

СУПЕРКОМПЬЮТЕРНЫЕ И GRID-ТЕХНОЛОГИИ

С.М. Абрамов, А.А. Московский, В.А. Роганов (ИПС РАН),

П.Е. Велихов (SDSC, США)

В докладе рассматриваются общие вопросы, связанные с суперкомпьютерными и Grid-технологиями. Особенное внимание уделяется работам, выполняемым (или планируемым к выполнению) в Институте программных систем Российской академии наук (ИПСРАН).

Суперкомпьютеры и их использование

На каждом этапе развития средств вычислительной техники всегда находятся задачи, требующие для своего решения вычислительных мощностей, которые превышают мощности «обычных» компьютеров, доступных на рынке. Поэтому во все времена развития компьютерной отрасли ведущие страны мира занимались разработкой, реализацией и внедрением суперкомпьютеров.

Вокруг попыток дать определение термина «суперкомпьютер» всегда было много споров. Мы будем придерживаться следующего: компьютеры, включенные в некоторую редакцию мирового рейтинга пятисот самых мощных машин мира (Top500, [1]), безусловно, заслуживают применения к ним термина «суперкомпьютер». Рейтинг Top500 ведется с июня 1993 г. и обновляется два раза в год (в июне и ноябре). Для ранжирования вычислительных установок в рейтинге Top500 используется некая процедура вычисления «реальной производительности» на тестовой задаче (программе) Linpack. Получаемые при этом оценки производительности мы в данной работе будем называть Linpack-производительностью. Тест Linpack по своей сути сводится к решению системы линейных уравнений с большим числом переменных.

Существуют различные подходы к достижению высокой мощности (производительности) вычислительных установок [2]. В последнее время все большую популярность завоевывают архитектуры MPP и кластерный принцип построения суперкомпьютеров:

— используются широкодоступные компоненты — самые обычные процессоры, материнские системные платы (весьма часто — двухпроцессорные), модули памяти, жесткие диски;

— из этих компонентов собирают большое число (как правило, одинаковых) вычислительных узлов;

— вычислительные узлы соединяются между собой системной сетью, для этого используются либо существующие технологии высокоскоростных локальных сетей (например, Gigabit Ethernet), либо специализированные высокопроизводительные сетевые технологии (Myrinet, SCI, Infiniband и т.п.);

— системную сеть обычно используют только для интеграции вычислительной мощности вычислительных узлов, обычно это делается за счет реализации при помощи аппаратуры системной сети примитивов MPI [3];

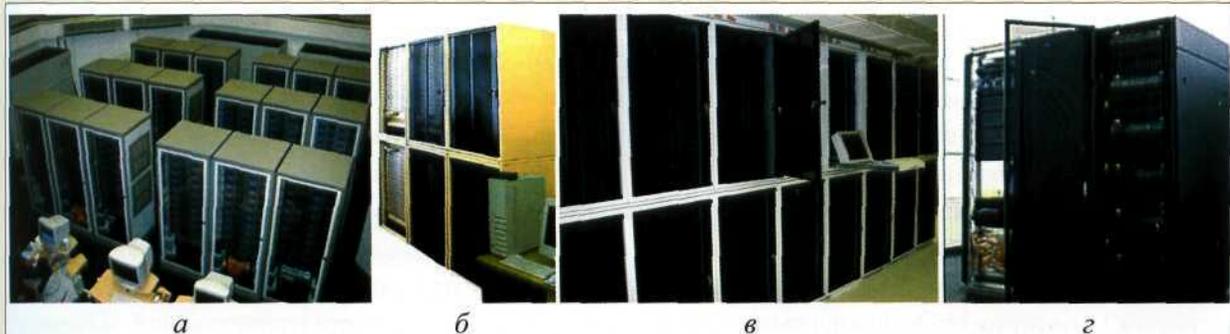
— часто кроме системной сети узлы связывают еще различными сетями: вспомогательной сетью, как правило, с протоколом TCP/IP (используется для передачи файлов и управления узлами) и сервисной сетью (например, для управления электропитанием, мониторинга и управления вычислительными узлами и т.п.).

