

РОССИЙСКИЕ СУПЕРКОМПЬЮТЕРЫ ТРИ ГОДА ПРОГРЕССА

М.Шейкин max.shaking@ya.ru

Первый национальный суперкомпьютерный форум, проходивший в Переславле-Залесском в ноябре 2012 года, безусловно, стал важнейшим событием для российской суперкомпьютерной отрасли. Несмотря на то, что Россия в целом отстает от ведущих стран в области суперкомпьютерных технологий, это мероприятие показало, что суперкомпьютерная отрасль в России в течение последних лет развивается достаточно интенсивно. Одна статья не может охватить все доклады и дискуссии форума, поэтому мы выбрали несколько тем, по которым можно судить о развитии российской суперкомпьютерной отрасли.

29 и 30 ноября 2012 года в Переславле-Залесском, в Институте программных систем (ИПС) им. А.К.Айламазяна РАН проходил Первый национальный суперкомпьютерный форум (НСКФ-2012). Он был посвящен вопросам создания и практики применения суперкомпьютерных технологий и проводился под эгидой Российской академии наук и Российского фонда фундаментальных исследований.

В течение двух ноябрьских дней гости форума могли не только послушать выступления ученых, разработчиков и коммерческих представителей фирм и институтов, имеющих отношение к суперкомпьютерным технологиям, но и обсудить интересующие темы на круглых столах. Иллюстрациями к выступлениям и обсуждениям служили экспонаты проходившей в рамках форума выставки.

По словам академика С.М.Абрамова, директора ИПС им. А.К.Айламазяна РАН, российская суперкомпьютерная отрасль, наконец, "выросла из коротких штанишек" и стала на ноги. Возникла естественная потребность фирм и институтов показывать свои разработки, общаться и делиться опытом. Такие возможности предоставляются на

форумах. Подобные мероприятия регулярно проводятся за рубежом, специально для них ведущие фирмы готовят новости и презентации своих новых продуктов. Теперь же мероприятие аналогичного уровня появилось и в России, и можно совершенно точно сказать, что Первый национальный суперкомпьютерный форум крайне важен для России, и "первый блин" определенно удался.

В целом успехи России в области суперкомпьютерных технологий (как это не обидно) пока что можно измерять лишь степенью отставания от передовых в этой области стран. Такая ситуация обусловлена и тяжелым "наследием" 1990-х годов, и отсутствием четких программ развития суперкомпьютерной отрасли за предыдущие 10 лет. Определенные работы в этой отрасли, конечно, велись, но суперкомпьютерные технологии не могут развиваться без координации усилий всех разработчиков аппаратных компонентов и ПО.

Тем не менее, за последние три года ситуация кардинально изменилась. По словам директора Департамента научно-производственной базы ядерно-оружейного комплекса ГК "Росатом" С.Е.Власова, если в 2009 году отставание в этой

области было приблизительно в 100 раз, то сейчас эта величина снизилась на порядок. Более того, три года назад 90% от всего промышленного ПО для моделирования поведения сложных изделий было американским, но, как заявил С.Е.Власов, к 2012 году отечественные пакеты ПО уже используются для решения 70% инженерных задач (однако осталось неясным, какие именно задачи имелись в виду. Применительно ко всей инженерной деятельности это утверждение звучит излишне оптимистично).

Так же были упомянуты базовые комплекты компактных суперЭВМ терафлопсной производительности, которые сейчас работают на 70 предприятиях. Существует концепция экзафлопсных вычислительных систем (они активно разрабатываются в США и Европе), которая была одобрена межведомственной рабочей группой по применению суперкомпьютеров в отраслях промышленности, организованной в рамках реализации программы создания информационного общества.

МИКРОПРОЦЕССОРЫ "ЭЛЬБРУС" – ДОГНАТЬ И ПЕРЕГНАТЬ X86

Пожалуй, самая известная и широко освещаемая прессой российская разработка в области компьютерной техники – процессоры "Эльбрус" ЗАО "МЦСТ". К сожалению, содержание докладов представителей МЦСТ на конференциях нескольких последних лет изменялось не сильно. О процессорах "Эльбрус" и возможностях создания суперЭВМ на их основе можно узнать, например, в [1], здесь же мы вкратце опишем особенности архитектуры этих процессоров.

