



СУПЕРКОМПЬЮТЕР «ЛОМОНОСОВ»

**для Московского Государственного Университета
им. М.В. Ломоносова**

МГУ им. М.В. Ломоносова – крупнейший вуз Российской Федерации, один из ведущих центров отечественной науки и образования. Университет проводит множество исследований для стратегически важных отраслей российской экономики, таких как энергетика и энергосбережение, транспорт, медицина, авиация и космонавтика, нанотехнологии. Без мощных вычислительных систем современные исследования в любой из этих областей занимают долгие годы и требуют серьезных и вложений в экспериментальную базу. Моделирование позволяет не только сократить сроки и стоимость исследований, но помогает добиться более глубокого понимания реальных процессов и явлений.

В МГУ установлены несколько суперкомпьютеров. Самым мощным из них еще недавно являлся суперкомпьютер СКИФ МГУ «Чебышев» с пиковой производительностью 60 Тфлопс, разработанный и построенный компанией «Т-Платформы». Однако с момента запуска этой машины весной 2008 года стало ясно, что потребности пользователей вычислительного центра в ресурсах намного превышают возможности «Чебышева». К 2009 году расширение центра стало насущной необходимостью.

ЗАДАЧА

«Мы должны всячески стимулировать востребованность суперкомпьютеров – не потому, что это модная тема, а просто потому, что по-другому не создать конкурентоспособную продукцию, которую будут воспринимать правильным образом наши потенциальные покупатели....».



Дмитрий Медведев
Совещание с членами Совета
Безопасности по вопросам
создания и применения
суперкомпьютеров.
28 июля 2009

Руководством МГУ им. М.В. Ломоносова было принято решение о необходимости приобретения нового, более мощного суперкомпьютерного комплекса, который обеспечил бы высокопроизводительными ресурсами большее число научных коллективов и позволил проводить больше исследований. Кроме того, суперкомпьютер большей мощности способен вывести исследования на качественно новый уровень: требования к вычислительным ресурсам растут пропорционально разрешению компьютерных моделей и числу учитываемых ими параметров.

Ключевые требования МГУ к новому суперкомпьютерному решению включали наилучшее соотношение цены и производительности, высокую масштабируемость большинства приложений, высокую плотность установки в условиях ограниченных площадей, расширяемость и отказоустойчивость. Вычислительная мощность новой установки должна была достичь не менее 400 Тфлопс в рамках заданного бюджета, при этом инженерная инфраструктура электропитания и охлаждения должна была обеспечивать возможность дальнейшего увеличения производительности компьютера. Архитектура вычислителя должна была стать максимально универсальной и гибкой, чтобы обеспечить высокую реальную производительность широкого спектра приложений. Открытый конкурс на поставку суперкомпьютерного комплекса для МГУ выиграла компания «Т-Платформы».

РЕШЕНИЕ

Суперкомпьютерный комплекс, поставленный компанией «Т-Платформы» для МГУ им. М.В. Ломоносова в 2009 году обладал пиковой производительностью 420 Тфлопс, в реальная производительность системы на тесте Linpack составила 350 Тфлопс. Таким образом, эффективность суперкомпьютера, то есть соотношение реальной и пиковой производительности, составляла 83%. В 2009 году суперкомпьютер Ломоносов стал двенадцатым в мире по мощности: Россия никогда не поднималась так высоко в мировых рейтингах. Подобные машины сегодня есть только в США, Германии и Китае.

В 2010 году, в рамках планового расширения системы, пиковая производительность суперкомпьютера была увеличена до 510 Тфлопс.



Суперкомпьютер «Ломоносов» относится к системам высшего диапазона производительности: установка содержит 5 130 вычислительных узлов и 44 тысяч процессорных ядер.

Впервые в истории столь мощную вычислительную систему удалось разместить на площади всего 252 квадратных метра: по вычислительной плотности «Ломоносов» сегодня не имеет себе равных в мире. При этом вычислитель потребляет не более 2 МВт электроэнергии. Однако помимо высокой плотности и оптимального энергопотребления, вычислитель такого масштаба должен обеспечивать высокую скорость решения реальных прикладных задач: архитектура системы должна позволять приложениям использовать вычислительную мощь процессоров максимально эффективно. Справиться со всеми этими задачами компании «Т-Платформы» позволили собственные разработки в области аппаратного и программного обеспечения. В суперкомпьютере используется 3 вида вычислительных узлов и процессоры с различной архитектурой, что позволяет получать высокую производительность максимально широкого спектра приложений. Это первый гибридный суперкомпьютер такого масштаба в России и Восточной Европе.

