УДК 681.32

ПОВЫШЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ КЛАСТЕРОВ РАБОЧИХ СТАНЦИЙ С ИСПОЛЬЗОВАНИЕМ ВЕЕРНОГО РАСПРЕДЕЛЕНИЯ ДОПОЛНИТЕЛЬНЫХ ЗАДАНИЙ НА ПРОСТАИВАЮЩЕЕ ОБОРУДОВАНИЕ

© 2012 В. М. Довгаль[1], С. Г. Спирин[2]

[1]профессор каф. программного обеспечения
и администрирования информационных систем
e-mail: vmdovgal@yandex.ru

Курский государственный университет

[2]инженер-программист
e-mail: hemmul.seto@gmail.com

ОАО «АСТОР-ТРЕЙД», г. Курск

В статье рассмотрена концептуальная схема, которая открывает возможности использования вычислительных ресурсов простаивающего оборудования, объединенного в кластер. Такой подход позволит экономить средства при решении внутренних сложных задач или предоставлять свои ресурсы сторонним физическим или юридическим лицам.

Ключевые слова: распределенные вычисления, ресурс, производительность, простои компьютеров, загрузка кластера.

Оперативное решение некоторых задач научно-технической сферы возможно только на высокопроизводительных вычислительных мощностях, такие мощности располагают в центрах обработки данных (ЦОД) [Воеводин 2002]. Организация и обслуживание собственного ЦОД требует существенных затрат [Мельник 2009; Воеводин 2008], поэтому для решения некоторых задач пользуются услугами организаций, которые предоставляют собственные вычислительные ресурсы. Основное преимущество такого подхода — сокращение времени решения задач и использование уже имеющихся ресурсов вместо высоко затратных вложений на создание новых аппаратных средств. Между тем нехватка высокопроизводительных вычислительных ресурсов для решения объема задач, существующего на настоящее время [Мельник 2009], устанавливает их достаточно высокую цену.

В настоящее время существует GRID технология, в которой средой распределенных вычислений являются простаивающие рабочие станции [Воеводин 2008]. Эта технология демонстрирует возможность предоставления простаивающих ресурсов для сервисов по обработке данных.

Поиск простаивающих ресурсов производится среди кластеров рабочих станций организаций, основываясь на оценке отношения их реальной производительности к пиковой. На текущий момент это отношение составляет 5–20 % [Ивашко; Дубова 2005], что позволяет прийти к заключению о недостаточной загруженности (простоях) вычислительных узлов. Под понятием загруженности узла p с одним исполнительным устройством понимается отношение значения времени непосредственной работы узла за определенный промежуток времени к значению этого промежутка. Реальная производительность R кластера рабочих станций при пиковой производительности r_i рабочей станции i, где $i \in \{1,2, \ldots, I\}$, и объеме совокупности рабочих станций I вычисляется по формуле [Воеводин 2002]

$$R = \sum r_i \cdot p_i . \quad (1)$$

Цель работы заключается в изложении одной из концептуальных схем повышения производительности распределенных вычислений путем реализации предварительно разделенных на два класса заданий на множестве удаленных исполнительных устройств. К первому классу относятся задания, которые требуют высоких затрат времени, а ко второму – быстро решающиеся задания; при этом предлагается, используя асинхронную динамику завершения заданий первого класса, загружать свободные исполнительные устройства и каналы связи кластера заданиями второго класса.

Нужно отметить, что тогда, когда интервалы простоя узлов кластера вызваны интенсивной работой с памятью или ограничениями коммуникационной составляющей, которая непосредственно обеспечивает связь с каждой конкретной рабочей станцией, нет возможности использования такого ресурса [Воеводин 2009]. Тем не менее если простои возникают в результате неоптимального планирования распределения заданий в кластере при неустранимых ошибках оценки времени завершения обработки данных, то такие интервалы простоя целесообразно использовать для повышения производительности системы обработки данных, применяя веерную дополнительную загрузку исполнительных устройств кластера быстро выполняющимися заданиями второго класса.

Реальная производительность при максимальной загрузке оценивается с помощью стандартных тестов [Воеводин 2000; Пересветов 2010]. Значение отношения реальной производительности к пиковой соответствует значению загрузки P кластера рабочих станций:

$$P = \sum_{i=1}^{I} \alpha_i \cdot p_i , \ \alpha_i = r_i / \sum_{i=1}^{I} r_i , \quad (2)$$

где r_i – пиковая производительность i-той рабочей станции , p_i – значение загрузки для каждой рабочей станции [Воеводин 2002].

Организация распределенных вычислений в общем случае состоит из организации очереди заданий, способа выбора следующей задачи из очереди, алгоритма назначения на вычислительные ресурсы и менеджера ресурсов. Использование доступных ресурсов зависит в первую очередь от организации менеджера ресурсов и механизма (алгоритма) назначения.