Первые микропроцессоры серии "Эльбрус" появились в 2000 году. В отличие от RISC-процессоров семейства "МЦСТ-R", совместимых со SPARC, процессоры "Эльбрус" имеют архитектуру VLIW (Very long instruction word – очень длинная машинная команда). По заявлению разработчиков "Эльбруса", эта архитектура может стать основой для создания экзафлопсных вычислительных систем.

Как следует из ее названия, одна машинная команда для VLIW-процессора содержит несколько инструкций, которые должны выполняться одновременно. Таким образом, распараллеливание вычислительного процесса происходит на этапе компиляции. Это позволило упростить сам процессор, отказавшись от блоков аппаратного распределения задач, уменьшив тем самым его энергопотребление. Однако эффективность выполнения вычислений теперь очень сильно

зависит от качества оптимизации кода компилятором. Поэтому, говоря о VLIW-процессоре, правильней иметь в виду аппаратно-программный комплекс "процессор+компилятор".

Текущая реализация архитектуры "Эльбрус" позволяет выполнять в каждом такте до четырех независимых операций с плавающей точкой с двойной точностью. Одновременно производятся целочисленные вычисления, загрузка операндов из памяти, проверка логических условий, подготовка и осуществление переходов – всего до 18 операций в одной "длинной" команде на линейных участках кода и до 23 операций – при выполнении циклов.

Кроме этого, в архитектуре "Эльбрус" реализована асинхронная подгрузка данных. Процессор не содержит аппаратных блоков предварительной загрузки, вместо этого на этапе компиляции в код добавляются инструкции подгрузки данных еще до того момента, когда они будут необходимы. Эти данные хранятся в отдельном буфере подкачки, откуда извлекаются по мере выполнения кода.

Для ускорения выполнения переходов применяется механизм подготовленных переходов. Кроме основного конвейера инструкций, параллельно запускаются три дополнительных. Если принимается решение о переходе, процессор может переключиться на один из дополнительных конвейеров без потери тактов либо остаться в текущем. Это позволяет существенно ускорить циклы со сложными условиями.

Важная особенность процессоров "Эльбрус" – возможность работы в защищенном режиме с использованием дескрипторов (указателей на область памяти) длиной 128 бит, в которых, кроме самого указателя, хранятся его тип и диапазон допустимых значений. В этом случае невозможно преобразовывать в дескриптор целочисленное значение (адресная информация отделена от прочей), а сам дескриптор может указывать на строго определенную область памяти. У программы, исполняющейся в режиме защищенных вычислений, контролируются все обращения в память на предмет выхода за разрешенный диапазон адресов, блокируется некорректная работа с указателями на локальные (временные) объекты и диагностируется использование неинициализированных данных. Такой подход позволяет защитить систему от возможных атак и ошибок и в разы упростить отладку программы. Плата за надежность в этом случае – некоторое (до 20%) замедление выполнения программы.

Кроме ОС на базе Linux, прикладных программ и инструментов разработчика, для "Эльбруса" создана специальная программная "прослойка" для преобразования кода для процессоров архитектуры x86 в код "Эльбруса". Производительность системы с процессором "Эльбрус-2С+" с тактовой частотой 500 МГц в режиме выполнения кода x86 примерно равна производительности процессора Pentium-M, работающего на частоте 1000 МГц (рис.1).

Сегодня МЦСТ предлагает процессор "Эльбрус-2С+" с частотой 500 МГц, имеющий два ядра "Эльбрус" общей производительностью до 16 GFlops и четыре ядра ЦОС - до 12 GFlops при рассеиваемой мощности 25 Вт. Для работы в паре с этим процессором также создан "южный мост" КПИ с частотой 250 МГц и имеющий все необходимые современному компьютеру порты ввода-вывода.

