Ganglia о загруженности кластера

Для управления заданиями пользователей используется система Cleo [6], которая определяет порядок запуска задач, время их работы, а также распределяет вычислительные ресурсы между задачами и предоставляет информацию о состоянии очередей. Используя сведения, собираемые Cleo, на сайте ЦКП КарНЦ РАН в удобном виде выводится информация о заявках, обрабатываемых кластером в текущий момент. Это помогает пользователям правильно оценить объем свободных ресурсов, а также узнать какие заявки обрабатываются в текущий момент и оценить время до завершения их обработки (см. рис. 3).

Рис. 3. Текущая загруженность кластера

В августе 2009 г. на вычислительный кластер был получен сертификат Intel Cluster Ready.

Научная и образовательная деятельность

С момента запуска вычислительного кластера в штатном режиме, сотрудниками ЦКП были предприняты активные усилия по привлечению сотрудников КарНЦ РАН к использованию вычислительных ресурсов кластера для решения собственных научных задач. На текущий момент список активных пользователей кластера насчитывает 12 человек, из них 9 сотрудников ИПМИ КарНЦ РАН, 1 сотрудник и 2 студента Петрозаводского государственного университета. Далее представлены описания наиболее важных научных работ, проводимых с использованием вычислительных ресурсов кластера.

- Математическое моделирование задач водородного материаловедения (д.ф.-м.н., проф. Ю. В. Заика, к.ф.-м.н. Н. И. Родченкова)

Проект нацелен на разработку математических моделей и эффективных вычислительных методов, которые позволяют моделировать взаимодействие водорода с твердым телом (в том числе и в экстремальных условиях) с учетом новых физико-химических представлений и оценивать различного рода кинетические параметры. Возникает новый класс нелинейных краевых задач, характеризующийся неклассическими динамическими граничными условиями (дифференциальные уравнения не только в объеме, но и на поверхности), свободными границами раздела фаз и условиями сопряжения на стыках слоев (защитные покрытия). Для прямых задач развиваются численные методы, эффективные в классе жестких задач. Разрабатываются устойчивые параллельные алгоритмы решения обратных задач параметрической идентификации моделей и программный комплекс. Алгоритмы ориентированы на экспериментальные методы проницаемости, концентрационных импульсов и термодесорбционной спектрометрии.

На основе разностных аппроксимаций построена неявная разностная схема метода переменных направлений, называемая продольно-поперечной (схемой Писмена-Рэчфорда). Разработан итерационный алгоритм численного моделирования дегазации цилиндрического образца (соответствующего ГОСТа), содержащего растворенный водород, когда нагрев происходит ступенчатым образом (так называемый дискретный ТДС-спектр). Проведено тестирование алгоритма при различных наборах параметров модели.

Рис. 4. Десорбционный поток

Рис. 5. Дискретный спектр

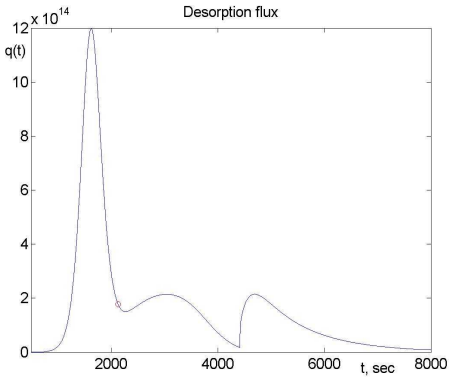

Рис. 4. Десорбционный поток

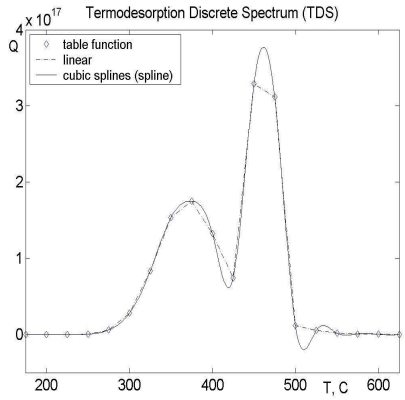

Рис. 5. Дискретный спектр

- Исследование температурных зависимостей в модели водородопроницаемости с нелинейными граничными условиями (объемная десорбция) (д.ф.-м.н., проф. Ю.В.Заика, к.ф.-м.н. Е. П. Борматова)

