

Гибридные вычислительные системы на основе графических процессоров NVIDIA Tesla

Российский рынок высокотехнологичной продукции сегодня особенно остро нуждается в недорогих решениях, обеспечивающих высокую вычислительную производительность. При нынешнем уровне финансирования большинство научных организаций не могут себе позволить иметь столь необходимые для решения расчетных задач вычислительные ресурсы в достаточном объеме, а промышленные предприятия не в полной мере оправились еще от экономических потрясений. Для многих существенным ограничением являются и эксплуатационные затраты на содержание больших вычислительных систем. Одним из наиболее эффективных решений этих проблем может стать использование при создании высокопроизводительных систем графических процессоров, построенных на архитектуре параллельных вычислений CUDA компании NVIDIA.

Гибридные технологии используют самые мощные суперкомпьютеры в мире, программирование на CUDA преподают в институтах, производители серверных решений поддерживают GPU-вычислители в качестве стандартных компонентов, а производители программного обеспечения поддерживают массивно-параллельную архитектуру GPU для ускорения работы приложений.

Графические процессоры NVIDIA Tesla на архитектуре Fermi, обеспечивающие значительно более высокую производительность на операциях с вещественными числами двойной точности (по сравнению с предыдущим поколением), ECC-защиту памяти и ряд других серьезных усовершенствований, позволяют использовать GPU для широкого круга научных и прикладных задач и промышленных вычислений. Кроме того, гибридные вычислительные системы на базе NVIDIA Tesla существенно более компактны и энергоэффективны по сравнению со стандартными вычислительными системами. Например, гибридный вычислитель с производительностью 80 Тфлопс в сравнении с традиционным вычислителем на блэйд-серверах с современными процессорами Intel Xeon (Westmere/Nehalem) потребляет в 5-6 раз меньше электроэнергии, занимает в 3-4 раза меньше стоек и стоит в 6-7 раз дешевле, не говоря о разнице в эксплуатационных затратах. Такое решение в итоге экономит заказчику миллионы долларов. Конечно, не всем нужны такие мощные вычислители, но важность разницы в затратах при бюджетах в десятки сотни тысяч долларов только возрастает!

Кому это нужно и зачем

Не станем утомлять читателя полным перечнем задач, которые можно решать на гибридных вычислительных системах с GPU-ускорением (для этого есть web-сайты NVIDIA и ее партнеров), и тем более тех, которые можно решать на стандартных x86-серверах, входящих в состав гибридного кластера. Гибридные вычислительные системы представляют безусловный интерес для:

■ широкого круга научно-исследовательских и образовательных учреждений при решении теоретических, экспериментальных и прикладных расчетных задач различных областей науки – на таких системах может работать не только ПО собственной разработки, но и популярные приложения, такие, к

NVIDIA CUDA – это программно-аппаратная архитектура, позволяющая использовать GPU для вычислений общего назначения. Архитектура CUDA поддерживает популярные языки программирования и API, включая Microsoft DirectCompute, OpenCL, CUDA C, C++, CUDA Fortran и др. NVIDIA продолжает развивать среду разработки CUDA, увеличивая спектр инструментов и библиотек для исследователей и ученых. Тысячи разработчиков программного обеспечения, ученых и исследователей широко используют CUDA в различных областях, включая астрофизику, вычислительную биологию и химию, моделирование динамики жидкостей, электромагнитных взаимодействий, полученных путем компьютерной томографии, сейсмический анализ, трассировку лучей, обработку видео, восстановление изображений и многое другое.

Вся документация для программирования на CUDA находится в свободном доступе на сайте NVIDIA.

Программно-аппаратная модель CUDA поддерживается последними версиями широко известных промышленных, промышленных и финансовых приложений. Среди них: приложение для расчетов в биологии, квантовой физике и механике – Multi-GPU-симулятор молекулярной динамики AMBER версии 11; пакет прикладных программ для решения задач технических вычислений MATLAB версии R2010b от MathWorks; универсальная программная система конечно-элементного анализа ANSYS Mechanical версии R13; научно-инженерный пакет Mathematica 8 компании Wolfram Research. Совсем недавно NVIDIA анонсировала C-компилятор от The Portland Group (PGI) с поддержкой CUDA и x86 одновременно.

Крупнейшие суперкомпьютерные центры мира начали активное внедрение решений на GPU. Согласно последней, 36-ой, редакции рейтинга TOP 500 самых быстрых в мире суперкомпьютеров, графические процессоры лежат в основе трех из пяти самых мощных систем.

Место	Название	Страна	Производительность в Linpack	Кол-во GPU	Производитель GPU	Энергопотребление
1	Tianhe-1A	Китай	2.507 Пфлопс	7168	NVIDIA	4.04 МВт
2	Jaguar	США	1.75 Пфлопс	–	–	6.95 МВт
3	Nebulae	Китай	1.27 Пфлопс	4640	NVIDIA	2.55 МВт
4	Tsubame 2.0	Япония	1.192 Пфлопс	4200	NVIDIA	1.340 МВт
5	Hopper	США	1.05 Пфлопс	–	–	2.93 МВт

Интересно, что Национальная Лаборатория Окриджа (США), которая до последнего времени располагала самым быстрым в мире суперкомпьютером (№ 2 в текущей версии рейтинга TOP500), также заявила о том, что ее новый суперкомпьютер, который должен стать в 10 раз мощнее, чем существующий, будет построен именно на графических процессорах.

примеру, как MATLAB и Mathematica, которые уже поддерживают GPU-ускорение;

- ▶ конструкторских и дизайн-бюро промышленных предприятий, занимающихся промышленным дизайном, расчетами, анализом и моделированием, а также с целью ускорения CAE-систем, таких как, например, пакеты ANSYS и MSC.