Поддерживается возможность объединения до четырех процессоров (и, при необходимости, "южных мостов"), использующих общую память, без каких-либо дополнительных узлов. Связь осуществляется по схеме "каждый с каждым". Многопроцессорные системы, в свою очередь, могут объединяться через когерентные межпроцессорные каналы со скоростью передачи данных 1 Гбит/с.

На основе процессора "Эльбрус-2С" и RGB разработаны и созданы плата "Монокуб" для моноблочных ПК и двух- и четырехпроцессорные промышленные модули формфактора CompactPCI 6U (рис.2). В декабре 2012 года совместно с фирмой Kraftway была выпущена пробная партия настольных моноблочных компьютеров на базе платы "Монокуб".

В 2013 году намечено завершить разработку четырехъядерного процессора "Эльбрус-4С" по 65-нм технологии производительностью

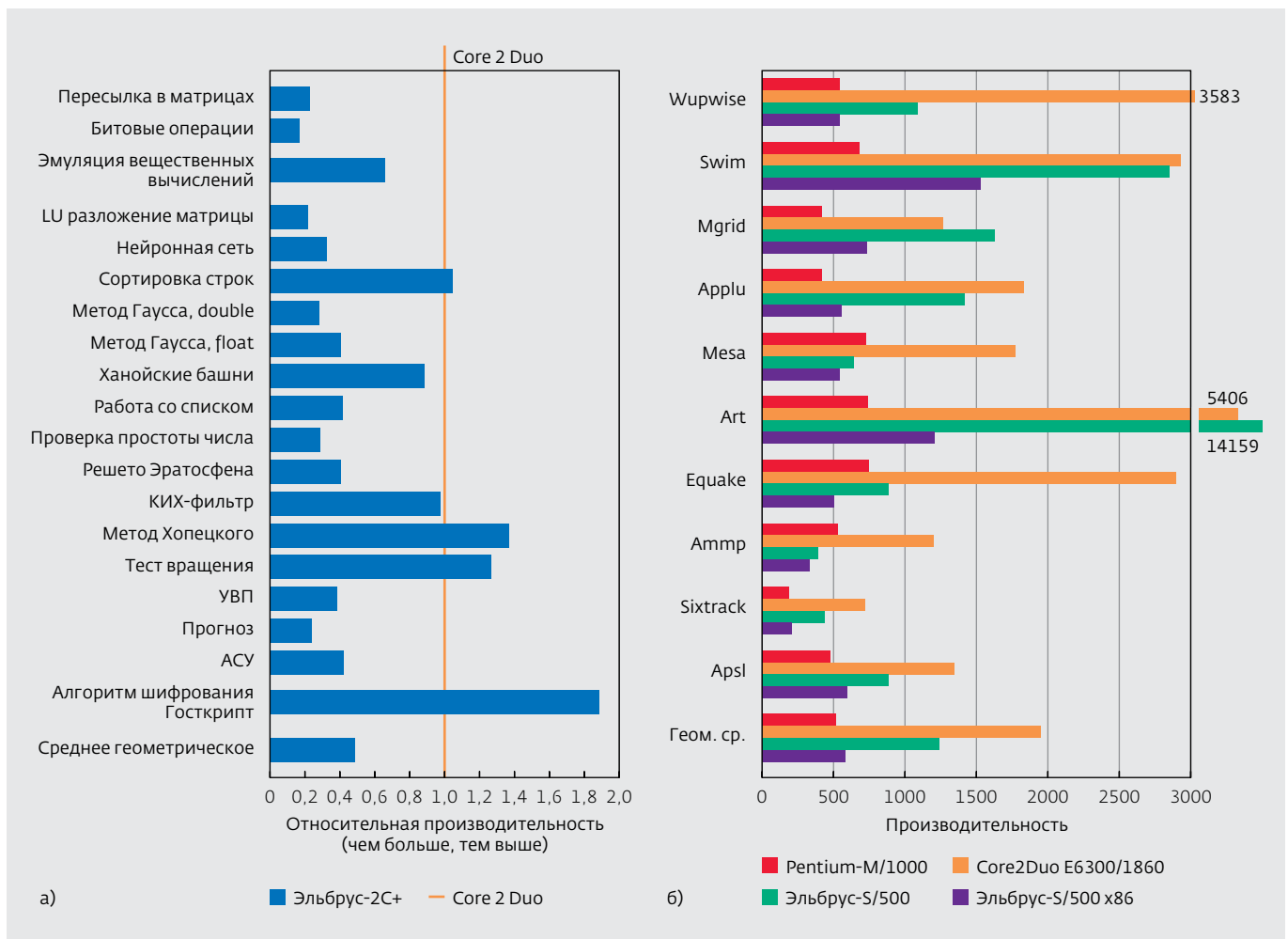


Рис.1. Сравнение производительностей процессоров "Эльбрус" и x86 при выполнении тестов SPEC CPU2000 (а) и пользовательских задач (б)

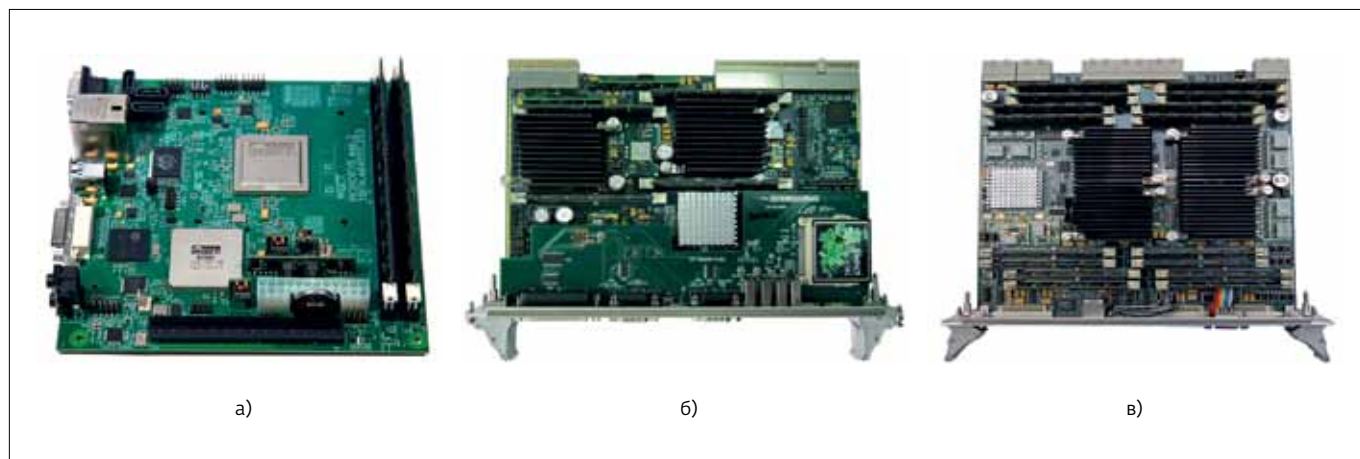


Рис.2. Плата МЦСТ "Монокуб" (а), модули "МВКУБ-С" (б) и "МВК-С" (в)

