РАЗВИТИЕ КРАСНОЯРСКОГО ЦЕНТРА ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ*

С. В. Исаев, А. В. Малышев, В. В. Шайдуров
Институт вычислительного моделирования СО РАН, Красноярск, Россия
e-mail: si@icm.krasn.ru, amal@icm.krasn.ru, lena@icm.krasn.ru

The paper reviews how parallel computations are organized at the Krasnoyarsk Scientific Center of the Siberian Branch of the Russian Academy of Science (KSC SB RAS). It also discusses its existing computing and telecommunication resources and how they will be developed.

Красноярский центр параллельных вычислений создан в 2002 году с началом эксплуатации и обслуживания приобретенной Институтом вычислительного моделирования (ИВМ) СО РАН многомашинной вычислительной системы кластерной архитектуры МВС-1000/16, разработанной в Научно-производственном объединении "Квант" и ИПМ РАН.

Сотрудники центра решают следующие основные задачи:

— регистрация, администрирование и консультирование пользователей кластера;

— организация подключения кластера к сетям общего пользования, в том числе к сети Интернет;

— настройка системы безопасности на управляющем узле кластера для организации телекоммуникационного доступа с использованием протоколов шифрования;

— обеспечение бесперебойной работы системы, в том числе организация системы электропитания, охлаждения, аварийной остановки, мониторинга неисправностей;

— своевременный ремонт и замена выходящих из строя комплектующих, по возможности без остановки процесса вычислений;

— установка дополнительного программного обеспечения и настройка существующего.

Проблемы по запуску, подключению к сети, настройке программного обеспечения кластера были успешно разрешены, и к середине 2002 года система работала в круглосуточном режиме. Доступ к кластеру получили пользователи корпоративной сети, объединяющей все научные учреждения Красноярского научного центра СО РАН (КНЦ СО РАН), а также Красноярского государственного технического университета, Красноярского государственного университета и других вузов, подключенных к Красноярской научно-образовательной сети (рис. 1).

Благодаря программной совместимости МВС-1000/16 с МВС-1000М Межведомственного суперкомпьютерного центра РАН у пользователей Красноярского научного центра появилась возможность переносить на МВС-1000М программы, требующие большей вычислительной мощности.

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 05-07-90201-в).

В настоящее время МВС-1000/16 функционирует в непрерывном режиме. Ведутся работы по оптимизации использования двух компьютерных классов Красноярского государственного технического университета в качестве вычислительного кластера, состоящего из 22 компьютеров. Многомашинная вычислительная система МВС-1000/16 ИВМ СО РАН имеет кластерную архитектуру и состоит из 16 узлов в общей стойке, один из которых является управляющим, а остальные — вычислительными (рис. 2). Пиковая производительность данной системы составляет около 14.0 млрд оп/с.

Рис. 1. Логическая схема соединений научно-образовательной оптоволоконной сети.

Рис. 2. МВС-1000/16 Института вычислительного моделирования СО РАН.

На суперкомпьютере установлено следующее программное обеспечение:

— коммуникационные среды-реализации MPI (MPICH ver 1.2.0, LAM ver 6.3.3);

— компиляторы (GNU C/C++/F77 2.96, Intel C/C++ compiler 7.0 noncommercial);

— пакет Intel Math Kernel Library v. 5.2 SP1, включающий в себя следующие оптимизированные библиотеки — BLAS — Basic Linear Algebra Subprograms, библиотека элементарных векторных и матричных операций; FFTs — библиотека для выполнения одно- и двумерного быстрого преобразования Фурье; LAPACK — библиотека для выполнения матричных преобразований, решения СЛАУ и проблемы собственных значений; Vector Math Library (VML) — библиотека для вычисления элементарных функций над векторами;

— системы параллельного программирования (пакет SCALAPAC, ver. 1.7 — параллельная версия пакета LAPACK; система DVM, ver. 3.82).

Наиболее распространенные расчеты — исследование свойств наноструктур и молекулярных соединений — велись в трех институтах КНЦ СО РАН с разными версиями, в том числе параллельными, готового программного комплекса HyperChem, предоставленными зарубежными партнерами или взятыми в открытом распространении. Исследования велись в области кристаллов (ИФ — Институт физики СО РАН), свойств химических соединений (Институт химии и химической технологии СО РАН), молекулярных свойств биологических соединений (ИФ СО РАН, ИВМ СО РАН). Без МВС-1000/16 исследования таких сложных молекул были бы невозможны. С применением МВС-1000/16 и МВС-1000М велись расчеты давления солнечного ветра и теплового давления Земли на искусственные спутники Земли с целью уточнения параметров их орбит в реальных условиях с учетом переотражения и специальных индикатрис рассеяния и поглощения, а также начато решение приливных уравнений Лапласа для Мирового океана. Приемлемая точность расчетов стала возможной лишь при использовании многопроцессорной вычислительной техники.