ВЫЧИСЛИТЕЛЬНОЕ ЯДРО

В качестве основных узлов, обеспечивающих свыше 90% производительности системы, используется инновационная blade-платформа, получившая название TB2. По вычислительной плотности на квадратный метр занимаемой площади – 30 Тфлопс/м² – эта система превосходит все мировые аналоги. TB2 на базе процессоров Intel® Xeon X5670 обеспечивает производительность 27 Тфлопс в стандартной стойке высотой 42U. В суперкомпьютере также задействованы blade-системы TB1.1 с увеличенным объемом оперативной памяти и локальной дисковой памятью для выполнения специфических задач, особенно требовательных к этим параметрам системы. Третий тип узлов – платформы на базе многоядерного процессора PowerXCell 8i, используемые в качестве мощных ускорителей для ряда задач. Все три типа вычислительных узлов являются разработками компании «Т-Платформы». Все вычислительные узлы установки, а также систему хранения данных, связывает высокоскоростная коммуникационная сеть QDR InfiniBand с пропускной способностью до 40Гб/сек. В качестве дополнительных сетей используются 10G Ethernet и Gigabit Ethernet, а также выделенные сети поддержки коллективных коммуникаций разработки «Т-Платформы».

РОССИЙСКИЕ СУПЕРКОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ

Платформа TB2 была сконструирована инженерами «Т-Платформы» «с нуля» - все платы и механические компоненты являются собственными разработками компании. TB2 выходит за рамки традиционных возможностей blade-систем. Новая платформа позволяет создавать суперкомпьютеры высшего диапазона производительности, не уступающие специализированным нестандартным вычислителям с массивно-параллельной архитектурой (MPP). Такие системы обеспечивают лучшую масштабируемость приложений по сравнению с традиционными кластерами и число процессоров в них может достигать десятков тысяч.

Суперкомпьютеры на базе TB2 обеспечивают высокую эффективность реальных приложений за счет принципиально новых технологий, таких как выделенные сети барьерной синхронизации и глобальных прерываний. Они позволяют за минимальное время синхронизировать вычислительные процессы на отдельных узлах и ускорить обмен данными между процессорами.

СИСТЕМНАЯ ПЛАТА ВЫЧИСЛИТЕЛЬНОГО МОДУЛЯ



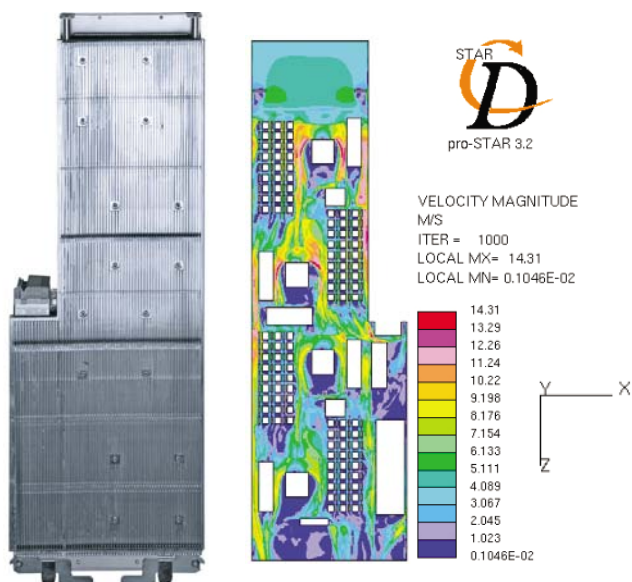
Основой решения TB2 является вычислительный модуль на основе системной платы разработки «Т-Платформы». Уникальная вычислительная плотность решения достигается прежде всего за счет оригинального дизайна 14-слойной системной платы, на которой расположены 4 процессора Intel® Xeon® серии 55xx или 56xx, 4 трехканальных модуля памяти DDR3 разработки «Т-Платформы» и интегрированные контроллеры системной сети QDR InfiniBand.

КОММУТАТОР СИСТЕМОЙ СЕТИ QDR INFINIBAND

Дизайн коммутаторов системной сети в системах TB2 является оригинальной разработкой «Т-Платформы» на основе эталонного дизайна Mellanox InfiniScale IV. Два коммутатора, интегрированных в задней секции шасси, обеспечивают неблокируемую пропускную способность системной сети 1,6 Тбит/сек для бесконфликтной передачи данных в инсталляциях, включающих десятки тысяч узлов. Два коммутатора имеют 32 внутренних порта для подключения всех вычислительных узлов и 40 внешних портов, 6 из которых используются для подключения систем хранения данных по сети InfiniBand или создания гетерогенных вычислительных систем – например, с узлами на базе PowerxCell 8i или GPGPU NVidia.