Отечественные суперкомпьютеры. За всю историю развития отечественной суперкомпьютерной отрасли только четыре вычислительные установки — МВС-1000М, МВС-5000БМ, СКИФ К-500, СКИФ К-1000 — смогли попасть в мировой рейтинг Top500, четыре раза отечественные установки входили в «первую сотню» мирового рейтинга (МВС-1000М — 3 раза, СКИФ К-1000 — 1 раз) (см. таблицу и рисунок). Эти суперкомпьютеры разрабатывались и изготавливались:

— МВС-1000М, МВС-5000БМ — силами НИИ «Квант», ИПМ им. М.В. Келдыша РАН и

Отечественные суперкомпьютеры в мировом рейтинге Top500

Супер-компьютер	Разработчики и изготовители	Завершение изготовления	Место установки	Пиковая/ Linpack-производительность (млрд операций в секунду)	Тип процессора/тип системной сети	Число узлов/число процессоров	Место/выпуск рейтинга Top500
МВС-1000М	ФГУП «Квант», ИПМ РАН, МСЦ	Июнь 2002 г.	Межведомственный суперкомпьютерный центр (МСЦ, Москва, Президиум РАН)	1024/564, с сентября 2002 г. 1024/734,6	Alpha EV67 667 MHz/ Myrinet	384/768	64/июнь 2002 г., 74/ноябрь 2002 г., 95/июнь 2003 г., 189/ноябрь 2003 г., 392/июнь 2004 г.
СКИФ К-500	Суперкомпьютерная программа «СКИФ» Союзного государства. ОИПИ НАН Беларуси (Минск), НИИ ЭВМ (Минск), «Т-Платформы» (Москва), ИПС РАН (Переславль-Залесский)	Сентябрь 2003 г.	ОИПИ НАН Беларуси, Минск	716,8/423,6	Pentium IV Xeon 2,8 GHz/ SCI 3D	64/128	407/ноябрь 2003 г.
МВС-5000БМ	ФГУП «Квант», ИПМ РАН, МСЦ	Июнь 2004 г., модернизация в сентябре 2004 г.	Межведомственный суперкомпьютерный центр (МСЦ, Москва, Президиум РАН)	1075,2/722,1, с сентября 2004 г. 2112/1401	IBM eServer BladeCenter JS20 PowerPC970 1,6 GHz/ Myrinet	84/168, с сентября 2004 г. 168/336	399/июнь 2003 г., 210/ноябрь 2004 г.
СКИФ К-1000	Суперкомпьютерная программа «СКИФ» Союзного государства. ОИПИ НАН Беларуси (Минск), НИИ ЭВМ (Минск), «Т-Платформы» (Москва), ИПС РАН (Переславль-Залесский)	Сентябрь 2004 г.	ОИПИ НАН Беларуси, Минск	2534,4/2032	AMD Opteron 2,2 GHz/ InfiniBand	288/576	98/ноябрь 2004 г.



Внешний вид суперкомпьютеров МВС-1000М (а), СКИФ К-500 (б), СКИФ К-1000 (в), МВС-5000БМ (г)

Межведомственного суперкомпьютерного центра (МСЦ);

— СКИФ К-500, СКИФ К-1000 — в рамках суперкомпьютерной программы «СКИФ» Союзного государства (Россия и Беларусь) силами ОИПИ НАН Беларуси (Минск), НИИ ЭВМ (Минск), компании «Т-Платформы» (Москва) и ИПС РАН (Переславль-Залесский).

Полное название программы «СКИФ» — «Разработка и освоение в серийном производстве семейства моделей высокопроизводительных вычислительных систем с параллельной архитектурой (суперкомпьютеров) и создание прикладных программно-аппаратных комплексов на их основе».

С одной стороны, вхождение в список Top500 отечественных вычислительных установок говорит о том, что мы владеем технологиями, которые доступны только весьма ограниченному списку стран:

— суперкомпьютеры из списка Top500 установлены всего лишь в пяти десятках стран мира;

— суперкомпьютеры из списка Top500 способны изготавливать только страны, входящие в весьма узкий (из полутора десятка стран) «элитный клуб разработчиков суперкомпьютеров»;

— суперкомпьютеры из «первой сотни» способны изготавливать только в США, Японии, Китае и Союзном государстве.

С другой стороны, всего четыре суперкомпьютера за два с половиной года — это прискорбно мало для такой страны, как наша. Для примера, только в текущую редакцию Top500 (ноябрь 2004 г.) включено 17 установок, произведенных в КНР, причем три из них входят в «первую сотню».

Области применения суперкомпьютеров.

Если в начале развития суперкомпьютерной отрасли их использование было связано только с решением задач оборонного значения (как правило, задач, связанных с проектированием ядерного оружия), то сегодня суперкомпьютеры широко используют в самых различных отраслях народного хозяйства, науки, государственного управления, искусства и индустрии развлечений. Дело идет к

тому, что скоро будет проще назвать отрасль народного хозяйства или науки, где суперкомпьютеры не применяются, чем перечислять области применения суперкомпьютеров.

