

УДК 681.322

И.А. Жуков, А.В. Гуменюк

ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ КЛАСТЕРНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ В АВИАЦИИ

Рассмотрены проблемы применения высокопроизводительных вычислительных систем в авиации, принципы построения кластерных вычислительных систем, примеры их использования и перспективы развития.

Для решения важных трудоемких задач, стоящих сегодня перед такими отраслями, как машиностроение, энергетика, аэрокосмическая отрасль, метеорология, геология, экология, астрономия и другими, требуются вычислительные системы (ВС) с высоким быстродействием и надежностью функционирования.

В области авиации возникают задачи, для решения которых указанные свойства ВС особенно важны, поскольку эти задачи связаны с вопросами безопасности полетов. Примером такой задачи может служить система глобального спутникового позиционирования, контролирующая одновременно сотни воздушных судов. Также в авиации существует множество задач, не связанных напрямую с полетами воздушных судов, но требующих значительных вычислительных мощностей, например, моделирование полета или некоторых физических процессов.

Основными проблемами построения ВС для критически важных приложений, связанных с обработкой транзакций, управлением базами данных и обслуживанием телекоммуникаций, являются обеспечение высокой производительности и продолжительного функционирования систем [1]. Наиболее эффективный способ достижения заданного уровня производительности - применение параллельных масштабируемых архитектур.

Массивно-параллельная система (МРП) состоит из однородных вычислительных узлов, включающих: один или несколько центральных процессоров (обычно RISC), локальную память, коммуникационный процессор или сетевой адаптер, иногда - жесткие диски и (или) другие устройства ввода-вывода [2]. Узлы связаны через некоторую коммуникационную среду (высокоскоростная сеть, коммутатор и т.п.). Общее число процессоров в реальных системах достигает нескольких тысяч. Используется в таких вычислительных системах: IBM RS/6000 SP2, Intel PARAGON/ASCI Red, SGI/CRAY T3E, транспьютерные системы Parsytec. Вычислительные системы с такой архитектурой обладают высоким быстродействием, однако высокая цена делает их недоступными для широкого применения.

Наряду с этим в последние годы значительно выросла производительность персональных компьютеров. Одновременно стала приобретать все большую популярность операционная система (ОС) Linux - бесплатно распространяемая версия UNIX. В настоящее время в кластерных ВС используются ОС Tru64 UNIX, Linux и Windows NT.

Возникшая ситуация привела к созданию и активному использованию параллельных ВС (кластеров) из общедоступных компьютеров, недорогих Ethernet-сетей и установленных на эти компьютеры Linux. Оказалось, что на многих классах задач и при достаточном числе

узлов такие системы дают производительность, сравнимую с производительностью суперкомпьютеров.

Первой концепцию кластерной системы анонсировала компания DEC [2] (VAX-кластер), определив ее как группу объединенных между собой вычислительных машин, представляющих собой единый узел обработки информации. По существу VAX-кластер представляет собой слабосвязанную многомашинную систему с общей внешней памятью, обеспечивающую единый механизм управления и администрирования. При этом VAX-кластеры предлагают проверенный набор решений, который устанавливает критерии для оценки подобных систем. VAX-кластер обладает следующими свойствами.

Разделение ресурсов. Все компьютеры в кластере могут обращаться к отдельным файлам данных как к локальным.

Высокая готовность. Если происходит отказ одного из компьютеров, задания его пользователей автоматически могут быть перенесены на другой компьютер кластера. Если в системе имеется несколько контроллеров внешних накопителей и один из них отказывает, другие контроллеры автоматически подхватывают его работу.

Высокая пропускная способность. Ряд прикладных систем могут пользоваться возможностью параллельного выполнения заданий на нескольких компьютерах кластера.

Удобство обслуживания системы. Общие базы данных могут обслуживаться с единственного места. Прикладные программы могут устанавливаться только однажды на общих дисках кластера и разделяться между всеми компьютерами кластера.

