

ТЕХНОЛОГИЯ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ В НАУЧНЫХ ИССЛЕДОВАНИЯХ И УЧЕБНОМ ПРОЦЕССЕ ВОРОНЕЖСКОГО ГОСУНИВЕРСИТЕТА

С.А. Запрягаев, С.Д. Кургалин
Воронежский государственный университет

Введение

Высокопроизводительные параллельные вычислительные системы в последнее время получают распространение и широко используются при решении сложных научных задач. Такими системами все в большей степени оснащаются университеты. В значительном числе учебных заведений уже имеются и системы дистанционного обучения. Актуальной в настоящее время становится задача объединения этих систем для эффективной подготовки современных специалистов высокой квалификации. Целью настоящей работы является рассмотрение проблем интеграции и совместного использования этих систем в учебном процессе и научных исследованиях Воронежского государственного университета (ВГУ).

1. Параллельный компьютерный кластер Воронежского государственного университета

В 2002 г. в ВГУ при поддержке гранта VZ-010 фонда CRDF создан 20-процессорный высокопроизводительный вычислительный кластер в составе четырехпроцессорного сервера SUPERSERVER 8060 (4×CPU Pentium III Xeon, 700 МГц, 4 Гб SDRAM), четырех двух-

процессорных серверов SUPERSERVER 6010H (2×CPU Pentium-III, 1,13 ГГц, 1 Гб SDRAM), восьми рабочих станций (Pentium-4, 2,0 ГГц, 512 Мб RIMM). Пиковая производительность кластера - 28 Гфлопс. На кластере установлена операционная система Windows 2000. Узлы кластера первого уровня (четырёх- и двухпроцессорные) связаны по протоколу Gigabit Ethernet, для связи остальных узлов (однопроцессорных) реализован протокол Fast Ethernet. В кластере используется высокопроизводительное сетевое коммутационное оборудование (Rapier G6f и Rapier 24) компании Allied Telesyn Intern. Кластер связан оптоволоконным каналом с разветвленной внутренней сетью ВГУ и имеет высокоскоростной выход во внешнюю сеть, объединяющую ряд вузов г. Воронежа.

Система дистанционного обучения ВГУ создана на базе портала "Воронежский виртуальный университет" - "<http://Voronezh.OpenNet.ru>" [1] и программно-инструментальной среды Lotus LearningSpace 5. Действующая в ВГУ лаборатория дистанционного образования обеспечивает разработку на основе этой среды новых электронных курсов. В настоящее время в дистанционный учебный процесс включено 26 таких курсов. Используется и сформированная на сервере