В настоящее время на кластере зарегистрировано 25 пользователей. С учетом резервирования под отладочные задачи пяти вычислительных узлов из 15 реальная суммарная загрузка процессоров составляет около 60 % и близка к максимально возможной.

Главными задачами МВС-1000/16 в настоящее время в связи с недостаточной производительностью являются обучение параллельному программированию и отладка параллельных программ, которые затем могут запускаться на кластере МВС-1000М Межведомственного суперкомпьютерного центра. Институт координирует подключение пользователей КНЦ СО РАН к последнему кластеру. В настоящее время зарегистрировано семь пользователей, работающих над четырьмя проектами.

В связи с невозможностью модификации существующего кластера и имеющейся потребностью в вычислительных ресурсах в 2004 году начались работы по проектированию и началу сборки новой многопроцессорной вычислительной системы МВС-1000/25 для замены имеющейся.

Произведено исследование архитектуры и конструктивных решений современных многопроцессорных систем. Сделаны выводы и разработаны конструктивные решения по созданию в ИВМ СО РАН кластера на основе 24 вычислительных и одного управляющего узла двухсетевой архитектуры.

Для выбора процессорной платформы выполнен ряд тестов на примерно одинаковых по стоимости конфигурациях AMD Athlon 64 3400 — далее AMD64 и Pentium4 3000 — далее P4. Использовалась операционная система Gentoo Linux, версия дистрибутива — 2004.2.

Для тестирования применялось следующее программное обеспечение:

— компилятор GNU C/C++/Fortran77 v.3.3.4 для обеих платформ;

— компилятор Intel C/C++/Fortran9x v.8.1 для обеих платформ;
— компилятор PathScale C/C++/Fortran9x v.1.4 для AMD64.

Категории тестов:
— вычисления с плавающей точкой;
— целочисленные вычисления;
— интегральный тест;
— тест сетевой производительности.

Операционная система Gentoo Linux — это относительно молодой проект, представляющий собой открытый, компактный, быстрый, популярный и быстро развивающийся дистрибутив. К его важным особенностям относится явная поддержка 64-битных архитектур, в частности AMD64.

Компилятор GNU C/C++/Fortran 3.3.4 — последняя из выпущенных на момент проведения тестов стабильных версий базового компилятора Unix-систем, поддерживает компиляцию как 32-битных, так и 64-битных приложений.

Компилятор Intel C/C++/Fortran 8.1 — последняя на момент проведения тестов версия свободно распространяемого для некоммерческого использования компилятора фирмы Intel, поддерживает компиляцию как 32-битных, так и 64-битных приложений.

Компилятор PathScale — коммерческий компилятор, ориентированный на AMD64 архитектуру (Athlon64, Athlon64FX, Opteron), предоставляется 30-дневная пробная версия с полной функциональностью. Стоимость PathScale: 1495 долл. на год, 2695 долл. на два года, 3795 долл. на три года (на одного пользователя). Использовался только на платформе AMD.

1. Вычисления с плавающей точкой

Использовались следующие тесты:

Название	Описание	Язык
LINPACK	Тест состоит в решении системы линейных уравнений с помощью LU-факторизации. Результаты теста используются в Top500	Fortran
CLINPACK	Тест LINPACK, переписанный на языке C	C
NAS	Последовательная версия NAS CFD (computational fluid dynamics)	Fortran
Whetstone	Синтетический тест, ориентированный на численное программирование (с плавающей запятой). Не учитывает кэш	C
ICM	Многократное составление и решение трех различных систем линейных алгебраических уравнений итерационным методом (последовательная версия задачи о вязком теплопроводном газе). Четыре версии с различными размерами задачи	C

Итоги тестирования

Компилятор GNU. Во всех тестах, кроме Whetstone и NAS, AMD64 показал бóльшую, чем P4, производительность (AMD64 был быстрее в 1.13–1.61 раза).

Тест NAS не удалось корректно откомпилировать.

Тест Whetstone пройден P4 быстрее, чем AMD64, в три раза. Не исключена возможность некорректной работы теста.

Компилятор Intel. Во всех тестах на языке C AMD64 показал бóльшую производительность, чем P4 (AMD64 был быстрее в 1.24–1.63 раза). В обоих тестах на языке Fortran,

напротив, P4 был быстрее в 1.5–1.73 раза. В целом на платформе AMD64 Intel незначительно проигрывает GNU в тестах на C и существенно выигрывает в тестах на языке Fortran. На платформе P4, напротив, Intel выигрывает во всех тестах, кроме ICM.

