ОБ ОПЫТЕ ИСПОЛЬЗОВАНИЯ КЛАСТЕРОВ COWS ДЛЯ РЕСУРСОЕМКИХ РАССЧЁТОВ

С.А. Лупин, И.В. Подкопаев, Е.С. Дорошенко
Московский институт электронной техники
Россия, 124498, Москва, Зеленоград, проезд 4806, 5
E-mail: lupin@miee.ru, ilyamiee@gmail.com, len-doroshenko@mail.ru

Рассматривается возможность использования в качестве платформы для проведения высокопроизводительных вычислений компьютерных классов MS HPC Server 2008 с телекоммуникационной средой, основанной на Gigabit Ethernet. В качестве примера приведен опыт использования подобной системы в МИЭТ (ТУ).

ABOUT EXPERIENCE OF USE CLUSTERS COWS FOR RESOURSE-INTENSIVE CALCULATIONS / S.A. Loopin, I.V. Podkopaev, E.S. Doroshenko (Moscow Institute of Electronic Technology, Pas. 4806, 5, Zelenograd, Moscow, 124498, Russia.). A possibility of using computer classes governed MS HPC Server 2008 with interconnects, based on Gigabit Ethernet as a high performance computing platform is described. As an example authors describe experience of using such system in TU MIET.

Введение.

В настоящее время высокопроизводительные вычисления занимают все более важное место в промышленности, науке, образовании. В июле 2009 года на заседании Совета безопасности России президент Д.А.Медведев заявил о важности суперкомпьютеров для инновационного развития экономики страны. Современные суперкомпьютеры в подавляющем большинстве являются дорогостоящими blade-кластерами. Ведущие университеты устанавливают у себя подобные системы для обеспечения НИР, обучения студентов параллельному программированию и навыкам работы с высокопроизводительными вычислительными системами. Подобные ресурсы используются преимущественно в режиме разделения. Система управления, тем или иным образом выделяет время и процессорные мощности для каждой задачи, поступающей на выполнение. При этом каждый пользователь получает в свое распоряжение только часть системы производительностью порядка нескольких сотен GFlops. Этого оказывается вполне достаточно для большинства приложений.

Альтернативный подход к реализации ресурсоемких приложений связан с использованием гораздо более дешевых вычислительных комплексов типа кластеров рабочих станций (Cluster of WorkStations, CoWS). Их можно создавать на основе учебных компьютерных классов, без которых не обходятся ни одно высшее учебное заведение. Считалось, что подобные системы могут использоваться в режиме высокопроизводительных вычислений (High Performance Computing, HPC) только в свободное от учебных занятий время, когда на рабочих станциях никто не работает.

Исследования, проведенные в Московском институте электронной техники, позволяют опровергнуть эти утверждения. Сегодня можно использовать в режиме HPC и кластер CoWS, на узлах которого решаются локальные задачи пользователей, т.е. во время проведения лабораторных работ. Если в рабочих станциях используются многоядерные процессоры, то во время проведения занятий свободные ресурсы узлов могут быть использованы для проведения ресурсоемких расчетов, поскольку использующиеся в учебном процессе приложения, загружают только 1-2 ядра, оставляя 6-7 ядер свободными.

Исследование параметров кластера

Практические исследования предлагаемого подхода проводились на кластерной системе МИЭТ2008, установленной в ходе реализации Инновационной образовательной программы в МИЭТ. В основе кластера лежит архитектура CoWS, в 26 узлах установлены многоядерные процессоры, со следующими параметрами:

- Процессор: 2 x Intel XEON E5335, quad-core, 2.0 GHz
- Оперативная память: 4x1GB FBDIMM-667MHz
- HDD: SATA 250Gb
- Video card: GF8600GT
- ОС: MS HPC Server 2008

Все узлы объединены двумя сетями Gigabit Ethernet с помощью двух коммутаторов Hewlett-Packard. Топология кластера представлена на рис.1.