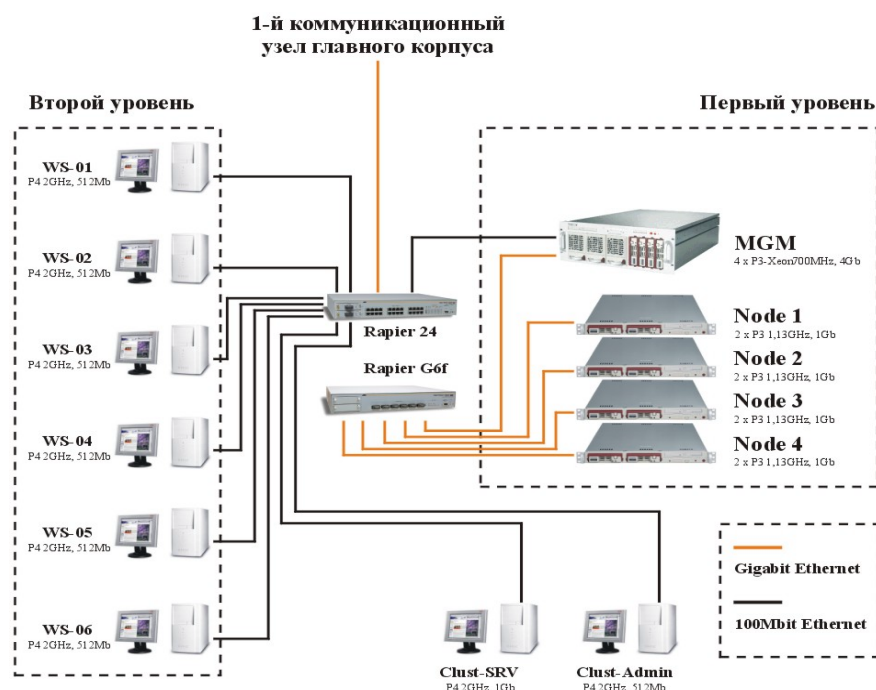


Рис. 1. Компьютерный кластер ВГУ

портала электронная библиотека учебно-методических ресурсов - более 500 учебных и методических пособий и 100 учебных видеоматериалов (Web-страница "<http://www.main.vsu.ru/video.html>").

Создание параллельного кластера существенно увеличивает возможности компьютерного моделирования сложных процессов и проведения ресурсоемких расчетов. Образование лаборатории высокопроизводительных вычислений ВГУ и включение ее в информационно-образовательную среду открывает новые перспективы для научных исследований в Воронеже, области и регионе [2]. Эта лаборатория интегрируется с лабораторией дистанционного образования ВГУ на основе высокоскоростного канала доступа к их ресурсам. На их базе организован учебно-научный комплекс для обеспечения научных исследований и учебного процесса. Комплекс используется и в системе повышения квалификации и профессиональной переподготовки (в институте повышения квалификации ВГУ).

Студентам ВГУ читаются курсы "Параллельные алгоритмы обработки данных", "Высокопроизводительные компьютерные системы", "Параллельное программирование". Параллельные алгоритмы изучаются также в курсах "Математическая логика и дискретная математика", "Вычислительный эксперимент и методы вычислений" и др. Выполняются курсовые и дипломные работы в области создания параллельных программ, использования возможностей кластера.

Для проведения лабораторных работ используется система "ПараЛаб" [3]. Она предназначена для изучения и исследования параллельных методов решения сложных вычислительных задач и позволяет моделировать работу параллельных компьютерных систем: задавать их топологию, устанавливать число процессоров и производительность каждого из них, определять характеристики коммуникационной среды, решать задачи сортировки, операций с матрицами, обработки графов с выбором метода решения задачи и определения объема исходных данных. Система дает возможность проводить и реальные параллельные эксперименты в режиме удаленного доступа к компьютерному кластеру.

Объединение параллельного кластера и регионального информационно-образовательного Интернет-портала "Воронежский виртуальный университет", созданного и администрируемого ВГУ и интегрированного в общероссийскую информационно-образовательную среду, позволяет обеспечить условия для дистанционного проведения сложных вычислительных экспериментов в научных и учебных целях.

2. Дистанционный доступ к кластеру по сети Интернет

Для дистанционного доступа по сети Интернет к вычислительным ресурсам кластера на базе про-

граммного обеспечения Microsoft Internet Information Services (MSIIS) создан специализированный Web-сервер [4], размещенный на 4-процессорном внутреннем узле кластера - SUPERSERVER 8060. Интерфейс сервера разработан на основе технологии Active Server Pages (ASP) на языке Java Script. Для удаленного доступа к кластеру необходима регистрация, осуществляемая администратором кластера средствами управления учетными записями операционной системы Windows 2000 Server. Зарегистрированным пользователям предоставляется возможность через Web-браузер размещать свои программы для выполнения на кластере, дистанционно компилировать и исполнять их. Первый вариант системы доступа обеспечивает функционирование программ, созданных в среде Microsoft Visual Fortran (MVF). Перед пересылкой и размещением на сервере программы архивируются. В каждом архиве может быть одно задание. Если в архиве находятся файлы проекта, созданного средой MVF, то параметры компиляции определяются содержимым этих файлов. При отсутствии файлов проекта запускаемая программа считается состоящей из одного модуля и компилируется в соответствии со спецификацией среды MVF как Win32 Release. Файлы с заданием после размещения в кластере разархивируются в каталог, имя которого совпадает с именем архива. Пользователь имеет возможность просматривать любой файл из этого каталога, удалять его или изменять содержимое. Кроме того, при создании новых файлов во время работы программы, например, с результатами расчетов, пользователь также может просматривать и эти файлы. Защита данных на Web-сервере осуществляется средствами MSIIS и файловой системы NTFS. Программы выполняются в пакетном (автоматическом) режиме без взаимодействия с администратором или пользователем. Результаты работы программы переносятся на компьютер пользователя средствами Web-браузера.