Чтобы дать представление о широте использования суперкомпьютеров сегодня, перечислим области, в которых используются вычислительные установки, создаваемые по суперкомпьютерной программе «СКИФ» Союзного государства России и Беларуси:

— космические исследования — обработка результатов дистанционного зондирования Земли (обработка радиолокационных и оптических сигналов), моделирования широкополосных пространственно-временных радиолокационных сигналов;

— физические процессы, аэродинамика и аэромеханика — расчеты обтекания плохообтекаемых тел, гиперзвуковое движение космического тела в плотных слоях атмосферы, параметров лазерного факела у поверхности твердотельной мишени в воздухе;

— химическая промышленность и фармакология — проектирование химических реакторов, создание веществ с прогнозируемыми свойствами, прогноз различных свойств веществ, исследование органических соединений (синтез лекарств), перспективных органических парамагнетиков, создание молекулярных парамагнитных кристаллов, расчеты люминесцентных характеристик веществ и т.п.;

— системы искусственного интеллекта — классификация текстов на естественном языке, извлечение знаний из текстов на естественном языке, интеллектуальное управление сложными системами;

— гидрометеорология и экология — моделирование атмосферных явлений, прогнозирование погоды, прогноз ветрового переноса загрязнений при лесном пожаре;

— различные инженерные расчеты — проектирование изделий в автомобильной промышленности (расчет несущих конструкций карьерных самосвалов БелАЗ и шахтных крепей, турбокомп-

рессоров для наддува дизельных двигателей и т.п.), анализ процессов в авиационных двигателях (детонационное горение, распространение пламени форсунки камеры сгорания газотурбинной установки) и оценка прочности авиационных газотурбинных двигателей), расчет механических характеристик сельскохозяйственных орудий и изделий легкой промышленности¹;

— геомеханические задачи — моделирование деформационных процессов на земной поверхности, устойчивости подземных сооружений, напряженно-деформированного состояния подработанной толщи земли;

— расчеты в интересах обороны и правоохранительных органов — идентификации личности по голосу, оптимизация назначения частот в группе РЭС, решение задач перебора большой размерности;

— медицина и здравоохранение — кардиологический комплекс нового поколения, моделирование процессов лазерного спекания порошковых материалов (для медицинских изделий).

Отметим, что данный список касается работ по весьма скромной по бюджету и продолжительности программе.

Проблемы суперкомпьютерной отрасли.

В процессе создания и применения суперкомпьютерных технологий легко обнаружить следующие проблемы:

— суперкомпьютерные системы весьма дорогие, но они довольно быстро «морально стареют». Например, суперкомпьютер, вошедший сегодня в «первую сотню» списка Top500, через 2—3 года вообще будет вытеснен из списка. Конечно, это не значит, что через 2—3 года суперкомпьютер выбрасывают на свалку, но для той организации (проблематики), для которой он создавался, скорее

всего, он уже станет непригодным. В этот момент самое правильное — передать такую технику в другую организацию (например, из НИИ передать в вуз). Однако разборка, транспортировка, сборка и наладка такого сорта установок — настолько хлопотное дело, что неплохо было бы всего этого избежать;

— дорогостоящая техника не должна простаивать. Но если не озаботиться объединением всех суперкомпьютеров страны в единую систему, то вполне возможны такие ситуации, когда будут наблюдаться перегрузка (огромные очереди на счет) суперкомпьютеров в одной части страны и простой, недогрузка вычислительных систем — в другой;

— какой бы мощный одиночный суперкомпьютер мы ни создали, всегда найдется задача, которую надо решить именно сегодня и для которой нужна большая мощность, чем производительность любого одиночного суперкомпьютера. При этом суммарная производительность всех существующих сегодня супер-ЭВМ вполне может быть достаточной для решения данной задачи.

Все эти проблемы довольно очевидны. А поиск их решения однозначно приводит к идее интеграции разрозненных компьютерных ресурсов в единую территориально-распределенную систему. Это и есть основная идея Grid-систем и Grid-технологий.

Grid-технологии

В данном докладе под Grid-технологиями мы будем понимать набор средств и технических решений (аппаратных, программных, организационных и т.п.), которые позволяют разрозненные разнородные компьютеры и суперкомпьютеры объединять в территориально-распределенную гетерогенную информационную и вычислительную систему. При этом основной задачей будет интеграция всех вычислительных ресурсов компьютеров, входящих в систему.