Расширяемость. Увеличение вычислительной мощности кластера достигается подключением к нему дополнительных VAX-компьютеров. Дополнительные накопители на магнитных дисках становятся доступными для всех компьютеров, входящих в кластер.

В настоящее время кластерные ВС создаются на основе ОС UNIX. В современных кластерных системах набор рабочих станций (или даже ПК) общего назначения используется в качестве недорогого варианта массивно-параллельного компьютера. Для связи узлов используется одна из стандартных сетевых технологий (Fast/Gigabit Ethernet, Myrinet) на базе шинной архитектуры или коммутатора. При объединении в кластер компьютеров разной мощности или разной архитектуры говорят о гетерогенных (неоднородных) кластерах. Узлы кластера могут одновременно использоваться в качестве пользовательских рабочих станций. В случае, когда это не нужно, узлы могут быть существенно облегчены и (или) установлены в стойку.

Рассмотрим некоторые примеры кластерных ВС и кластерных технологий, появившихся в последнее время.

21 июля 1999 года корпорация IBM объявила, что с использованием массивно-параллельной системы RS/6000 SP ей удалось установить новый рекорд в скорости сортировки сверхбольших объемов данных. Суперкомпьютер, установленный в Teraflex Integration Center (Poughkeepsie, шт. Нью-Йорк), справился с задачей сортировки 1 Тб данных за 17 мин. Предыдущий рекорд, равный 50 мин, был установлен в Sandia National Laboratories на кластере из 72 двухпроцессорных серверов от Compaq. В данном эксперименте были задействованы 488 SP-узлов на базе процессоров PowerPC 604e/332 МГц, из которых 432 были сконфигурированы как вычислительные узлы и еще 56 – как узлы ввода-вывода. Ключевым моментом в достижении данного успеха специалисты IBM считают ис-

пользование параллельной файловой системы (GPFS), которая позволяет всем SP-узлам иметь быстрый доступ ко всем подключенным к системе дисковым массивам [2].

3 августа 1999 года Корнельский университет (Итака, шт. Нью-Йорк) объявил об установке в Cornell Theory Center (CTC) нового 256-процессорного вычислительного кластера. Данная система состоит из 64 SMP-серверов PowerEdge от DELL, работающих под управлением Windows NT. Каждый сервер включает четыре процессора Pentium III Xeon/500 МГц с 2 Мб кэш-памяти, 4 Гб оперативной и 54 Гб дисковой памяти. Для объединения узлов между собой используется сетевая технология cLAN от Giganet. Кластер в CTC, получивший название "AC³ Velocity", будет служить в качестве вычислительной платформы для ученых CTC и исследовательского консорциума AC³ (Advanced Cluster Computing Consortium) [2].

9 августа 1999 года компания MPI Software Technology, Inc. (MSTI) объявила о своем соглашении с компанией Giganet, согласно которому MSTI будет распространять ее сетевые адаптеры и коммутаторы семейства cLAN для серверов на базе Linux и Windows NT. Со своей стороны MSTI предлагает программный продукт MPI/Pro – высокопроизводительную реализацию MPI для кластеров на базе cLAN. Объединение этих продуктов предлагается пользователям в составе кластерной технологии GigaCluster. Заявлены следующие параметры производительности Gigacluster, т.е. MPI/Pro с использованием cLAN: пропускная способность - от 80 до 100 Мб/с и более (пиковая производительность cLAN составляет порядка 150Мб/с), латентность (задержка сообщений) – порядка 30 мкс. MPI/Pro обеспечивает поддержку безопасной многопоточности (thread-safety), а также сравнительно небольшую нагрузку на центральный процессор. Технология cLAN является аппаратной реализацией стандартизированной архитектуры Virtual Interface, разработанной совместно корпорациями Intel, Compaq и Microsoft. Примерная стоимость сетевого адаптера cLAN1000-LNX составляет 795 долл. США, а стоимость 8-портового коммутатора cLAN5000-LNX - 6250 долл. США. Для сравнения, стоимость адаптера Myrinet M2F-PCI32C для 32-битной шины PCI - 1095 долл. США, а соответствующего 8-портового коммутатора M2F-SW8 – 2400 долл. США [2].