Для выявления ограничений максимальной загрузки узлов проведем анализ архитектуры вычислительного кластера. Кластер – это объединенная в локальную вычислительную сеть группа компьютеров, работающих как целостный вычислительный ресурс [Воеводин 2000]. Коммутация каналов ЛВС производится специализированными коммутаторами, благодаря чему из каналов образуются маршруты для связи между двумя любыми узлами кластера.

В общем виде маршрут между любым источником данных и целевым узлом состоит из n каналов, где n принимает значения из 1,2 ..... N, при N – максимальной длине произвольного маршрута. В каждый момент времени любой из каналов маршрута может быть полностью загружен, что исключает возможность его использования в рассматриваемый момент времени.

Если маршрут между планировщиком и целевым узлом состоит из l каналов и значение l минимально, то вероятность U того, что этот маршрут свободен, будет максимальна:

$$U = \prod_{y=0}^{l} u_y , \quad (3)$$

где u_y – вероятность того, что канал с номером $y \in [0,l]$ маршрута будет не загружен.

Всякий узел через собственный канал соединяется с одним из коммутаторов, каждый из которых формирует группу вычислительных узлов. Группы, в свою очередь, объединяются через каналы между коммутаторами или с помощью маршрутизатора. Любой коммутатор обеспечивает доступ к каждому простаивающему узлу, с которым имеется непосредственная связь, если этот простой, как отмечалось выше, не вызван передачей данных на этот узел или работой исполнительного устройства узла с внутренней памятью.

При передаче данных объемом V_d через любой конкретный маршрут, время передачи t_d зависит от значений задержек f_i каждого канала маршрута, номера $i \in (0,1 .. l)$, и минимальной скорости передачи s_{min} для маршрута:

$$t_d = \sum_{i=1}^{l} f_i + V_d / s_{\min} . \qquad (4)$$

Из этой формулы видно, что время задержки с увеличением длины маршрута увеличивается.

Объединим в одном устройстве реализацию функции коммутатора, функции менеджера интервалов простоев исполнительных устройств и функцию планировщика заданий дополнительной нагрузки (далее брокеров). В результате представляется возможным коммутатор заместить компонентом, который включает выполнение всех перечисленных выше функций. Обозначим его как устройство ***A***. Такой подход, при наличии заданий дополнительной нагрузки в распоряжении устройства ***A*,** позволяет использовать простои группы рабочих станций, которые возникают при выполнении основных задач используемого кластера. Поэтому целесообразно вводить подкласс заданий дополнительной нагрузки для каждого устройства ***A*** в моменты простоев соответствующих маршрутов.

После проведенных преобразований информация о простаивающих узлах каждой группы локализована в соответствующих им устройствах ***A***. В общем случае минимально достижимое время выполнения любого задания может быть определено при отсутствии ограничения на использование простоев одной группы. Исходя из этого, в функции устройства ***A*** целесообразно включить функцию обмена информацией о текущих простоях между устройствами ***A*** и возможность использования доступных ресурсов, возникающих в результате простоев вычислительных узлов в других группах. Механизм использования простоев узлов в других группах и простоев собственной группы обеспечивает менеджер простоев.

По мере возникновения простоев данные о них передаются от соответствующего устройства ***A***, образующего группу, всем другим устройствам ***A*** кластера.

Для реализации трассировки заданий дополнительной нагрузки посредством замещения каждого коммутатора устройством ***A*** целесообразно решить задачу выбора маршрута передачи данных, с учетом состояния каналов, от источника данных к целевому простаивающему узлу. Эту задачу в сетях связи решает протокол OSPF (RFC 2328), и на текущий момент он удовлетворяет поставленной задаче по критериям скорости сходимости алгоритмов, достижимости узлов сети и оптимальном использовании пропускной способности каналов, поэтому передача параметров простоев основана на OSPF.

Данные о простоях узлов включают в себя номер группы, номера простаивающих вычислительных узлов и время окончания простоев (начала выполнения новых заданий в соответствии с планами реализации), пиковую производительность узлов, маршрут передачи данных, время задержки передачи данных в группу целевых узлов и пропускную способность маршрута. Относительно

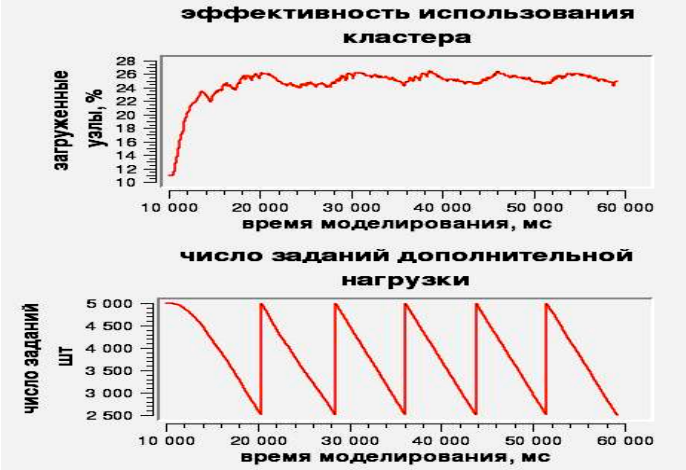
различных устройств ***A*** различаются время задержки и пропускная способность, которые корректируются с учетом параметров маршрута.