Основные идеи Grid-систем. Основные идеи Grid-технологий заимствованы из практики

¹ Речь идет о реальных заказах на расчеты динамических моделей для швейных изделий, и это еще раз подчеркивает, что в ближайшем будущем в любой отрасли народного хозяйства без использования суперкомпьютеров будет невозможно создавать конкурентоспособную продукцию.

работы национальных электрических сетей (power Grid), которые включают в себя распределенных по огромным территориям потребителей электроэнергии, линии передачи и генерирующие мощности различных (разнородных по своему устройству) электростанций. Потребитель электроэнергии обслуживается этой гигантской системой, и ему не важно знать, какого типа электростанция в настоящее время питает его оборудование (тепловая, атомная, гидроэлектростанция или другая), какие линии передачи электроэнергии задействованы для этого. Электрическая сеть поддерживает разные аспекты такого обслуживания: эффективное использование в национальном масштабе имеющихся генерирующих мощностей, переброса избытка мощности из одного региона в другой, использование резервных линий для нейтрализации последствий аварий на линиях передачи электроэнергии или на электростанциях и т.п.

Сегодня цели и планы разработки Grid-технологий и Grid-систем не менее важны и грандиозны, чем цели и планы национальных электрических систем, создаваемых в первой половине двадцатого века. Необходимо объединить в единую систему различные по технической реализации и типам компьютерные ресурсы (вычислительные ресурсы, ресурсы хранения и передачи информации) и донести совокупный ресурс до потребителя. Потребитель должен получать услуги от системы в целом, ему не важно знать, где и какая установка хранит или обрабатывает его информацию, какого типа данная установка, какие линии передачи информации при этом задействованы и т.п. Так же, как и в электрических сетях, высокопроизводительная вычислительная Grid-система должна обеспечивать эффективное использование всей совокупности ресурсов (вычислительных, хранения информации, исходных данных, т.е. источников данных), нейтрализацию последствий аварий на линиях передачи, в устройствах хранения или обработки информации. Создание подобных систем должно кардинально улучшить эффективность использования совокупных компьютерных ресурсов страны.

Таким образом, Grid — это технология создания эффективных территориально-распределенных гетерогенных сетей, объединяющих компьютеры с самыми различными аппаратными и программными системами. Основная задача Grid — реализация гибкого, защищенного, скоординированного вычислительного пространства для совместного использования ресурсов между динамически меняющимися сообществами пользователей.

Данная технология призвана осуществлять хранение информации и высокопроизводительные параллельные вычисления в сети территориально-распределенных вычислительных средств и других ресурсов: суперкомпьютеров, отдельных серверов, мэйнфреймов, систем хранения и баз данных (с разной реализацией этих баз).

На основе Grid-технологий можно создавать мощные информационно-вычислительные среды, обладающие также уникально низким соотношением «эксплуатационные расходы/производительность», за счет повышения коэффициента использования всей совокупности ресурсов, включенных в Grid-сеть. Например, средняя загрузка 6000 процессоров собственной Grid-сети компании Sun Microsystems, находящейся в эксплуатации, составляет 98%, в то время как ПК-серверы используются в среднем на 5—20%.

Различные аспекты интеграции ресурсов в Grid-системах. Итак, основная идея Grid-систем — интеграция ресурсов, и здесь важно помнить, что ставится задача интеграции самых различных ресурсов, включаемых в систему:

— интеграция вычислительной мощности. Grid-система «в пике» должна предоставлять пользователю вычислительную мощность, равную сумме мощности всех компьютеров, включенных в состав системы;

— интеграция источников данных (баз данных). Grid-система должна интегрировать все базы данных в единую и предоставлять ее пользователю (конечно, с учетом прав доступа). Эта единая база данных должна позволять наиболее полно разрешать запросы на доступ к данным. При этом

сложность возможного запроса и полнота выборки в «интегральной базе данных» должны значительно превосходить то, что может обеспечить каждая отдельная база данных, включенная в состав Grid-системы;

— интеграция ресурсов хранения. Grid-система должна предоставлять пользователю возможность хранить данные в объеме, значительно превышающем емкость средств хранения данных любой из установок, включенных в состав системы;

— интеграция ресурсов передачи информации. Grid-система должна предоставлять пользователю возможность передавать данные с лучшими показателями (задержка и скорость передачи), чем передача данных в рамках одного сеанса (по одному маршруту связи);

— интеграция ресурсов взаимодействия персонала. Grid-система должна предоставлять пользователю широкие возможности взаимодействия групп людей (Access-Grid, многопользовательские телеконференции, распределенное взаимодействие в процессе решения некоторой задачи).