Готовые кластеры – экономия ресурсов заказчика

Для адресного решения таких задач в условиях жесткой борьбы за ИТ-бюджеты компанией “Открытые технологии” были разработаны стандартные конфигурации компактных гибридных вычислительных систем на GPU – экономичные, высоконадежные, сбалансированные и полностью законченные системы, занимающие всего одну монтажную стойку высотой 24U (чуть более метра) или 42U (чуть менее двух метров).

Охарактеризуем кратко основные особенности этих решений:

- ▶ стоимость системы сопоставима со стоимостью нескольких профессиональных рабочих станций при несоизмеримо более высокой совокупной производительности и возможностях;
- ▶ сложность выбора конфигурации сведена к минимуму – во всех решениях используются стандартные компоненты, оптимальные для боль-

шинства задач, заказчику необходимо указать только номер конфигурации (или пиковую производительность), используемый тип GPU-вычислителей, необходимость высокоскоростной сети Infiniband (например, при использовании MPI), необходимость сетевой системы хранения данных (NAS), тип операционной системы (Linux либо MS Windows). При этом будут рассмотрены и любые альтернативные конфигурации и при необходимости разработано индивидуальное решение любой сложности;

- ▶ система поставляется в виде готовой к работе платформы, имеющей в своем составе все необходимые компоненты для функционирования и управления – соответственно, решение вопросов интеграции системы в инфраструктуру сводится к минимуму;
- ▶ возможно использование серверного варианта GPU-вычислителя NVIDIA Tesla серии S для обеспечения повышенной надежности, удобства обслуживания и гибкости в подключении к серверам кластера, что особенно важно при необходимости долгой и непрерывной работы. Благодаря партнерству с NextIO, поставщиком серверных решений на базе NVIDIA Tesla GPU, компания “Открытые технологии” предлагает в качестве GPU-вычислителя решение VCORE Express 2050 (NVIDIA Tesla S2050), а также эксклюзивное решение на российском рынке – VCORE Express 2070 (NVIDIA Tesla S2050)





21 апреля,
Москва, отель
„Ренессанс“

IDC Datacenters Transformation Roadshow 2011

Конференция
«Системы хранения данных,
виртуализация и облачные вычисления»

Темы конференции:

- Виртуализация: от серверов к СХД. Перспективы технологии.
- Дата-центры. Какими будут ЦОДы в ближайшем будущем?
- СХД следующего поколения. Иерархические системы хранения, новое ПО, SSD
- Управление ИТ-сервисами, аутсорсинг и жизненный цикл приложений
- Облачные вычисления
- Решения plug&play

Дополнительная информация и регистрация на сайте www.idc.com/russia и по телефону 495 661 6166

Золотой партнер:



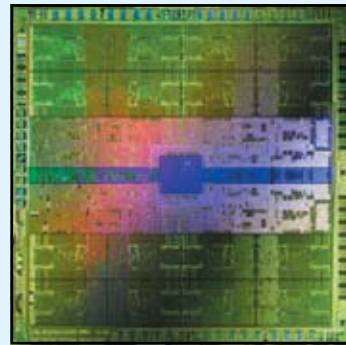
Генеральный медиа-партнер:



Информационная поддержка:



Исключительная вычислительная мощь GPU объясняется особенностями его архитектуры. В отличие от CPU, который состоит из нескольких ядер (на большинстве современных систем от 2-х до 4-х), графический процессор изначально создавался как многоядерная структура, в которой количество ядер измеряется сотнями. Разница в архитектуре обуславливает и разницу в принципах работы. Если архитектура CPU предполагает последовательную обработку информации, то GPU исторически предназначался для обработки компьютерной графики,



Новейшая архитектура NVIDIA Fermi располагает 512-ю вычислительными ядрами

поэтому рассчитан на массивно параллельные вычисления.

Каждая из этих двух архитектур имеет свои достоинства. CPU лучше работает с последовательными задачами. При большом объеме обрабатываемой информации очевидное преимущество имеет GPU. Условие только одно – в задаче должен наблюдаться параллелизм.

с увеличенным до 6 ГБ на GPU объемом памяти, что для ряда задач может быть критически важным – в частности при параллельных вычислениях (большие размерности задач, большие наборы данных/элементов и т.д.).

С подробным описанием возможных конфигураций предлагаемых гибридных решений можно ознакомиться на сайте компании “Открытые технологии” – http://ot.ru/download/doc/whitepaper/whitepaper_gpu-clusters.pdf.

Партнерство с компаниями NVIDIA и NextIO и собственный опыт в построении больших вычислительных кластеров, таких как МСЦ РАН, позволяет компании “Открытые технологии” предлагать своим заказчикам решения по построению гибридных суперкомпьютеров на GPU любого уровня производительности. Среди решений NextIO – не только GPU-вычислители на графических процессорах NVIDIA (VCORE Express и VCORE Extreme), но и решения для эффективного построения высокоскоростной системы хранения для вычислительных кластеров на основе SSD-дисков с высокоскоростным доступом по интерфейсу PCI-Express.

В компании создан центр компетенции CUDA, в котором проводится подготовка специалистов по вопросам программирования и оптимизации программного кода для GPU-вычислений, а также осуществляется адаптация существующего ПО для архитектуры CUDA.

Опыт компании “Открытые технологии” в построении центров обработки данных и ИТ-инфраструктуры помогает реализовать полностью законченные решения, включая решение задач оптимального размещения компонентов вычислительных комплексов, энергообеспечения и охлаждения, управления и мониторинга, информационной безопасности.

По материалам компании “Открытые технологии”