Исследования проводились в рамках проекта на 2010 г. «Математическое моделирование задач водородного материаловедения» (рук. д.ф.-м.н., проф. Ю.В.Заика) по Программе фундаментальных исследований Отделения математических наук РАН «Современные вычислительные и информационные технологии решения больших задач» (2009–2011 гг.)

Работа заключается в моделировании экспериментов по исследованию водродопроницаемости мембраны (метод установления стационарного потока) в диапазоне температур от 0С до максимально допустимых значений для данных сплавов.

Установлено, что при определенных значениях параметров исследованная модель качественно не соответствует экспериментальным данным для рекристализованного сплава на основе железа. Модель качественно соответствует поведению аморфного сплава при высоких температурах (от 200С), но изменение стационарного потока (и других вычисляемых величин) с ростом температуры незначительно. Так, например, падение потока не превосходит погрешности эксперимента.

- Численное моделирование крупномасштабной гидродинамики Белого моря (к.ф.-м.н. И. А. Чернов)

Модель крупномасштабной циркуляции Арктики, разработанная Яковлевым Н.Г. (ИВМ РАН, г. Москва) адаптирована для Белого моря — водоема, существенно отличающегося по своим гидрологическим характеристикам от Арктического региона. Модель представляет собой систему дифференциальных уравнений в частных производных для полей скоростей течений, температуры и солености в области сложной формы и требует значительных вычислительных ресурсов.

Проект направлен на:

- верификацию модели с целью добиться согласия между вычисляемыми и наблюдаемыми величинами, как в отдельных точках, так и усредненными;
- разработку параллельной версии программной реализации модели.

Численными экспериментами установлен диапазон устойчивости шагов сетки крупномасштабной циркуляции Белого моря. Проведены расчеты с учетом ветра и приливов, а также учтены солнечная радиация, переменная температура воздуха и образование/таяние льда. Работоспособность модели можно считать установленной. Вводится учет стока рек, приливных неравенств, влияния распределения температуры воздуха и атмосферного давления (ветра) по акватории моря и облачности. Выявлен параллелизм расчетной программы. Проводятся работы по переводу последовательного кода программы в параллельный.

На рис.6 и 7 представлены результаты численного моделирования толщины льда и поля скоростей Белого моря, проведенного с использованием вычислительных ресурсов кластера.

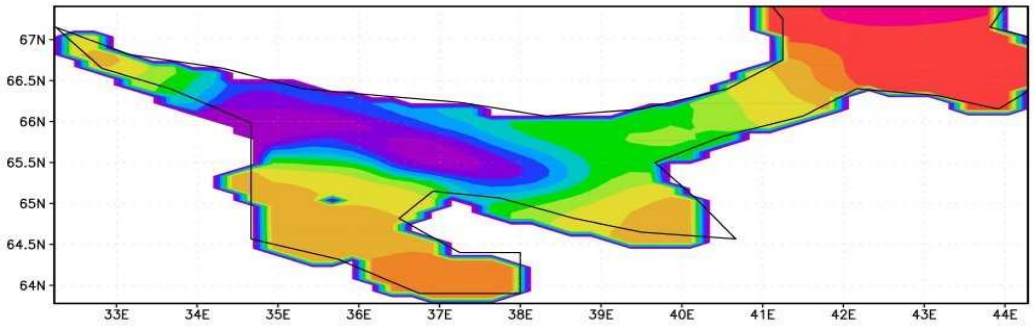

Рис. 6. Толщина льда Белого моря (февраль)