32 GFlops. Он будет работать на частоте 1 ГГц и выделять до 80 Вт тепла. Немаловажно, что его производство планируется начать на российской фабрике в Зеленограде. В перспективе – создание в 2015 году восьмиядерного процессора с частотой более 1 ГГц и выполненного по 28-нм технологии.

"АНГАРА" – ПЕРСПЕКТИВНАЯ РОССИЙСКАЯ ПЛАТФОРМА НА БАЗЕ ПРОЦЕССОРОВ AMD

Мощность отечественных микропроцессоров пока еще недостаточна для построения высокопроизводительных по современным меркам суперкомпьютерных систем. Поэтому разработчики отечественных вычислительных платформ используют зарекомендовавшие себя решения зарубежных производителей.

В научно-исследовательском центре электронной вычислительной техники (ОАО "НИЦЭВТ", Москва) разрабатывается суперкомпьютерная система транспетафлопсной производительности на основе созданной там же вычислительной платформы, получившей название "Ангара". В ней используются процессоры AMD семейств Bulldozer и Piledriver. В один вычислительный модуль устанавливаются четыре процессора AMD G34 и 16 модулей памяти DDR3 (до 512 ГБ памяти с максимальной пропускной способностью на один сокет 58 Гбайт/с). 22-слойная печатная плата модуля имеет размеры 520×411,5 мм (рис.3а). Энергопотребление вычислительной платформы составляет не более 650 Вт.