Особого внимания заслуживают первые два аспекта: интеграция вычислительной мощности и интеграция баз данных.

Развитие работ по Grid-технологиям за рубежом. Grid — интенсивно развивающаяся перспективная технология. Поскольку Grid-технологии сегодня находятся в начале своего жизненного цикла, фактически на стадии научно-исследовательских разработок, страны, обладающие необходимым научно-техническим потенциалом и практическим заделом, при целенаправленной государственной поддержке имеют все необходимые предпосылки включиться в разработку Grid-технологий и их своевременное освоение, тем самым не допустить отставания в этом новом направлении.

В настоящее время практически в каждой развитой стране развернуты национальные Grid-проекты, имеющие целью создание соответствующей инфраструктуры и развитие технологий, обеспечивающих удаленный доступ к разнообразным вычислительным ресурсам независимо от места

расположения потребителя (например, подобную программу недавно приняла Болгария).

В Европе все работы по этой технологии координируются в рамках проекта Data Grid (в него уже вложено 9,8 млрд евро), а в Великобритании для разработок по Grid создан отдельный государственный орган.

Объем ежегодных инвестиций в Grid-проекты с 2003 по 2008 г. вырастет по прогнозам в 20 раз (с 250 млн до 4,9 млрд долл. США).

Объясняется это тем, что институты современного общества, такие, как правительственные организации, занимающиеся вопросами национальной безопасности и долгосрочного планирования, организации здравоохранения, мониторинга окружающей среды, научно-исследовательские институты, предприятия электронного бизнеса имеют распределенную природу и нуждаются в открытой и стандартизированной программно-телекоммуникационной инфраструктуре, обеспечивающей гибкое, безопасное и скоординированное динамически разделяемое использование распределенных вычислительных ресурсов.

Использование Grid-технологий позволяет решать научные, научно-практические, инженерные, медицинские и социально значимые задачи высокой информационной сложности, к решению которых до появления Grid невозможно было подступиться.

В США созданы четыре мощные национальные сети, обслуживающие основные учреждения, занятые научными исследованиями и информационным обслуживанием, — «Компьютерная сеть национального фонда исследований» (NSF Computational Grid), «Информационная сеть поддержки НАСА» (NASA Information Power Grid), «Глобальная информационная сеть Министерства обороны» (DOD Global Information Grid), «Сеть суперкомпьютерной инициативы Министерства энергетики» (DOE ASCI Grid).

В частности, проект ASCI Grid, объединяющий три самых мощных суперкомпьютера программы ASCI, предназначен в первую очередь для обеспечения расчетов, необходимых при решении

задач совершенствования и поддержки боевого ресурса ядерных зарядов без проведения натуральных испытаний (ядерных взрывов).

Под руководством Пенсильванского университета на базе Grid-технологии создан Национальный цифровой центр маммографии с объемом хранимых данных 5,6 петабайт ($5,6 \cdot 10^{15}$), что позволяет тысячам больниц и госпиталей хранить маммограммы в цифровой форме, а медицинским специалистам получать практически мгновенный доступ к записям пациентов.

В компании Sun Microsystems находится в эксплуатации собственная Grid-сеть, включающая 6000 процессоров и 210 Тбайт данных, которая ежедневно обчисляет более 50 000 задач электронного проектирования.

В 2004 г. в США объявлена президентская стратегическая Grid-программа (Strategic Grid Computing Initiative), основной целью которой является создание единого национального пространства высокопроизводительных вычислений (National High Performance Computing Environment). Американские эксперты оценивают инвестиции в создание высокопроизводительных компьютерных сетей стратегическим вкладом государства в ускоренное развитие своего научно-технического, промышленного и оборонного потенциала.

Самым последним подтверждением значимости развития Grid-технологий на базе суперкомпьютерных систем явилось принятие 30 ноября 2004 г. Сенатом США закона 108-423 о создании при Министерстве энергетики Центра высокопроизводительных вычислительных систем (High-End Software Development Center) с бюджетным финансированием на период 2005—2007 г. в объеме 165 млн долларов.

Основными задачами этого центра являются:

— проведение исследований и перспективных разработок математического обеспечения для высокопроизводительных вычислительных систем Министерства энергетики США;

— развитие и внедрение результатов разработки матобеспечения для высокопроизводитель-

ных вычислительных систем в перспективные прикладные научные и инженерные проекты, реализуемые Министерством энергетики.