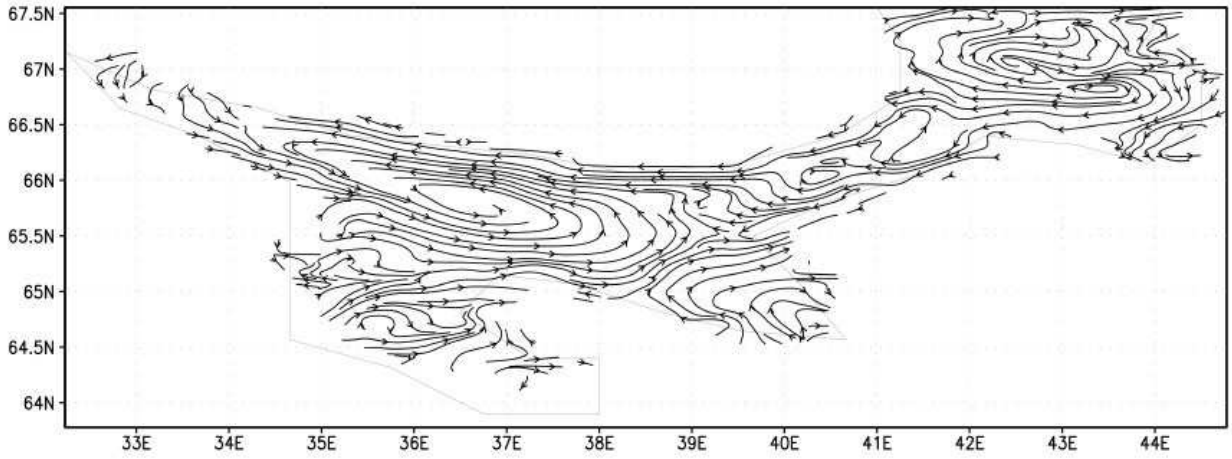

Рис. 7. Поле скоростей Белого моря (прилив, без ветра)

- Разработка алгоритма минимизации энергии Гиббса (д.ф.-м.н., проф. Ю. В. Заика, Д. Г. Вилаев)

Задача определения равновесного термодинамического состояния химических систем является одной из основных задач физической химии; она широко встречается как в научно-исследовательской деятельности, так и на практике при проектировании технологических процессов. В настоящее время наибольший интерес представляют численные методы нахождения равновесного состояния, основным методом решения задачи при этом является нахождение состава реакционной смеси, характеризующегося минимумом энергии Гиббса (МЭГ).

Существующие алгоритмы для нахождения МЭГ не всегда приводят к получению оптимального решения, из-за чего результаты расчетов становятся ненадежными и малоприменимыми на практике.

Характерными особенностями исходной задачи также являются:

- вычислительная неустойчивость целевой функции при появлении компонент с исчезающе малыми мольными долями;
- появление в ходе вычислений матриц, близких к сингулярным.

В рамках научно-исследовательского проекта проводится работа по созданию более надежного алгоритма для нахождения минимума энергии Гиббса. Был разработан базовый вариант алгоритма, позволяющий находить оптимальное решение для ряда случаев, в которых алгоритма Уайта-Джонсона-Данцига (являющийся основой современных программ для решения задачи МЭГ) не может найти решение. Однако указанный вариант алгоритма является более ресурсоемким по сравнению с алгоритмом Уайта-Джонсона-Данцига и пока не решает всех проблем (в частности, не решен вопрос работы алгоритма для задачи с вырожденной матрицей стехиометрических ограничений).

На следующем рисунке представлен пример графика количества вещества в равновесной смеси для задачи, некорректно решаемой алгоритмом Уайта-Джонсона-Данцига и корректно решаемой разработанным алгоритмом.