На следующем этапе развития программного обеспечения кластера предполагается создать ряд средств управления заданиями (прерывание или приостановка выполнения задания, изменение его параметров без перекомпиляции и др.).

Внедрение системы удаленного доступа к ресурсам кластера наряду с объединением кластера и портала "Воронежский виртуальный университет" позволяет обеспечить дистанционное проведение вычислительных экспериментов. Образование лаборатории высокопроизводительных вычислений кафедры цифровых технологий факультета компьютерных наук ВГУ и включение ее в информационно-образовательную среду открывает новые перспективы для развития научных исследований. На ее базе ведутся учебные занятия, проходят научные семинары, а также разрабатывается программное обеспечение для проведения исследований с использованием возможностей компьютерного кластера.

3. Разработка параллельных алгоритмов и программ. Использование кластера для компьютерного моделирования

Решение ряда актуальных задач в приоритетных научных направлениях зачастую не может быть осуществлено на традиционных последовательных компьютерах из-за слишком большого времени выполнения программ, поэтому использование высокопроизводительных компьютерных систем часто становится единственной возможностью для получения новых научных результатов.

Для эффективного использования компьютерного кластера возникает необходимость создания новых параллельных алгоритмов или переработки имеющихся последовательных программ. В процессе формирования библиотеки параллельных алгоритмов, которая должна обеспечить широкое использование кластера для расчетов в различных областях исследований, были разработаны или реализованы в возможных топологиях кластера ВГУ параллельные алгоритмы умножения квадратных матриц, расчета многомерных интегралов и многократных сумм, зависящих от нескольких переменных, решения "задачи коммивояжера", параллельных сортировок и др.

Актуальной является разработка методов параллельного компьютерного моделирования для решения сложных научных задач. К таким задачам, решаемым в настоящее время, относится, например, расчет вероятности процесса внутреннего тормозного излучения (ВТИ) [5] при α -распаде, моделирование спектров α -частичных состояний в ядрах и др. Созданный ранее комплекс последовательных программ для решения этих задач на традиционных компьютерах не обеспечил проведение полного цикла расчетов из-за катастрофического увеличения времени вычислений с ростом числа входных данных, поэтому имеющиеся последовательные алгоритмы были подвергнуты распараллеливанию. Применялись как известные приемы распараллеливания алгоритмов, так и вновь созданные для обеспечения эффективной работы конкретной параллельной программы комплекса. При этом использовались среда параллельного программирования MPICH 1.2.5 и язык Fortran.

На базе использования реалистических ядерных потенциалов создана новая теоретическая модель расчета спектров ВТИ, универсальная для α -, кластерного и протонного распадов [6, 7]. На ее основе было проведено компьютерное моделирование процесса ВТИ [8, 9] на параллельном кластере, получены и исследованы спектры ВТИ, найдены сходные черты и различия в поведении вероятностей и в областях формирования γ -излучения при α -, кластерном и протонном распадах.

4. Электронные библиотеки в информационно-коммуникационной среде вуза

Анализ ситуации в мире, проведенный в исследовании [10], показал, что в образовательных стандартах учебных заведений растет приоритет фундаментальных знаний, необходимых для достижения высоких результатов в последующей работе выпускников вузов. Подготовка современного специалиста проводится на основе использования последних достижений в области информационных технологий, в частности, методов математического и компьютерного моделирования в различных областях, причем число таких областей возрастает. При моделировании все шире применяются мультикомпьютерные системы (параллельные компьютерные кластеры), которые начинают использоваться в учебном процессе. В связи с этим возникает потребность в информационном обеспечении данного направления и создании соответствующих Интернет-ресурсов.