По положению закона центр должен координировать свои усилия в области разработки математического обеспечения для высокопроизводительных вычислительных систем с другими федеральными агентствами и ведомствами США, такими, как Национальный научный фонд, Агентство перспективных оборонных исследований и проектов, Национальная администрация ядерной безопасности, Агентство национальной безопасности, Национальное агентство по авиации и космонавтике, Национальная администрация по исследованию мирового океана и атмосферы, Национальный институт стандартов и технологий, Агентство по охране окружающей среды, Национальные институты здравоохранения, что в целом показывает масштабы развернутых в США работ по созданию высокопроизводительных вычислительных систем и разработке Grid-технологий.

Также планируется широко использовать Grid-технологии в качестве одного из главных (системообразующих) средств в борьбе с современными вызовами и угрозами. Так, например, Grid-технологии планируется применять в системах идентификации личности по биометрическим параметрам, которые будут включать в себя базы биометрических данных и аппаратные средства снятия и обработки биометрической информации в реальном масштабе времени на базе высокопроизводительных гетерогенных распределенных Grid-систем. Эти проекты осуществляются для контроля за перемещениями иностранцев через границы США и стран Западной Европы, а также для сбора биометрической информации на всех «подозрительных» гражданах этих стран и стран, входящих в список так называемой «оси зла», что обусловлено, в первую очередь, необходимостью противостояния мировому терроризму и необходимостью борьбы с криминальным оборотом наркотиков и распространением оружия.

В Китае с 2000 г. ведутся работы по освоению Grid-технологий и к настоящему времени в еди-

ную сеть высокопроизводительных вычислений объединены основные научно-производственные центры страны, расположенные в четырех провинциях.

Инфраструктура для реализации Grid-технологий. Для создания Grid-систем необходимо развивать соответствующие инфраструктуры: суперкомпьютерные центры и центры данных, а также высокоскоростные национальные системы передачи данных. В большинстве стран в качестве телекоммуникационной платформы для Grid-проектов часто используются сети для науки и образования (Science & Education Networks), среди которых необходимо отметить:

— крупнейшую национальную сеть для образования и науки NSFNet в США;

— магистральную европейскую сеть для образования и науки GEANT (Pan-European Backbone, [4]), объединяющую научные организации в 30 европейских странах (наиболее часто здесь используются каналы емкостью 10 и 2,5 Гб/с, реже — 622 и 155 Мб/с);

— сеть для образования CSTNet, объединяющая 800 научно-исследовательских институтов и вузов Китая;

— российская магистральная сеть науки и образования RBNet (Russian Backbone Network, [5]).

Особое место среди сетей для науки и образования занимает сеть GLORIAD (Global Ring Network for Advanced Applications Development, [6]). Это волоконно-оптическое кольцо, опоясывающее северное полушарие Земли, призвано объединить между собой национальные научно-образовательные сети России, Китая и США. Сеть GLORIAD создана в 2003 г. по инициативе академика Е.П. Велихова. Сейчас рассматривается вопрос вхождения в этот проект национальных научно-образовательных сетей еще нескольких стран. Сегодня GLORIAD обеспечивает емкость каналов 155 Мб/с, планируется расширение данной емкости до 10 Гб/с.

Таким образом, важнейшими частями аппаратных платформ для Grid-систем являются суперкомпьютеры, суперкомпьютерные центры

и центры данных, а также объединяющие их и потребителей высокоскоростные сети передачи данных. Необходимо отметить, что в деле создания Grid-систем техническая сложность и нерешенность большого числа проблем сегодня наиболее остро чувствуются в части программного обеспечения, нежели в аппаратной части. Особенно это верно для поддержки интеграции вычислительной мощности и источников данных.

GEMS: интеграция баз данных. Средства интеграции источников данных (интеграции баз данных) ИПС РАН разрабатывает в проекте GEMS (Grid-Enabled Mediation Service), который выполняется в кооперации с Суперкомпьютерным центром Сан-Диего (США) и ИСП РАН.

Цель проекта — разработка и реализация системы интеграции данных и приложений, основанной на открытых стандартах (XML и Grid) и следующих ключевых идеях:

— используется декларативный, функциональный язык XQuery, в котором в качестве модели данных принимается XML;

— язык XQuery дает возможность интегрировать данные и приложения (экспортируемые в Grid в качестве сервисов);

— использование XML в качестве стандарта обмена данными дает возможность одновременно работать со структурированными и неструктурированными данными.