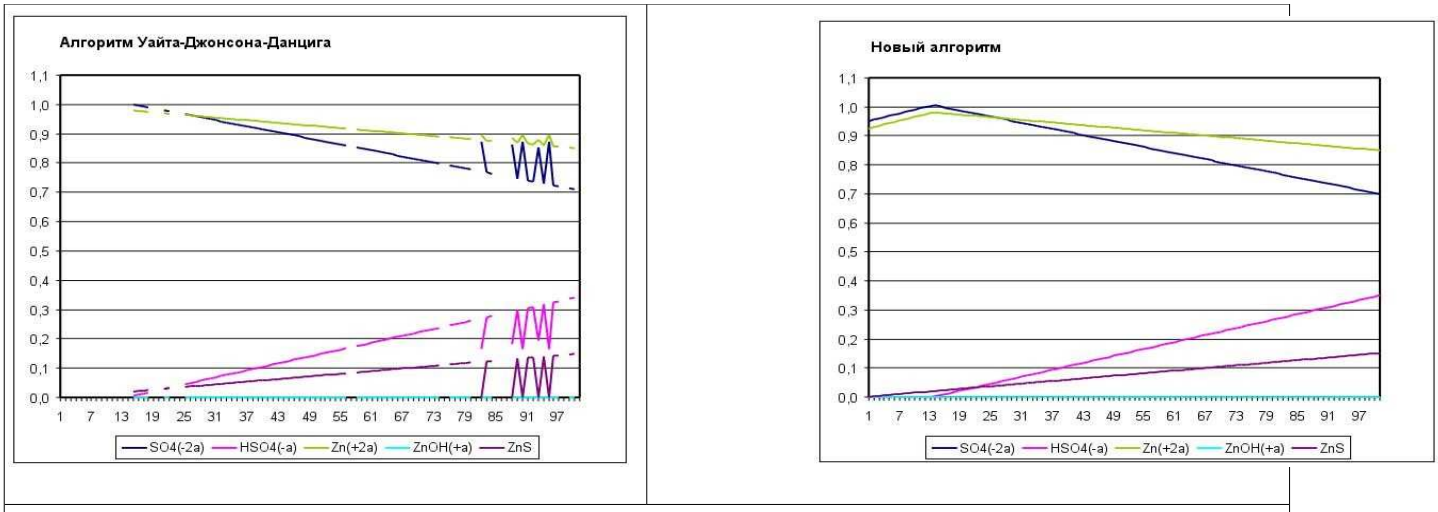

Рис. 8. Нахождение количества вещества в равновесной смеси алгоритмом Уайта-Джонсона-Данцига и разрабатываемым алгоритмом

- Исследование и разработка методов вероятностно-статистического графового моделирования систем передачи и анализа данных (д.ф.-м.н., проф. Ю. Л. Павлов, к.т.н. М. М. Лери, к.ф.-м.н. И. А. Чеплюкова)

Быстрое развитие и широкое использование глобальных сетей передачи данных, одной из которых является сеть Интернет, за два последних десятилетия привело к появлению множества работ как теоретического, так и прикладного характера, направленных на описание структуры и функционирования таких сетей. Предыдущими исследователями было предложено описание глобальных сетей с помощью случайных графов, степени вершин которых представляют собой независимые одинаково распределенные случайные величины, распределение которых является дискретным аналогом распределения Парето. Такие графы иногда называют графами Интернет-типа. Наряду с теоретическими подходами, одним из средств изучения такого рода объектов является имитационное моделирование.

В рамках исследования была создана имитационная модель случайного графа Интернет-типа. С помощью метода Монте-Карло изучалось поведение следующих структурных характеристик Интернет-графов: сумма степеней вершин, максимальная степень вершины, объем гигантской компоненты, объемы следующих по размеру компонент, общее число компонент, число вершин, находящихся на определенном расстоянии от вершины максимальной степени и т.д.

Вычислительные эксперименты проводились при различных значениях параметра распределения степеней вершин на графах различных объемов (от 1 до 100 тысяч вершин). Проводится исследование устойчивости случайных графов Интернет-типа к процессу разрушения в виде направленного удаления вершин максимальной степени. Также изучались возможности использования критерия согласия Пирсона для исследования структуры Интернет-графов.