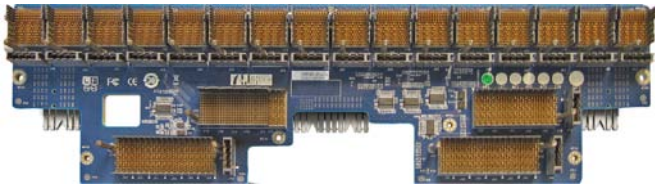


РАДИАТОР СИСТЕМОЙ ПЛАТЫ



Высокая степень интеграции компонентов на системной плате, выделяющей порядка 570 Вт тепла, требует эффективного охлаждения. Оптимальный вариант дизайна радиатора, отводящего это тепло, был найден благодаря моделированию на суперкомпьютере мощностью 10 Тфлопс. В ходе двухмесячного компьютерного анализа из 6 вариантов был выбран один с наилучшим соотношением веса и энергоэффективности. Это композитный алюминиевый радиатор с медными вставками, полностью закрывающий системную плату и обеспечивающий эффективное воздушное охлаждение блейд-системы. Выбранная конструкция позволила уменьшить вес шасси до 153 кг и снизить таким образом нагрузку на фальшполы в суперкомпьютерном центре, что позволило оптимизировать затраты на подготовку инфраструктуры зала.

ОБЪЕДИНИТЕЛЬНАЯ ПЛАТА



Эта плата объединяет все подсистемы шасси: вычислительные модули, систему питания и охлаждения, сетевые интерфейсы и систему управления. Компактный дизайн объединительной платы оставляет в шасси достаточно пространства для эффективного воздушного охлаждения вычислительных модулей вентиляторами, расположенными в передней части шасси. 24-слойная плата обеспечивает горячую замену вычислительных модулей, а также подачу питания и информационных сигналов ко всем подсистемам.

МОДУЛЬ УПРАВЛЕНИЯ

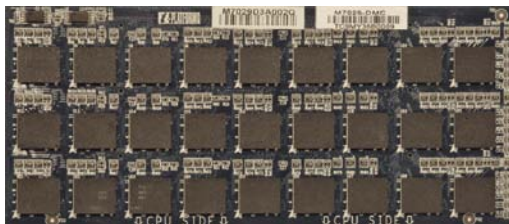
Модуль управления включает 4 функциональных блока, обеспечивающих мониторинг и управление системы, интеграцию управляющей и вспомогательной сетей стандарта 10GbE/ GbE/Ethernet, интеграцию специализированных сетей барьерной синхронизации и глобальных прерываний, а также сети внешней частотной синхронизации вычислительных узлов.

Специализированные сети снижают задержки, возникающие при синхронизации выполнения параллельных операций на крупных инсталляциях. Управление этими сетями осуществляется специально запрограммированной микросхемой FPGA.



МОДУЛЬ ПАМЯТИ

В отличие от стандартных DIMM-модулей, оригинальные высокоинтегрированные модули памяти разработки «Т-Платформы» позволяют обеспечить очень плотное размещение вычислительных модулей в блейд-шасси. Каждая микросхема интегрирует функционал трех DIMM-модулей и горизонтально размещается на системной плате, позволяя экономить место. Модули памяти легко снять и заменить, что облегчает и удешевляет обслуживание по сравнению с другими компактными решениями, где память интегрирована на системную плату.



ШАССИ ТВ2

Шасси ТВ2 на сегодняшний день обеспечивает максимальную плотность вычислительной мощности на квадратный метр площади дата-центра. Воздушное охлаждение упрощает обслуживание и улучшает ремонтопригодность системы. Дизайн шасси обеспечивает высокую компактность, небольшой вес, простоту инсталляции в стандартные шкафы 19" и возможность развертывания в ЦОДах с традиционной инфраструктурой. Такое решение позволяет снизить капитальные затраты при реализации комплексных суперкомпьютерных центров на базе ТВ2.



СИСТЕМА ХРАНЕНИЯ ДАННЫХ

Суперкомпьютер «Ломоносов» использует многоуровневую систему хранения данных: высокоскоростное дисковое хранилище T-Platforms SAN 7998 объемом 500 Тбайт под управлением параллельной файловой системы Lustre, скоростное дисковое хранилище повышенной надежности объемом 300 Тбайт, а также роботизированную ленточную библиотеку ёмкостью 1Пбайт.

СПЕЦИАЛИЗИРОВАННОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Задачу ускорения реальных приложений в суперкомпьютере «Ломоносов» решают не только аппаратные технологии, но и новое системное ПО. Суперкомпьютер работает под управлением пакета Clustrx разработки компании T-Massive Computing, входящей в состав холдинга «Т-Платформы». Clustrx ОС устраняет критические ограничения масштабируемости, присущие современным операционным системам, обеспечивая более эффективное использование ресурсов крупных инсталляций. Пакет Clustrx содержит все необходимые компоненты для управления суперкомпьютером и организации удобного доступа пользователей к системе. Clustrx обеспечивает ежесекундный мониторинг до 150 метрик на каждом вычислительном узле, использует технологии агрессивного энергосбережения и автоматического реагирования на критические ситуации.

ОТКАЗУСТОЙЧИВОСТЬ

Добиться требуемого уровня отказоустойчивости позволило резервирование всех критических подсистем и компонентов суперкомпьютерного комплекса – от вентиляторов и блоков питания в вычислительных узлах до систем электропитания и охлаждения. Высокую надежность blade-систем обеспечивает отсутствие кабельных соединений и жестких дисков внутри шасси, а также целый ряд конструктивных решений, таких как специально разработанные разъемы для модулей памяти.

«Суперкомпьютерами такой мощности, согласно списку ТОП500, сейчас обладают всего 4 страны, среди которых теперь числится и Россия. Это громадный шаг вперед, который означает, что отечественные разработки находятся на высочайшем уровне, — отметил ректор МГУ имени М.В. Ломоносова академик В.А. Садовничий. — Подобных суперкомпьютеров сейчас нет не только в России, но и во всей Восточной Европе. С появлением этого вычислительного комплекса МГУ вошел в число крупнейших суперкомпьютерных центров мира»

ИНЖЕНЕРНАЯ ИНФРАСТРУКТУРА

«Ломоносов» потребляет около 2 МВт электроэнергии и надёжно защищён от возможных сбоев в энергосистеме. В случае аварии в источники бесперебойного питания обеспечат полноценную работу компьютера на время, гарантирующее корректное завершение всех программ и сохранение данных. КПД источников бесперебойного питания составляет около 97%, в то время как в системах подобного масштаба эта характеристика обычно составляет не более 92%.



Из-за высокой вычислительной плотности от одного шкафа с оборудованием необходимо отводить до 65кВт тепла – больше, чем в каком-либо другом вычислительном центре мира. Для этого инженерам компании «Т-Платформы» пришлось разработать особую систему теплоотвода. Суперкомпьютер охлаждается и за счет сложной инфраструктуры, которая занимает в общей сложности около 800 квадратных метров внутри и снаружи здания. Для охлаждения системы используется режим “free-cooling”: при низких температурах уличного воздуха водоохлаждающие агрегаты частично или полностью отключают работу компрессоров, что позволяет значительно экономить электроэнергию – в российских условиях система охлаждения около полугодя почти не потребляет электричество.

Серьёзное внимание разработчики уделили системе противопожарной безопасности. Здесь тоже не обошлось без инноваций – автоматическая система пожаротушения суперкомпьютера использует специальный газ. В случае возгорания, он всего за полсекунды заполнит помещение и ликвидирует пожар, не повредив дорогое оборудование. Ещё одна особенность нового газа в том, что он не снижает концентрацию кислорода, а потому практически безопасен для человека.

Суперкомпьютерные комплексы МГУ используются для реализации фундаментальных и прикладных проектов, в которых ученые Московского университета сотрудничают с множеством научных коллективов по всей стране и за рубежом. Например, одной из важнейших проблем современной науки является оценка будущих изменений климата и их последствий для окружающей среды. Наиболее перспективным средством получения таких оценок являются математические модели климатической системы, которые включают описание широкого круга физических, химических и биологических процессов, происходящих в атмосфере, гидросфере, криосфере и биосфере. По пространственному масштабу климатические модели разбивают на три класса: глобальные, региональные и локальные. В международном проекте изучения глобальных моделей Россия представлена проектом ИВЦ РАН. С моделью этого проекта проведены многочисленные расчеты изменений климата в 21 веке – например, роста среднегодовой температуры в том или ином регионе - при различных сценариях выбросов парниковых газов в атмосферу. Региональные климатические модели важны для планирования социально-экономического развития регионов. В Научно-исследовательском вычислительном центре МГУ развивается региональная атмосферная модель, которая позволяет рассчитывать динамику атмосферы вплоть до масштабов отдельных мощных кучевых облаков. С помощью локальных моделей климата изучают ветровые потоки, которые играют существенную роль в жизни современного города. Сильный порывистый ветер, образующийся между высотными зданиями, приводит к многочисленным неприятным последствиям, которых можно избежать, если спрогнозировать ветровые потоки на этапе застройки нового района. Такой прогноз можно осуществить на основе вихреразрешающей модели, развиваемой в ИВМ РАН и НИВЦ МГУ.