В 1999 году специалисты Sandia создали крупномасштабную массивно-параллельную вычислительную систему на основе кластерной технологии, пользуясь общедоступными компонентами и сетевыми технологиями. Корпорация Compaq Computer установила под этот проект почти 1400 рабочих станций с процессорами Alpha различной мощности, 16 серверов DS20 и четыре сервера AlphaServer 1200. Общий объем дисковой памяти составляет около 3 Тб. Суммарная пиковая производительность всех процессоров составляет 1.1 TFLOP/c, что позволяет надеяться на получение производительность по LINPACK порядка 600 GFLOP/c. При запуске теста LIPACK на 350 узлах кластера была получена производительность в 152 GFLOP/c. В качестве ОС для всех узлов кластера выбрана ОС Linux (только серверы AlphaServer 1200 работают под Tru64 UNIX в силу отсутствия необходимых сетевых драйверов). Базовым строительным блоком кластера Splant является так называемый SU (Scalable unit, "масштабируемый модуль"). SU состоит из двух стоек с вычислительными узлами (восемь рабочих станций XP1000 в стойке) и одного "диагностического" узла. В каждой стойке также выделен один "сервисный узел", который имеет дополнительный Ethernet-адаптер, подключенный к внешней сети; с этих узлов пользователи в пакетном режиме за-

пускают на выполнение свои задачи. В качестве основной сетевой технологии выбрана коммутируемая сеть Myrinet (1.28 Гбит/с). Используется один 16-портовый коммутатор Myrinet на каждый SU. Дополнительная сеть на базе Fast Ethernet предназначена для задач администрирования и диагностики проблем кластера. Вычислительные узлы кластера не имеют жестких дисков и загружаются через Ethernet [2].

12 августа 1999 года корпорация SGI объявила, что подписано соглашение об установке в суперкомпьютерном центре Ohio (OSC) 128-процессорного Linux-кластера, который будет первой для SGI высокопроизводительной системой класса Beowulf, а также самой большой кластерной системой, установленной в штате Ohio. Кластер, устанавливаемый в OSC, будет состоять из 32 недавно анонсированных четырехпроцессорных серверов SGI 1400L на базе Pentium III Xeon/500 МГц. Серверы будут работать под управлением RedHat Linux 6.0 и SGI Linux Environment. Ученые и специалисты штата Ohio получают доступ к данному кластеру через высокоскоростную сеть OARnet, организованную в рамках штата при участии OSC. В OSC уже установлены несколько суперкомпьютеров от SGI: это Cray T3E (128 процессоров Alpha 21164/300 МГц), Cray T94, два сервера Origin2000. Используются рабочие станции визуализации Silicon Graphics 320 и 540 [2].