На рис. 9 представлен один из результатов, полученных при исследовании объема гигантской компоненты Интернет-графа.

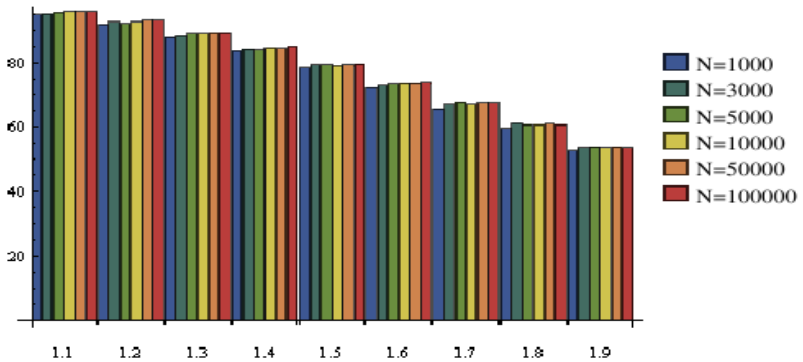

Рис. 9. Процента вершин, вошедших в гигантскую компоненту Интернет-графа, в зависимости от параметра распределения степеней вершин и объема графа N.

- Решение комбинаторных задач (А. М. Караваев)

Следующие задачи направлены на изучения комбинаторных свойств графовых структур.

1. Задача о количестве простых циклов на квадратной решетке.

Имеется прямоугольная решетка размером m × n. Простым циклом называется цикл, проходящий через каждый узел не более одного раза. Ранее эта задача была решена для n = 10. С помощью кластера удалось продвинуться до n = 20, что является значительным вычислительным результатом.

Результаты, связанные с решением данной задачи были представлены автором на международной конференции «Высокопроизводительные параллельные вычисления на кластерных системах» в г. Владимире в ноябре 2009 года [7].

n	**Число простых циклов**
1	0
2	1
3	13
4	213
5	9349
6	1222363
7	487150371
8	603841648931
9	2318527339461265
10	27359264067916806101
11	988808811046283595068099
12	109331355810135629946698361371
13	369549179620398849533878683346444 57
14	381577036885778453044455308512420552673 53
15	120285789340533859558405124213592877516931371715
16	11570932657359853016220237135357395703611284809490441 65
17	339538440233866215589923027136989220328064193307483128228754 35
18	303844278337980735850034087602907615697609016188856597131551124107474 5
19	828994780940754546935155193723999456397444612351458013833196066822882956458645
20	689454183113508056830743257694610073918027584884435945831932331941835810292154647051579

2. Распараллеливание явных формул для подсчета коротких циклов в графе.

Задан ориентированный граф G = (V, E), в котором V множество вершин (|V | = n), а E множество ребер (|E| = m). Требуется подсчитать в этом графе количество циклов длиной k (где 3 ≤ k ≤ n).

Ранее научному коллективу, в котором состоит автор, удалось вывести явные формулы для подсчета коротких циклов в графе (длиной до 12 в произвольном графе и длиной до 14 в двудольном).

В том случае, когда циклы короткие, формулы имеют аддитивную структуру. Задача была в проверки эффективности распараллеливания такой формулы. Оказалось, что циклы длиной до 12 (когда k ≤ 12) можно подсчитывать достаточно эффективно в произвольных графах из нескольких сотен вершин. Показано, что формулы допускают суперлинейное ускорение при распараллеливании, и что учет специфики графа может значительно повлиять на скорость расчета по ним.

Результаты исследований опубликованы в [8] и представлены авторами на международной конференции «Параллельные вычислительные технологии’ 2010» в г. Уфе в апреле 2010 года.

3. Задачи о расстановке ферзей и королей

Сколькими способами можно расположить 6 не бьющих друг друга шахматных ферзей на доске размером n × n?