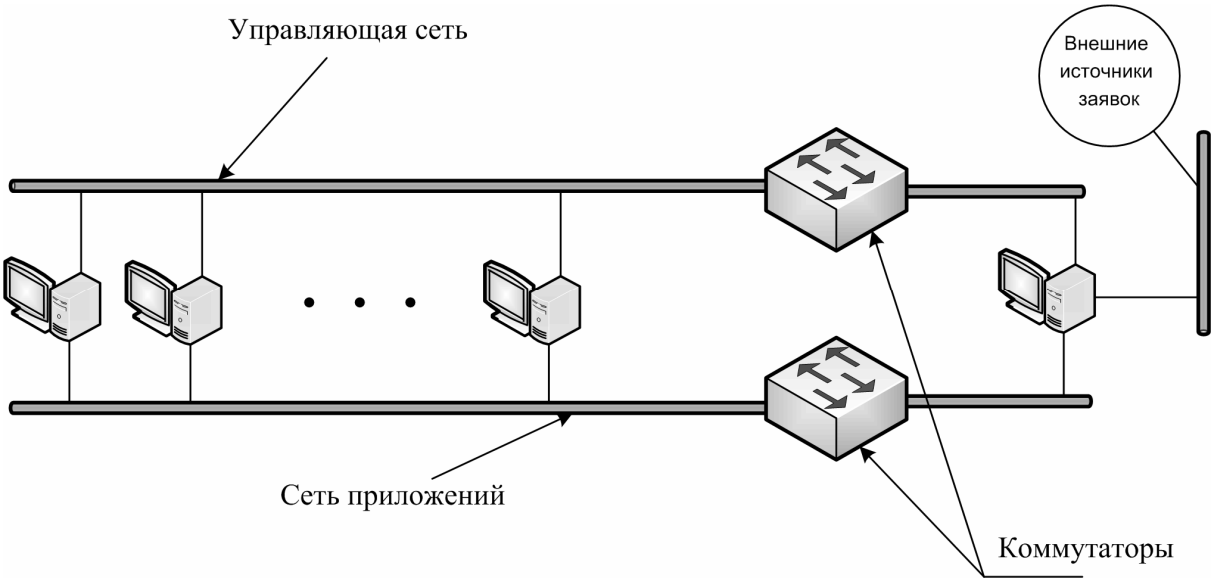

Рис. 1. Топология кластера МИЭТ2008

Пиковая производительность кластера составляет 1,6 Тфлопс, производительность на тесте Linpack достигает 871,5 Гфлопс (64,4%). Данный результат позволил кластеру занять 43-е место в списке ТОП50 суперкомпьютеров России и СНГ (редакция от 23.09.2008).

Отметим, что при увеличении числа узлов кластера, не наблюдается деградации производительности системы в целом.

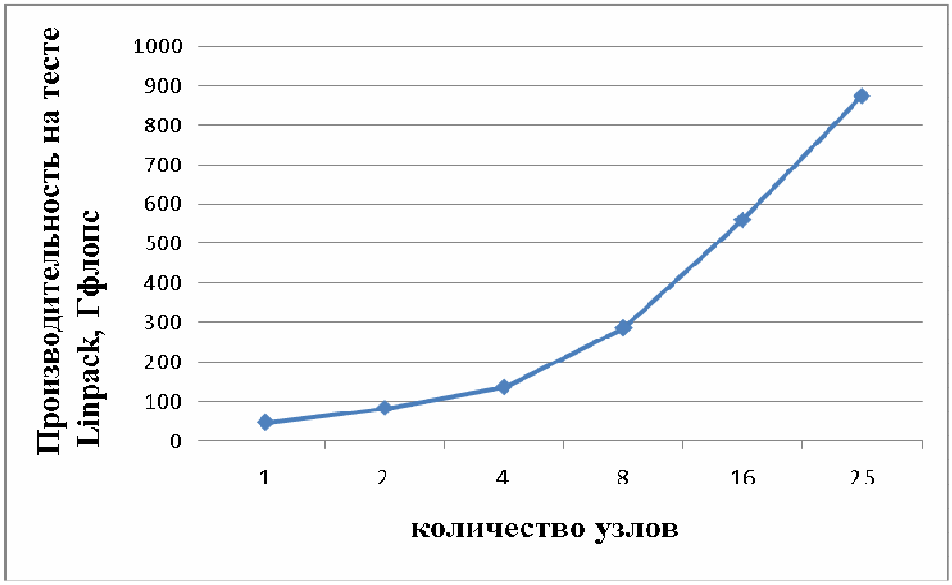

Рис. 2. Зависимость максимальной производительности кластера на тесте Linpack от числа узлов.

Для ответа на вопрос о наличии свободных вычислительных ресурсов при решении на узлах локальных задач был проведен мониторинг с помощью средств HPC Cluster Manager. Он показал, что загруженность процессоров рабочих станций не превышает 10-12%, т.е. 7 из 8 ядер остаются свободными. При этом на узлах использовалось следующее программное обеспечение: Altium Designer, Microsoft Office, Visual Studio.

Использование распределенных вычислительных ресурсов для запуска параллельных приложений возможно только при наличии в системе соответствующего менеджера, управляющего только заданиями, претендующими на свободные ресурсы узлов. Локальные задачи администрируются операционной системой рабочих станций.