При освоении образовательных программ, насыщенных информационными технологиями, требуются библиотеки различного уровня и назначения, которые или изначально создаются как электронные, или трансформируются в них в следствие назревшей необходимости. Одним из видов библиотек, обеспечивающих фундаментальную подготовку специалистов, являются библиотеки компьютерных программ и алгоритмов. Однако известные библиотеки этого типа остаются пока узкоспециализированными и тесно связанными с системами программирования на определенных языках. Для обычного (непараллельного) программирования разработаны огромные библиотеки программ, охватывающие практически все области науки. Без них уже невозможно представить работу современного специалиста высокой квалификации. Создание библиотек параллельных алгоритмов с хорошими коммутационными свойствами, которые, с одной стороны, были бы специализированными и обеспечивали развитие научных исследований в определенной области знаний, а, с другой стороны, имели бы универсальную компоненту, используемую в разных отраслях, является одной из актуальных задач, решаемых в ВГУ. Разработка и использование таких электронных библиотек в реальном учебном процессе становится методом повышения его научного уровня.

В ВГУ в рамках создания учебно-научного комплекса формируется информационно-коммуникативная среда, включающая дистанционную систему обучения. В представительстве ВГУ в портале "Воронежский виртуальный университет" создана обширная электронная библиотека разработанных преподавателями и сотрудниками университета информационных и учебно-методических материалов в электрон-

ной форме для всех видов обучения. В настоящее время в ней размещены различные учебные материалы, в состав которых входят как полнотекстовые учебно-методические пособия, электронные сетевые учебные курсы, так и рабочие программы курсов, лабораторные работы, контрольно-измерительные материалы и т.д. Ведется работа по формированию в электронном виде учебно-методических комплексов по всем дисциплинам учебных планов.

Переход к компьютерным технологиям обучения обостряет проблему повышения квалификации преподавателей и требует использования новых приемов педагогической работы. Значительно усложняется деятельность преподавателей по разработке курсов. Важными их функциями становятся поддержка обучающихся в их продвижении по траектории образовательной программы, помощь в решении возникающих проблем при освоении разнородной информации.

Стратегической задачей для университетов в настоящее время является разработка и накопление электронных образовательных ресурсов для информационной поддержки системы образования. На решение этой задачи направлены усилия подразделений ВГУ, осуществляющих учебный процесс и обеспечивающих его качество.

5. Перспектива развития параллельных компьютерных систем

Как известно, компания Orion Microsystems представила принципиально новый класс компьютерной техники - персональные и недорогие настольные суперкомпьютеры Orion Cluster Workstation, каждый из которых намного превосходит по своим вычислительным возможностям обычные компьютеры. Эта компания намерена заполнить нишу, разделяющую обычные персональные компьютеры и суперкомпьютеры, начав производство высокопроизводительных персональных рабочих станций с передовой архитектурой на основе широко распространенных процессоров.

В настоящее время различие в производительности обычных компьютеров и суперкомпьютеров достигает трех порядков. Ключевой инновацией, открывшей путь к широкому распространению бытовых суперкомпьютеров, явилась разработка технологии управления энергопотреблением отдельных элементов компьютерной системы. В результате энергопотребление удалось снизить до значений, позволяющих подключать суперкомпьютер к бытовой электросети.

Модель DT-12 представляет собой блок из 12 процессоров Transmeta. В системе установлена память DDR SDRAM (до 24 Гбайт) и хранилище данных на жестких дисках объемом до 1 Тбайта. Производительность системы составляет 18 Гфлопс (36 Гфлопс в пиковом режиме). Энергопотребление системы - 220 Вт. Сис-

тема легко расширяется до 48-процессорной конфигурации.

Модель DS-96 выполнена в форм-факторе desktide и построена на 96 процессорах. Ее производительность - 150 Гфлопс в стандартном и 300 Гфлопс в пиковом режиме, объем ОЗУ - до 192 Гбайт, суммарный объем жестких дисков - 9,6 Тбайт. При этом ее общее энергопотребление - менее 1,5 кВт.

Габариты модели DT-12 составляют 50 × 60 × 10 см, модели DS-96 - 60 × 40 × 70 см.

Стоимость персональных суперкомпьютеров не выходит за рамки цен специализированных непараллельных рабочих станций. Персональные суперкомпьютеры работают под управлением операционной системы Linux и используют стандартные библиотеки параллельного программирования MPI, PVM и SGE. На Orion Cluster Workstation может использоваться безо всяких модификаций программное обеспечение Linux-кластеров.