В апреле 2010 года Vaclav Kotesovec (из Чехии) получил формулу для задачи о 5 ферзях, аналогичные формулы для 1, 2, 3 и 4 ферзях были известны ранее и получены другими авторами. Следующим шагом является задача о 6 ферзях, которая была успешно решена автором при помощи кластера КарНЦ.

Помимо научной работы, с использованием ресурсов вычислительного кластера ведется работа по подготовке специалистов по параллельным вычислениям. В 2009/2010 учебном году д.ф.-м.н., проф. Соколовым А. В. был прочитан учебный курс «Методы и алгоритмы параллельных вычислений», посвященный теоретическим и практическим аспектам реализации параллельных алгоритмов с помощью технологий MPI, OpenMP и др. Учебный курс прослушали студентов 6-го курса математического факультета Петрозаводского государственного университета.

Перспективы развития

На текущий момент налажена работа кластера и ряд исследователей привлечен для использования вычислительных ресурсов для решения своих научных задач. Загруженность вычислительных ресурсов кластера за последние три месяца достигла 35% процентов. Для удовлетворения возрастающих потребностей в вычислительных ресурсах на базе ЦКП планируется создание GRID, в которой (на первом этапе) будут задействованы компьютеры КарНЦ РАН. Вычислительные возможности GRID и кластера будут объединены для увеличения производительности и оптимального использования ресурсов.

Большое внимание следует уделить помощи исследователям в разработке программ. Широкое использование специализированных математических библиотек и технологий параллельного программирования позволит снизить время счета каждой из задач, а значит, эффективнее использовать вычислительные ресурсы кластера.

В своей научной работе вычислительные ресурсы кластера используют, в основном, сотрудники Института прикладных математических исследований. Это связано как со спецификой решаемых задач, так и с большей подготовленностью исследователей ИПМИ в области программирования. В будущем планируется привлекать к работе на кластере и сотрудников других институтов, при этом программы будут разрабатываться с помощью сотрудников ЦКП.

На текущий момент удаленный доступ к кластеру осуществляется по протоколу ssh. Однако в некоторых случаях такой доступ бывает неудобным для пользователей. Планируется постепенный перевод основных функций взаимодействия с кластером на web-интерфейс.

ЛИТЕРАТУРА:

1. Intel® Cluster Toolkit Compiler Edition 3.2.2 Release Notes [электронный ресурс] / Intel — URL: http://software.intel.com/sites/products/documentation/hpc/ictce/ictce_release_notes.pdf
2. Boost C++ Libraries [электронный ресурс] / URL: http://www.boost.org/
3. TotalView - Parallel and Thread Debugger on Multi-Core Linux, Mac OS X & Unix [электронный ресурс] / TotalView Technologies — URL: http://www.totalviewtech.com/products/totalview.html
4. Центр высокопроизводительной обработки данных ЦКП КарНЦ РАН [электронный ресурс] / Институт прикладных математических исследований Карельского научного центра РАН — URL: http://cluster.krc.karelia.ru
5. Ganglia (software) [электронный ресурс] / Wikipedia, the free encyclopedia — URL: http://en.wikipedia.org/wiki/Ganglia_%28software%29
6. Система управления заданиями Cleo / PARALLEL.RU — URL: http://www.parallel.ru/cluster/batch-system.html
7. Караваев А. М. Количество простых циклов на двумерной квадратной решетке : материалы / Высоко производительные параллельные вычисления на кластерных системах : материалы IX международной конфернции-семинара. г. Владимир. : Владимирский государственный ун-т., 2009. С. 202–207.
8. Караваев А. М., Воропаев А. Н. Эффективность распараллеливания явных формул для подсчета коротких циклов в графе : труды [электронный ресурс] / Параллельные вычислительные технологии’ 2010 : труды международная научная конференция. г. Уфа : Уфимский государственный авиационный технический ун-т. 1 CD-ROM. Заглавие с этикетки диска.