Компилятор PathScale показал наилучшие результаты во всех тестах на платформе AMD, кроме теста ICM при малых размерах задачи и теста CLINPACK, где его результат — средний между GNU и Intel.

2. Целочисленные вычисления

Использовались следующие тесты:

Название	Описание	Язык
Dhrystone	Тест целочисленной арифметики, показателен в системном программировании. Не учитывает производительность кэш-памяти	C
Heapsort	Целочисленная программа, сортирующая 2MB-массив из целых чисел	C

Итоги тестирования

Компилятор GNU. В обоих тестах AMD64 показал бóльшую производительность, чем P4 (AMD64 был быстрее в 1.15–1.9 раза).

Компилятор Intel. В обоих тестах AMD64 показал бóльшую производительность, чем P4 (AMD64 был быстрее в 1.2–1.04 раз). Однако в тесте Heapsort разброс между наихудшим и наилучшим временем для P4 был существенно меньше.

По обоим тестам на платформе AMD64 наилучшие результаты показал *PathScale* (в 1.15–1.18 раза быстрее ближайшего), Intel был вторым в тесте Dhrystone, а GNU — в тесте Heapsort. На платформе P4 ситуация та же, но без PathScale: Intel лучший в тесте Dhrystone, GNU — в тесте Heapsort.

Сводная таблица победителей тестов

Язык	Название	Лучший компилятор и платформа	
		с учетом компилятора PathScale	без учета компилятора PathScale
C	ICM	Path/AMD64	GNU/AMD64
C	Clinpack	GNU/AMD64	GNU/AMD64
F	linpack	Intel/P4	Intel/P4
F	NAS	Intel/P4	Intel/P4
C	Whetstone	Path/AMD64	Intel/AMD64
C	Dhrystone	Path/AMD64	Intel/AMD64
C	Heapsort	Path/AMD64	GNU/AMD64

3. Интегральный тест

Используемый тест — lmbench. Пакет тестов для оценки различных аспектов UNIX-систем включает тесты скорости работы с памятью, файлами, каналами и сетью (TCP/IP), тесты стандартных системных операций (переключение контекста, обработка сигнала, создание файла и т. п.).

Тестирование показало превосходство системы AMD64 на комплексных операциях: при работе с ядром операционной системы, в системных операциях, работе с ядром сетевой

подсистемы в 1.2–1.5 раза. Примерно такое же соотношение наблюдалось визуально при одновременном запуске компиляции крупных системных программых пакетов.

Скорость работы с жестким диском приблизительно одинаковая, результаты тестов памяти противоречивы, поэтому победителя выявить сложно.

4. Тест сетевой производительности

Для тестирования использовался пакет NetPerf 2.4. Узлы соединены через гигабитный коммутатор HP ProCurve 5148. Для раздельного тестирования использовался дополнительный компьютер с гигабитной сетевой картой.

Тест показал примерное равенство производительности обеих платформ.

Произведенный анализ возможных решений по организации обмена между вычислительными узлами показал целесообразность использования технологии Gigabit Ethernet в связи с небольшой стоимостью и удовлетворительной производительностью. Произведен сравнительный анализ характеристик имеющихся на рынке коммутаторов, выбраны и закуплены два коммутатора для организации сети управления и поддержки сетевой файловой системы, а также сети обмена данными между вычислительными модулями.

Выполненное комплексное сравнительное тестирование двух узлов различной архитектуры (Pentium-4 и AMD-64) с использованием трех различных компиляторов показало превосходство системы AMD-64 на комплексных операциях: при работе с ядром операционной системы, с ядром сетевой подсистемы, в системных операциях в 1.2–1.5 раза. На основе данных тестирования сделан выбор в пользу архитектуры AMD-64 в силу более высокой вычислительной производительности.

Рис. 3. МВС-1000/25 Института вычислительного моделирования СО РАН.

По состоянию на конец 2005 года созданная многопроцессорная вычислительная система МВС-1000/25 (рис. 3) имела пиковую производительность около 105.6 млрд операций в секунду, что в 7 раз превосходило имеющуюся систему МВС-1000/16 (14 млрд операций в секунду) и в 2 раза большую оперативную память и разрядность каждого вычислительного узла. По данным www.supercomputers.ru (в редакции за сентябрь 2005 года) система МВС-1000/25 была в числе 50-ти мощных суперкомпьютеров СНГ. Система находится в постоянной эксплуатации и ежегодно модернизируется. По состоянию на конец 2006 года производительность суперкомпьютера увеличена более чем в 2 раза (до 220.4 млрд оп/с).

Список литературы

[1] Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы. СПб.: Питер, 2001.

[2] Лацис А. Как построить и использовать суперкомпьютер. М.: Бестселлер, 2003.

Поступила в редакцию 26 сентября 2006 г.