Процессоры серии 6300 поддерживают технологию виртуализации AMD Virtualization (AMD-V) и работу с четырьмя каналами памяти на частоте до 1866 МГц. Связи между

процессорами организованы по системной шине Hypertransport 3.1 с пропускной способностью 102,4 Гбит/с. Для соединения вычислительных модулей применяется коммуникационная сеть собственной разработки. При использовании процессоров AMD Opteron 6380 удалось добиться пиковой производительности 640 GFlops.

Система охлаждения вычислительной платформы "Ангара" – комбинированная. Самые горячие элементы (процессор и СБИС маршрутизатора) охлаждаются водой с антигрибковыми присадками, а остальные – продувкой воздуха (рис.3б). Вычислительные блоки подключаются к жидкостной системе охлаждения с помощью бескапельных быстроразъемных соединителей. Вычислительная система из 32 стоек пиковой производительностью 1,3 PFlops занимает площадь 75 м² и потребляет 1,5 МВт электроэнергии.

Базовая система ввода-вывода (BIOS) вычислительной платформы создана на основе исходного кода проектов Coreboot и SeaBIOS. BIOS относительно быстро инициализируется и позволяет встраивать средства доверенной загрузки. В BIOS организована поддержка всех основных компонентов внешнего и встроенного оборудования, для нее проверено функционирование современных серверных ОС – Windows Server 2012, RHEL Server 6.3, SLES 11.

Опытные образцы платформы "Ангара" уже изготовлены и прошли тестирование. Во втором квартале 2013 года планируется завершить подготовку к ее серийному производству. На базе платформы "Ангара" уже разрабатывается высокопроизводительная транспетафлопсная суперкомпьютерная система. При создании систем такого рода приходится решать ряд важных