12 июля 1999 года компания Dell Computer анонсировала поддержку новой масштабируемой кластерной архитектуры - SEC (Scalable Enterprise Computing) и предложение ряда программных и аппаратных продуктов для обеспечения кластеризации. В данном случае под кластерной системой понимается объединение нескольких взаимодействующих серверов с целью повышения общей надежности системы, а также для одновременной поддержки большего числа пользователей, чем это возможно с любым одиночным сервером. Ряд вновь анонсированных продуктов поддержки кластеризации включает: программное обеспечение (ПО) администрирования кластеров Dell OpenManage Cluster Assistant with ClusterX, высокоскоростную коммуникационную технологию cLan от компании Giganet Inc., сетевые адаптеры GNN1000 и коммутаторы GNN5000 Cluster Switch. Основными строительными блоками SEC-кластеров Dell предлагает уже зарекомендовавшие себя серверы PowerEdge 4300, 4350, 6300 и 6350 и системы хранения данных PowerVault на базе шин SCSI и Fibre-Channel. Сервер PowerEdge 6300 включает до четырех процессоров Pentium III Xeon/500 МГц и до 4 Гб оперативной памяти, а PowerEdge 6350 - его модификация для монтирования в стойки (rack-mounting). RAID-системы PowerVault 650F/651F на базе шин Fibre-Channel поддерживают установку до 120 жестких дисков с общим объемом до 2.1 Тб. С помощью внутрисистемной сети (System Area Network) PowerVault SAN до четырех серверов PowerEdge подключаются к одной RAID-системе хранения данных PowerVault 650F и к ленточному накопителю PowerVault 130T. С помощью коммутаторов несколько SAN могут объединяться в одну большую кластерную систему. В настоящее время для использования с SEC-кластерами предлагается ОС Microsoft Windows NT 4.0 Enterprise Edition и ПО Microsoft Server Cluster. Поддерживается также параллельная система управления базами данных Oracle Parallel Server. В дальнейшем ожидается поддержка ОС Windows 2000 и UNIX-систем для процессоров Intel, а также 8-процессорных серверов PowerEdge и процессоров архитектуры IA-64, как только соответствующие продукты поступят на рынок [2].

В апреле 1999 года самая большая кластерная система, собранная корпорацией Compaq на базе процессоров Alpha, успешно прошла ряд тестов производительности на реальных задачах, предложенных крупнейшими национальными лабораториями США (Sandia, Los Alamos, Lawrence-Livermore), участниками суперкомпьютерной программы Министерства энергетики ASCI. Кластерная система, построенная Compaq, состоит из 128 двухпроцессорных серверов DS20 на базе Alpha EV6/500 МГц, каждый из которых включает 1 Гб оперативной памяти. Серверы работают под управлением 64-битной ОС Tru64 UNIX. Каждый узел обеспечивает пиковую производительность в 2GFLOP/c, а весь кластер - 256GFLOP/c. На тесте LIPACK кластер достиг производительность в 154.4GFLOP/c (60 % от пиковой), решая систему линейных уравнений со 120 тыс. переменных. Серверы DS20 в рамках кластера объединены с помощью 128-портового коммутатора QM-S128 (Elite) и сетевых PCI-адаптеров QM-400 (Elan), разработанных партнером Compaq, итальянской компанией Quadrics Supercomputers World (QSW). Для прикладных MPI-процессов коммуникационная среда обеспечивает пропускную способность около 200 Мб/с и латентность (задержку сигнала) порядка 5.5 мкс. Коммуникационное оборудование от QSW разгружает центральные процессоры от задач передачи данных, за счет чего MPI-процессы могут производить обмены асинхронно с вычислениями [2].

Приведенные примеры показывают перспективность применения кластерных технологий в авиации. Возможность построения кластерной системы на основе узлов разной мощности позволяет использовать для этих целей уже имеющуюся техническую базу, а также дает возможность постепенной смены узлов на более мощные в случае необходимости. Развитие сетевых технологий открывает еще более широкие возможности для повышения производительности таких систем.

Список литературы

1. Жуков И.А. Методология синтеза параллельных вычислительных систем для задач большой размерности // Проблеми інформації та керування.- К.: КМУЦА, 1997.- Вып.2.-С. 3-9.
2. Шнитман В. Современные высокопроизводительные компьютеры, информационно-аналитические материалы Центра информационных технологий, 1996.- <http://www.citmgu.ru>
3. Андреев А.Н. Новости мира высокопроизводительных вычислений. Лаборатория параллельных информационных технологий НИВЦ МГУ, 1999.- Информационно-аналитический Центр по параллельным вычислениям при поддержке Российского фонда фундаментальных исследований.- <http://parallel.srcc.msu.su/news/>

Стаття надійшла до редакції 16 жовтня 1999 року.