На первом этапе проекта, кроме проведения научно-исследовательских работ, планируется реализовать демонстрационное приложение, на котором можно было бы изучать сильные и слабые стороны предлагаемого подхода, выработать решения по совершенствованию системы.

T-система: интеграция вычислительных мощностей. T-система — система параллельного программирования, реализующая концепцию автоматического динамического распараллеливания программ. Разработка и первые подходы к реализации T-системы были выполнены в ИПС РАН в 1994—1999 гг. С 2000 г. к работам по развитию и реализации T-системы подключились новые

группы исследователей, в первую очередь из МГУ им. М.В. Ломоносова.

Реализованная сегодня версия Т-системы — Open T-System (OpenTS) — является в некотором смысле надстройкой над C++/MPI. Программы для OpenTS пишутся на языке T++, который является диалектом (расширением) языка C++. Программист при этом освобожден от таких работ для организации параллельного выполнения Т-программ:

- распараллеливание — выделение фрагментов кода в программе, который может выполняться параллельно;

- планировка вычислений—распределение параллельных фрагментов программы по вычислительным узлам кластера, обеспечение загрузки всех вычислительных узлов кластера;

- обмен данными — обеспечение передачи входных и выходных данных к фрагментам программы, распределенным по различным узлам кластера;

- синхронизация параллельных фрагментов кода — приостановка выполнения программного кода, если необходимые для счета данные пока еще не доступны, и возобновление счета при поступлении этих данных.

Все указанные действия выполняет Т-система без участия программиста (автоматически). Причем эти действия определяются и выполняются в динамике во время исполнения программы (а не планируются заранее, в статике, во время компиляции).

В перспективе очень важны будут следующие свойства Т-системы:

- алгоритм динамического планирования вычислений (распределения параллельных фрагментов программы по вычислительным узлам) отделен от пользовательской программы. Планировщик входит в состав ядра Т-системы. Можно использовать разные планировщики для одной и той же Т-программы. В том числе в планировщике можно учитывать неоднородность вычислительной установки: разные задержки и разные пропускные способности при передаче данных

между парами узлов, разные производительности узлов и т.п.;

- Т-система реализована таким образом, чтобы требования к задержкам во время обмена данными не были очень жесткими. Это поддерживается следующим: в каждом вычислительном узле можно ожидать достаточное количество параллельных фрагментов кода; используются асинхронные передачи данных.

Эти свойства сделали перспективными попытки использования Т-системы не только для кластеров, но и метаclusterных установок и территориально-распределенных неоднородных установок. По сути, имелись предпосылки для того, чтобы попробовать использовать Т-систему в качестве подхода к реализации интеграции вычислительных мощностей в Grid-системах.

Предпосылкой и базисом для развития данного подхода явились созданные в рамках программы «СКИФ» clusterные и метаclusterные конфигурации широкого диапазона производительности, а также технологии динамического распараллеливания вычислений (OpenTS), мониторинга и управления (Flame) этими конфигурациями. При этом для поддержки метаclusterных передач данных удалось использовать различные реализации MPI: MPICH-G2, IMPH, PACX-MPI. Первые эксперименты по метаclusterным вычислениям, проведенные в рамках программы «СКИФ», показали, что эти технологии могут быть расширены до полноценных Grid-решений.

T-Grid — испытательный стенд для отработки использования Т-системы для интеграции вычислительных мощностей в Grid-системах. Для отработки технологии использования OpenTS в качестве основы для интеграции вычислительных мощностей в Grid-системах в ИПС РАН организован испытательный стенд T-Grid. При этом:

- используется сетевая инфраструктура Переславля-Залесского — ИПС РАН объединил здесь скоростной (100/10 Мб/с) компьютерной сетью практически все учреждения и многих жителей этого города;

— используются компьютерные ресурсы ИПС РАН, Университета г. Переславля-Залесского и частично ресурсы «домашних компьютеров» добровольных участников проекта T-Grid.

Для поддержки разнородных вычислительных ресурсов (Windows-серверы, Linux-серверы, кластеры) и упрощения управления системой используется механизм виртуальных машин. В каждой реальной машине, часть ресурсов которой решено использовать в системе T-Grid, реализуется виртуальная машина с OS Linux. При этом для Windows-серверов используется Cooperative Linux, для Linux-серверов — UML. Далее в систему T-Grid уже объединяются эти самые виртуальные Linux-машины.