Одним из первых применений этих систем стали биологические исследования. Компания BioTeam, занимающаяся консалтингом в области биоинформатики, разработала совместно с Orion суперкомпьютерную систему Cluster Workstation for Bioinformatics, представляющую собой 12-процессорный настольный кластер с установленным на нем программным обеспечением iNquiry, в состав которого входит более 200 приложений.

Следовательно, одним из перспективных направлений развития параллельных компьютерных систем становится создание персональных суперкомпьютеров, которые будут создавать конкуренцию обычным персональным компьютерам и постепенно заменять их в различных областях применения.

6. Грид-технологии и компьютерный кластер ВГУ

Со второй половины 1990-х годов технологии программирования глобальных распределенных ресурсов развиваются под названием Грид-технологий [11]. На их базе строятся системы для решения крупномасштабных задач в области фундаментальных и прикладных исследований.

Основа технологии Грид состоит в создании набора стандартизованных служб для обеспечения надежного, дешевого и безопасного доступа к географически распределенным высокотехнологичным информационным и вычислительным ресурсам - отдельным компьютерам, кластерам и суперкомпьютерным центрам, хранилищам информации, сетям, научному инструментарию и т.д. Учитывая хорошие параметры компьютерного кластера ВГУ и его расположение в центральной части России, перспективной является его интеграция в масштабные российские

проекты по разработке, апробации и развитию Грид-технологий.

В мире активно реализуется ряд Грид-проектов. В европейском проекте участвуют крупнейшие научные центры 14 стран, в том числе и России, по разработке глобальной информационной среды нового поколения для высокопроизводительной обработки больших объемов информации [12].

Первоначальный трехлетний европейский Грид-проект EDG - European DataGrid успешно завершился 31 марта 2004 г. Проект EDG сформировал тестовую вычислительную инфраструктуру, способную обеспечить вычислительные ресурсы и обработку географически распределенных данных для всего европейского научного сообщества.

1 апреля 2004 г. начался новый европейский Грид-проект - Enabling Grids for E-Science in Europe (EGEE). Он использует достижения проекта EDG и должен привести к дальнейшему развитию Грид-технологий. Его результатом будет создание общеевропейской непрерывно действующей Грид-инфраструктуры и формирование концепции всемирного Грид-компьютинга.

EGEE - двухлетний проект, который является частью четырехлетней программы. Результаты первых двух лет послужат основой для выработки целей последующих разработок. Выбраны две прикладные области для оценки производительности и функциональных возможностей создаваемой Грид-инфраструктуры. Первая область - распределенная обработка данных экспериментов на Большом Адронном Коллайдере - БАК (LHC Computing Grid - LCG). Вторая - биомедицинский Грид, где имеются огромные потоки данных в областях биоинформатики и здравоохранения.

В LCG используется программное обеспечение "промежуточного уровня" (MiddleWare, MW) LCG-1, в основе которого лежат разработки европейского проекта EDG (<http://marianne.in2p3.fr>) и американского проекта VDT (<http://www.lsc-group.phys.uwm.edu/vdt/project.html>), с использованием интеграционных решений проекта GLUE. EDG и VDT MW базируются на программном обеспечении Globus Toolkit (<http://www.globus.org>).

Компьютерный кластер ВГУ включается в общеевропейскую сеть Грид распределенных высокопроизводительных вычислений для решения крупномасштабных задач в области фундаментальных и прикладных исследований путем организации Воронежского регионального сегмента Грид. Кластер подключается в качестве ресурсного центра (Resource Centre) к российской системе RDIG (Russian Data Intensive GRID) европейского проекта EGEE. Проводится установка программного обеспечения для работ в пилотной прикладной области EGEE - распределенной обработке данных экспериментов на БАК.

Заключение

Использование высокопроизводительного вычислительного кластера позволяет продвинуться вперед в решении сложных задач и приобрести опыт создания параллельных алгоритмов и программ, необходимый для применения новых методов компьютерного моделирования в различных направлениях научных исследований и учебном процессе.