Данными для решения задачи трассировки задания являются таблица данных о простоях узлов (таблица параметров), объем данных задания. Трассировка заданий сводится к выбору целевого ресурса, который удовлетворяет заданным критериям, включая информацию об отклонении в плане выполнения заданий, что также отражается в таблице параметров.

Приведенная выше концептуальная схема основывается на предположении, что задания дополнительной нагрузки равноправно занимают простаивающие ресурсы. Кроме того, вводится упорядочение по приоритетам, на основании чего используется механизм вытеснения на конкретном ресурсе заданий дополнительной нагрузки заданиями с приоритетом. Такой механизм требует определения ресурсов, занятых заданиями с меньшим приоритетом, по отношению к любому конкретному устройству ***A***. В результате некоторые задания могут занимать ресурс, который в рассматриваемый момент использует другое задание, но с меньшим приоритетом. Это обстоятельство приводит к необходимости вводить для каждого приоритета задания свою таблицу параметров. При использовании простоев устройство ***A***, в группе которого находится целевой узел, информирует об этом все остальные устройства ***A***. С учетом упорядочения заданий целесообразно также информировать и о приоритете задания, которое загружено в соответствующий узел.

Предложенный подход был использован при моделировании работы виртуального вычислительного кластера, структура которого была сформирована с использованием детерминированных хаотических рядов [Мун 1990] из групп, то есть значения параметров кластеров формировались на основе указанных рядов. Планировщик основных заданий кластера и менеджер ресурсов размещены на выделенном узле, выбор планировщиком целевого узла происходит с учетом состояния маршрута от этого выделенного узла до целевого. Реальная загрузка кластера для основных заданий оказалась около 11% используемых в каждый момент вычислительных узлов, что определяется недостатками централизованной схемы управления трассировкой. При проведении эксперимента в качестве приоритетных задач моделировалось выполнение связанных вычислительных процессов. Число приоритетных процессов и пересылаемый объем данных выбирались таким образом, чтобы в каждый момент времени общая загрузка каналов кластера достигала 80%.

Задания дополнительной нагрузки распространяются за счет свободных маршрутов до любого устройства ***A***. При реализации дополнительной нагрузки доля узлов, используемых в каждый момент времени, возросла до 26 % (см. рис.). Такое повышение эффективности работы кластера достигнуто при использовании узлов, освободившихся от выполнения приоритетных заданий раньше запланированного времени, причем маршрут между такими узлами и планировщиками был занят.



Результаты эксперимента моделирования работы кластера

Результаты реализации приведенных в статье предложений в ходе эксперимента позволили увеличить загрузку узлов с 11% до 26 %, что позволяет обосновать вывод о целесообразности веерной загрузки простаивающих вычислительных узлов кластеров. Ориентация на увеличение реальной производительности кластеров рабочих станций при использовании заданий дополнительной веерной нагрузки открывает новые возможности технико-экономического характера за счет решения сторонних задач, поступающих от физических или юридических лиц, или экономии средств для решения внутренних задач кластера, используя собственные ресурсы, порождаемые простоями оборудования.

Библиографический список

Воеводин В. В., Воеводин Вл. В. Параллельные вычисления. СПб.: БХВ-Петербург, 2002. 608 с.

Воеводин Вл. В. Суперкомпьютеры и парадоксы неэффективности // Открытые системы. 2009. № 10.– С. 37–45

Воеводин Вл. В. Кластеры и суперкомпьютеры – близнецы или братья? // Открытые системы. 2000. № 5–6. С. 48–53

Воеводин Вл.В. Экспонента суперкомпьютерных центров // Открытые системы. 2008. № 5. С. 32–39

Воеводин Вл. В., Филамофитский М. Суперкомпьютер на выходные // Открытые системы. 200. № 5. С. 42–48

Дубова Н. GRID: время пришло? //Директор информационной службы. 2005. № 4. С. 34–37

Ивашко Е. Е. Высокопроизводительные вычисления в каждый дом [Сайт]. URL: www.ibm.com/developerworks/ru/library/l-grid/index.html (дата обращения: 8.09.2012).

Мальковский С. И., Пересветов В. В. Исследование производительности четырехпроцессорных узлов в составе вычислительного кластера // Материалы международной научно-практической конференции «Суперкомпьютеры: вычислительные и информационные технологии». Хабаровск: Изд-во Тихоокеанского гос. ун-та, 2010. С. 77–84.

Мельник О. Е. Аутсорсинг ЦОД на фоне кризиса // СК Пресс. 2009. № 2–3

Мун Ф. Хаотические колебания: Вводный курс для научных работников и инженеров. М.: Мир, 1990. 312 с.