При разработке нанотехнологий компьютерное моделирование способно сыграть ведущую роль. Далеко не все процессы, происходящие на атомарном уровне, возможно изучить через физический эксперимент: все чаще и чаще для интерпретации данных экспериментов и более глубокого понимания причин процессов, происходящих в наносфере, ученые используют моделирование. В результате ученые получают возможность конструирования искусственных соединений с заданными свойствами и функциями. Например, с помощью методов молекулярной динамики ученые МГУ изучают новый перспективный класс соединений, обладающих полупроводниковыми свойствами и в тоже время способных к самоорганизации в так называемые «нанопровода», или «нанопровода». Атомистическое моделирование позволяет понять структуру, свойства и принципы самоорганизации этих нанопроводов. Благодаря уникальной способности этих соединений к самоорганизации, создание гибридных биосинтетических структур может вывести процесс создания нанoeлектронных устройств на принципиально новый уровень.

Начальный этап разработки нового лекарства составляет поиск молекул-ингибиторов для белков-мишеней. Существенно сократить затраты времени и средств на этапе поиска ингибиторов можно с помощью суперкомпьютерных технологий и методов компьютерного молекулярного моделирования. Это позволяет предсказывать новые органические молекулы, которые будут наиболее эффективно и избирательно связываться с активными центрами исследуемых белков. Разработке нового патентночистого синтетического ингибитора тромбина – причины смертельно опасного тромбоза - была посвящена совместная работа МГУ и Гематологического научного центра РАМН, а сам подход получил развитие в процессе поиска путей создания ряда других лекарственных препаратов. Моделирование позволило существенно сократить затраты на разработку лекарства: из 6000 перспективных соединений выбрали только самые лучшие, и вместо синтеза и измерений активности из всех 6000 соединений при использовании обычного метода экспериментального перебора синтезировали всего 20, чтобы найти новый ингибитор. Таким образом, применение суперкомпьютеров для разработки новых лекарств ускоряет стадию разработки в среднем до двух лет вместо восьми, удешевляет разработку в 10-100 раз, создает основу для регулярной конвейерной разработки новых лекарств и помогает снизить их стоимость.

Основные технические характеристики суперкомпьютера «Ломоносов»

Пиковая производительность	510 Тфлопс
Реальная производительность	408 Тфлопс
Число вычислительных узлов	5 130
Число процессоров	10 260
Число процессорных ядер	44 000
Число типов вычислительных узлов	3 (TB2, TB1.1, платформа на базе процессора PowerXCell 8i)
Основной тип вычислительных узлов	TB2
Процессор основного типа вычислительных узлов	Intel® Xeon X5670
Оперативная память	73 920ГБ
Общий объем дисковой памяти вычислителя	166 400ГБ
Занимаемая площадь	252 кв.м.
Энергопотребление вычислителя	2 МВт
Интерконнект	QDR Infiniband
Система хранения данных	Трехуровневая с параллельной файловой системой
Объем системы хранения данных	до 1 800ТБ
Операционная система	Clustrx T-Platforms Edition

«Т-Платформы» — крупнейший российский холдинг на рынке высокопроизводительных вычислений СНГ. Холдинг предоставляет полный комплекс решений и услуг в области суперкомпьютерных технологий и ресурсоемких расчетов:

- Разработка и производство суперкомпьютеров.
- Проектирование высокопроизводительных программно-аппаратных решений, обеспечивающих максимальную реальную производительность приложений заказчика.
- Аутсорсинг полного цикла высокопроизводительных вычислений, от аренды суперкомпьютерных мощностей до моделирования и экспертных расчетов.
- Разработка системного программного обеспечения для высокопроизводительных систем.
- Проектирование и создание суперкомпьютерных центров, а также управление ими для оптимизации затрат и получения прибыли.

Холдинг Т-Платформы
117198, г. Москва, Ленинский пр-т
д. 113/1, офис Е-307

тел: +7 (495) 956-54-90
факс: +7 (495) 956-54-15

sales@t-platforms.ru
<http://www.t-platforms.ru>

© Т-Платформы, 2011 г.