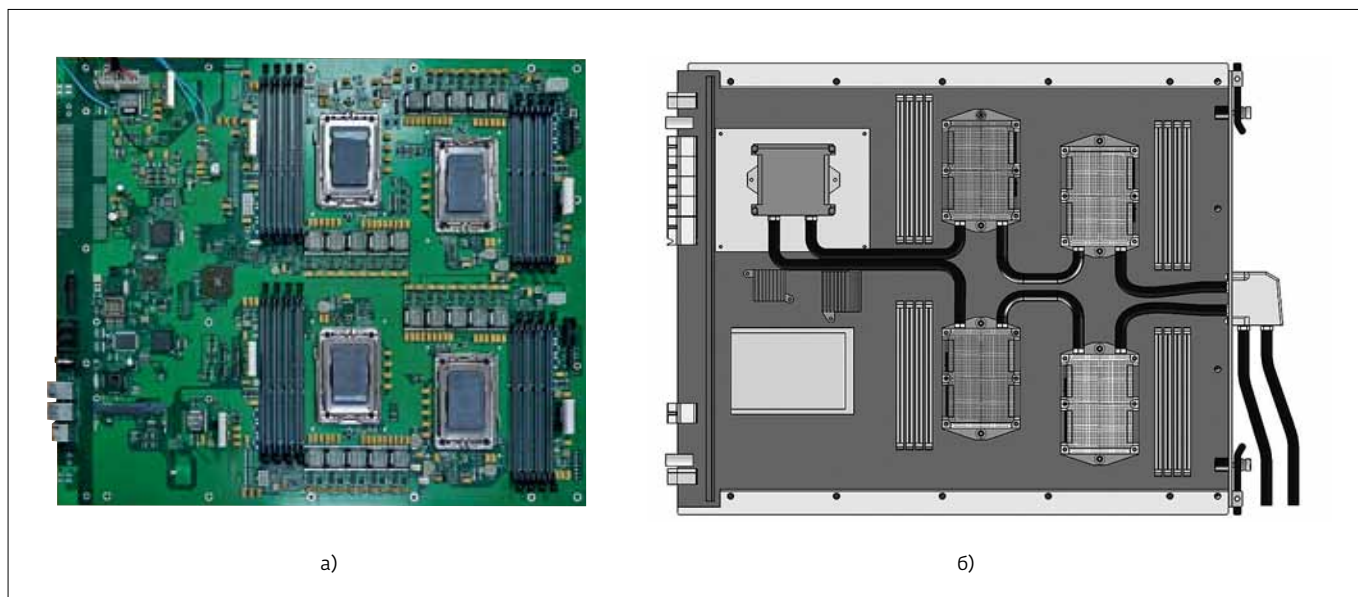


Рис.3. Опытная плата "Ангара" (а) и схема ее системы охлаждения (б)

задач. В первую очередь необходимо подготовить инфраструктуру комплекса – подсистемы питания, охлаждения, конструкцию стоек, шкафов и т.д. В частности, одно из требований при разработке конструкции серверной стойки суперкомпьютерных систем – достижение максимально возможной плотности компоновки электронных компонентов при оптимальном соотношении производительность/стоимость. Высокая плотность компоновки в конечном итоге позволяет существенно снизить затраты на производство и эксплуатацию вычислительного центра за счет сокращения суммарной площади помещений и стоимости системы охлаждения.

Для нового вычислительного комплекса была разработана ОС на базе модифицированного ядра Linux, написаны драйверы, библиотеки нижнего уровня, средства параллельного программирования, поддержки виртуализации, отладки, профилирования и т.д. Кроме системного ПО, требуются программы для выполнения задач на вычислительной системе – инфраструктура управления узлами системы, мониторинга и диагностики, запуска и планирования задач, комплекс ПО для автоматизации развертывания узлов кластера и поддержки отказоустойчивости. Для разработки прикладного ПО необходим комплекс программных средств. И, наконец, перед разработчиками стоит задача создать общую методологию эксплуатации вычислительной системы.

КОМПАКТНЫЕ СУПЕРЭВМ РФЯЦ-ВНИИЭФ

В ФГУП "РФЯЦ-ВНИИЭФ" с 2009 года ведутся научно-исследовательские и опытно-конструкторские работы по созданию высокопроизводительных компактных суперЭВМ для широкого круга научных и инженерных задач. На первом этапе были созданы две такие ЭВМ производительностью 1 TFlops на основе универсальных процессоров с арифметическими ускорителями и без них (рис.4а). В дальнейшем их производительность была увеличена до 3,5 TFlops.

В 2012 году разработана компактная суперЭВМ следующего поколения производительностью 5 TFlops без использования арифметических ускорителей. Прототип специализированной (с арифметическими ускорителями) компактной суперЭВМ обеспечивает производительность не менее 8 TFlops (рис.4б).

СИСТЕМА ПОГРУЖНОГО ОХЛАЖДЕНИЯ IMMERS

Без охлаждения невозможна работа ни одного современного компьютера, а тем более – мощных суперкомпьютерных систем. Сотрудники ИПС им. А.К.Айламазяна РАН совместно с группой компаний "Сторус" создали систему жидкостного охлаждения высокопроизводительных вычислительных систем, основанную на принципе полного погружения плат в диэлектрическую жидкость (рис.5). Все элементы вычислительной системы находятся в герметичном корпусе, заполненном специальной жидкостью. Ее циркуляция обеспечивает отвод тепла от компонентов



Рис.4. Компактные суперЭВМ, разработанные РФЯЦ-ВНИИЭФ

в окружающую среду. Такая система работает значительно тише, чем традиционные воздушные системы охлаждения. В перспективе появится возможность повторно использовать тепло. Важнейшее же преимущество системы охлаждения с полным погружением – ее экономичность. Для ее работы требуется всего несколько процентов от мощности, которую потребляет весь вычислительный комплекс. Для сравнения, гибридные (жидкостно-воздушные) системы потребляют до 40%, а воздушные – до 80% от общей мощности. Кроме этого, системы с погружным охлаждением очень гибкие – нет необходимости изготавливать специальные элементы системы (радиаторы, теплоотводящие элементы и т.д.) для новых плат и модулей.