На базе кластера организован универсальный учебно-научный комплекс, в котором ведутся учебные занятия, проходят научные семинары, а также разрабатываются программные средства для обеспечения научных исследований и учебного процесса. Наличие компьютерного оборудования высокого уровня открывает новый путь для развития учебных технологий, связанных с использованием параллельных вычислительных систем и включением их в образовательные программы. Применение удаленного доступа, обеспечиваемого Интернет-технологиями, дает возможность использовать кластер в учебном процессе и научных исследованиях не только в ВГУ, но и в городе и регионе.

Использование комплекса позволило создать систему подготовки в области применения суперкомпьютеров и разработки параллельных алгоритмов. На его основе в образовательные программы и различные направления научных исследований внедрены эффективные методы математического моделирования.

Литература

1. Запругаев, С.А. Воронежский виртуальный университет как региональная часть единого информационно-образовательного пространства России / С.А. Запругаев, С.Д. Кургалин, А.П. Толстобров // Информационные технологии в образовании : мат-лы междунар. конф. - Болгария, 2002. - С. 149-152.
2. Запругаев, С.А. Высокопроизводительный вычислительный кластер в учебном процессе Воронежского университета / С.А. Запругаев, С.Д. Кургалин // Математика, компьютер, образование : 10 междунар. конф. - Пушино, 2003. - С. 26.
3. Гергель, В.П. ПараЛаб. Программная система для изучения и исследования методов параллельных вычислений / В.П. Гергель, А.А. Лабутина. - Н. Новгород : Изд-во ННГУ им. Н.И. Лобачевского, 2003. - 125 с.
4. Кургалин, С.Д. Система удаленного доступа к параллельному компьютерному кластеру / С.Д. Кургалин, В.В. Расхожев. - "Телематика'2004" : тр. XI Всерос. науч.-метод. конф., С.-Петербург, 2004. - СПб. : Изд-во СПбГУИТМО. - Т.1. - С. 92-93.
5. Баткин, И.С. Внутреннее тормозное излучение, сопровождающее α -распад / И.С. Баткин, И.В. Копытин, Т.А. Чуракова // Ядерная физика. - 1986. - Т. 44, Вып. 6 (12). - С. 1454-1457.

6. Кургалин, С.Д. Внутреннее тормозное излучение при кластерном и протонном распадах / С.Д. Кургалин, Ю.М. Чувильский, Т.А. Чуракова // Изв. РАН, сер. физ.-2001. - Т. 65, № 5. - С. 666-670.

7. Кургалин, С.Д. Моделирование спектров внутреннего тормозного излучения при альфа-распаде ^{214}Po и ^{226}Ra / С.Д. Кургалин, Ю.М. Чувильский, Т.А. Чуракова // Изв. РАН, сер. физ. - 2004. - Т. 68, № 7. - С. 1045-1048.

8. Кургалин, С.Д. Компьютерное моделирование многокомпонентных ядерных систем / С.Д. Кургалин // Компьютерное моделирование 2004 : тр. 5-й Междунар. науч.-техн. конф., С.-Петербург. - Ч. 2. - СПб. : Нестор, 2004. - С. 102-104.

9. Кургалин, С.Д. Моделирование эффектов кластеризации в ядерных системах с использованием парал-

лельного компьютерного кластера / С.Д. Кургалин // Научный сервис в сети Интернет : тр. Всерос. науч. конф. - М. : Изд-во МГУ, 2004. - С. 234-236.

10. Основы открытого образования / А.А. Андреев, С.Л. Каплан, Г.А. Краснова и др. ; отв. ред. В.И. Солдаткин. - Т. 1. - РГИОО. - М. : НИИЦ РАО, 2002. - 676 с.

11. *The Grid: Blueprint for a New Computing Infrastructure* / Ed. by I. Foster and C. Kesselman // Morgan Kaufman Pub., San Francisco, CA.- 1999.

12. Ильин, В.А. Создание российского сегмента европейской инфраструктуры EU DATAGRID / В.А. Ильин, В.В. Кореньков // Электронные библиотеки: перспективные методы и технологии, электронные коллекции : тр. IV Всерос. науч. конф. - Т. 1. - Дубна : ОИЯИ, 2002. - С. 239-248.