Такой подход, среди прочего, решает проблему «отчуждения» ресурсов в Grid-систему. Передавая какие-то ресурсы в T-Grid, соответствующий администратор сервера и организация-владелец сервера передают всего лишь «виртуальную часть» сервера, сохраняя за собой все административные права на сервере и не передавая администраторам T-Grid ничего сверх обычных (непривилегированных) пользовательских прав.

Перспективы развития работ

Первые результаты, полученные по проектам GEMS и T-Grid, оказались достаточно обнадеживающими. Поэтому в планы ИПС РАН входит продолжение работ по данным проектам. В сотрудничестве ИПС РАН с белорусскими и российскими организациями были подготовлены предложения по формированию новой суперкомпьютерной программы Союзного государства «Разработка и использование программно-аппаратных средств Грид-технологий и перспективных высокопроизводительных (суперкомпьютерных) вычислительных систем семейства СКИФ (шифр СКИФ-Грид), включающие четыре направления работ:

— Grid-технологии — развитие, исследование и внедрение средств высокопроизводительных вычислений на основе Grid-технологий,

поддержка гетерогенных, территориально-распределенных вычислительных комплексов;

— суперкомпьютеры семейства СКИФ (ряд 3 и 4) — создание суперкомпьютеров СКИФ нового поколения на базе новых перспективных процессоров и вычислительных узлов, новых технических средств системной сети, управления системой, спецвычислителей и гибридных узлов, разработка соответствующего программного обеспечения;

— защита информации — реализация (аппаратных и программных) средств защиты информации в создаваемых вычислительных комплексах;

— пилотные системы — реализация прикладных систем в перспективных областях применения создаваемых вычислительных установок, решение актуальных задач на суперкомпьютерах и Grid-системах, усилия по подготовке и переподготовке кадров в области суперкомпьютерных и Grid-технологий.

Заключение

Работы, положенные в основу данной статьи, были выполнены в рамках:

— суперкомпьютерной программы «СКИФ» Союзного государства;

— программы фундаментальных научных исследований ОИВТС РАН «Высокопроизводительные вычислительные системы, основанные на принципиально новых методах организации вычислительных процессов»;

— программы фундаментальных исследований Президиума РАН «Разработка фундаментальных основ создания научной распределенной информационно-вычислительной среды на основе технологии Grid».

Особую благодарность авторы выражают академику-секретарю ОИВТС РАН, Е.П. Велихову за его интерес, постоянную помощь и поддержку наших работ в области суперкомпьютерных и Grid-технологий. Более того, справедливости ради надо отметить, что в России академик Е.П. Велихов является инициатором исследований в облас-

ти Grid-технологий и многих смежных разработок и проектов (глобальные и национальные научно-образовательные компьютерные сети, суперкомпьютерные технологии).

Авторы благодарны всем участникам суперкомпьютерной программы «СКИФ» (особенно белорусским коллегам — С.В. Абламейко, В.В. Анищенко, Н.Н. Парамонову, О.П. Чижу и др.), с которыми авторов связывает опыт плодотворной и успешной работы в программе «СКИФ». Надеемся, это сотрудничество будет продолжено и дальше в рамках программы «СКИФ-Грид».

Список литературы

1. TOP500 *Supercomputer Sites* — мировой рейтинг пятидесяти самых мощных компьютеров мира. — Информационный ресурс в сети Интернет, <http://www.top500.org/>.
2. Воеводин В.В., Воеводин Вл.В. *Параллельные вычисления*. — СПб.: БХВ-Петербург, 2002. 609с.
3. *Message Passing Interface Forum*. — Информационный ресурс в сети Интернет, <http://www.mpi-fomm.org/>.
4. The GEANT Website. — Там же, <http://www.geant.net/>.
5. Магистральная сеть науки и образования RBNet (Russian Backbone Network). — Там же, <http://www.riprn.mt:8081/rbnet/>.
6. Global Ring Network for Advanced Applications *Development (GLORIAD)*. — Там же, <http://www.gloriad.org>.
7. Абрамов С.М., Адамович А.И., Инюхин А.В., Московский А.А., Роганов В.А., Шевчук Е.В., Шевчук Ю.В. *T-система с открытой архитектурой*. — В сб.: Труды Международной научной конференции «Суперкомпьютерные системы и их применение» (SSA2004). Минск, ОИПИ ИАИ Беларуси, 26—28 октября 2004, с. 18—22.
8. Абрамов С.М., Адамович А.И., Коваленко М.Р. *T-система — среда программирования с поддержкой автоматического динамического распараллеливания программ. Пример реализации алгоритма построения изображений методом трассировки лучей*. — Программирование, 1999, т. 25(2), с. 100—107.