Одним из возможных инструментов управления потоками задач в кластерных системах является ОС MS HPC Server 2008, важным преимуществом которой является встроенный

планировщик. Отличительной особенностью планировщика HPC Job Scheduler является наличие разнообразных стратегий управления очередью заданий:

- FIFO с использованием приоритетов;
- С вытеснением;
- Адаптивное выделение ресурсов;
- Утилизация свободных ресурсов;
- Монопольное использование многоядерного узла.

Кроме того, существует возможность разделения кластера на подкластеры с целью разделения очереди и ограничения максимально доступного каким-либо источникам заявок числа узлов или ядер. Внутри очереди для каждого подкластера также может быть применена одна из стратегий, но стратегия эта должна быть едина для всех подкластеров.

Столь широкие возможности по выбору дисциплин управления, позволяют выделить под проведение лабораторных работ, к примеру, 2 ядра из 8-ми на каждом узле, оставив для ресурсоемких задач остальные.

Второй аспект проблемы связан с межузловыми коммуникациями. Был проведен анализ влияния сетевых помех в виде широковещательного трафика на производительность кластера.

Для оценки производительности сети использовалась кроссплатформенная клиент-серверная программа Jperf — генератор TCP и UDP трафика для тестирования пропускной способности сети.

Для создания широковещательного трафика один из узлов кластера, не участвующий в вычислениях, генерировал 16000 UDP пакетов в минуту размером 1500 байт.

Проведенные исследования показали, что реальную производительность кластера в большей степени ограничивает высокая латентность Ethernet, а не пропускная способность этой среды. Тем не менее, высокие результаты на тесте Linpack, весьма интенсивно использующем пересылки между узлами, говорят о том, что коммуникационная среда класса вполне позволяет реализовывать параллельные расчеты.

Результаты выполнения теста Linpack с широковещательным трафиком:

Число узлов	Размерность задачи	Время выполнения (с)	Производительность (GFlops)
8	40000	189	225,7
8	48000	295	249,8
16	62400	359	450,0

Результаты выполнения теста Linpack без широковещательного трафика:

Число узлов	Размерность задачи	Время выполнения (с)	Производительность (GFlops)
8	40000	182	234,8
8	48000	279	264,8
16	62400	330	489,8

Из приведенных тестов можно сделать вывод о том, что при наличии широковещательной загрузки сети производительность вычислений падает менее чем на 8%.

Другими словами, производительность сети кластера позволяет проводить высокопроизводительные вычисления даже при наличии интенсивной сетевой активности со стороны пользователей.

Приведенные результаты показывают, как влияет на производительность параллельных вычислений нагрузка, создаваемая локальными приложениями пользователей. Справедливо и обратное – параллельное приложение, запущенное на узлах кластера будет оказывать влияние на локальные процессы пользователей. Необходимо оценить, насколько снизится длительность отклика пользовательских узлов при выполнении интерактивных приложений.

В ходе исследований было выяснено, что запуск на узлах параллельных и локальных приложений в режиме разделения времени приводит к существенному увеличению времени отклика для интерактивных приложений пользователей.

Однако, если для параллельных приложений использовать на каждом узле не все доступные вычислительные ядра, а, допустим, семь из восьми, то это позволяет обеспечить время реакции системы на приемлемом уровне при сохранении доступа к весьма существенным вычислительным мощностям с пиковой производительностью:

$$R_{peak} = {}^{n_A}\!/_{n_C}\, R_{cluster\ peak} = {}^{7}\!/_{8} \cdot 64 \cdot 25 = 1400 \text{ GFlops},$$

где n_A - число доступных для проведения вычислений ядер на каждом узле.

Исследования подтвердили, что при решении на узлах локальных задач остаются незадействованными сетевые и вычислительные ресурсы, которые могут быть использованы для параллельных вычислений.

Заключение

Полученные результаты служат практическим подтверждением тезиса о том, что системы CoWS под управлением ОС MS HPC Server 2008 могут быть с успехом использованы и в роли обычного компьютерного класса, и в роли высокопроизводительной вычислительной системы для решения ресурсоемких задач.

Список литературы.

1. 9-я редакция списка ТОП50 суперкомпьютеров России и СНГ, http://supercomputers.ru/?page=archive&rating=9
2. www.microsoft.com/hpc
3. Дорошенко Е.С., Лупин С.А., Подкопаев И.В. Использование высокопроизводительных вычислительных систем в университетах. Известия высших учебных заведений. ЭЛЕКТРОНИКА. № 2(76), 2009г., стр. 74-81
4. Дорошенко Е.С., Лупин С.А., Подкопаев И.В. Использование свободных вычислительных ресурсов в многоядерных кластерах CoWS. «Телекоммуникации», - М: «Наука и технологии», 2010, №2, стр.16-20