На базе системы охлаждения Immers построен макетный образец суперкомпьютера производительностью 5 TFlops.

ВЫЧИСЛИТЕЛЬНЫЙ КЛАСТЕР В МФТИ – ПЕРВЫЙ С CUDA

Одно из перспективных направлений построения суперкомпьютерных систем – использование процессоров NVIDIA G80 и последующих поколений, поддерживающих программно-аппаратную архитектуру CUDA (Compute Unified Device Architecture). О таком опыте рассказал ассистент кафедры интеллектуальных электронных систем и технологий Московского физико-технологического института (МФТИ) Д.В.Пантюхин.



Рис.5. Система погружного охлаждения Immers

Созданный в МФТИ исследовательский 16-узловой кластер на базе процессоров G70 с экспериментальным прикладным программным обеспечением стал первым в России суперкомпьютером на базе процессоров NVIDIA. Экспериментальная система на процессорах NVIDIA G80 с архитектурой CUDA была создана в 2007 году. Опыт ее эксплуатации и анализ тенденций развития вычислительной техники в целом и процессоров NVIDIA в частности позволили разработать и реализовать проект суперкомпьютерного центра МФТИ.

Вычислительный комплекс состоит из четырех узлов (планируется его дальнейшее расширение), сервера хранения, к которому подключен дисковый массив, сервера визуализации и двух сетей коммутации Gigabit Ethernet и InfiniBand QDR. В состав комплекса также входят рабочие станции.

Сеть управления всеми компонентами комплекса построена на двух Gigabit Ethernet свитчах с резервированием. Для межузловых взаимодействий используется свитч InfiniBand QDR, к которому, кроме прочего, подключен сервер хранения. Общий объем доступного дискового массива – около 20 ТБ.

Один вычислительный узел содержит два шестиядерных процессора Intel Xeon X5670 и оперативную память DDR-3 (1066 МГц) ECC общим объемом 92 ГБ. К каждому узлу подключено по две вычислительные системы NVIDIA Tesla S2050 с 12 ГБ памяти GDDR5. Они потребляют не более 900 Вт энергии и при этом позволяют достичь производительности 2,5 TFlops с числами одинарной и двойной точности (в соответствии со стандартом IEEE 754). Таким образом, каждый

вычислительный узел может достичь производительности 5 TFlops, а суммарная производительность суперЭВМ – около 20 TFlops.

Сервер визуализации оснащен двумя графическими платами NVIDIA Quadro FX 5800. Он подключен к серверу хранения напрямую и отображает результаты измерений с высоким разрешением.

Рабочие станции с процессором NVIDIA Tesla C1060 выполняют некоторые вычисления без использования суперЭВМ, а также проводят разработку и отладку программного обеспечения.

При выборе ПО для суперкомпьютера было принято решение остановиться на CentOS 5.x, эта же система используется и на серверах хранения. В перспективе планируется исследование по сравнению работы системы на развивающихся ОС Ubuntu Server и openSUSE и на Windows HPC Server 2008 R2. Для рабочих и учебных станций была выбрана многосистемная конфигурация ОС – Ubuntu 10.10 и Windows 7.

Для организации вычислений на процессорах NVIDIA Tesla используется ПО CUDA Toolkit 3.2 с запланированным обновлением до версии 4.0.

Вычисления с применением нескольких серверов организованы с помощью Open MPI 1.4.3, который входит в состав ПО QLogic OFED+ из пакета QLogic Infiniband Fabric Suite. Планируется провести исследования технологии GPUDirect, позволяющей более эффективно проводить расчеты на графических платах, установленных в различных вычислительных серверах и соединенных с помощью коммуникационной сети Infiniband.

ПЛИС ВМЕСТО ПРОЦЕССОРА

Известно, что весь потенциал вычислительной системы раскрывается лишь при выполнении небольшого числа задач. Иными словами, заявленная пиковая производительность достигается только при определенных видах вычислений. При решении так называемых сильно связанных задач (требующих активного обмена информацией между процессорами) увеличение числа процессоров в системе не только не приводит к ускорению вычислений, а даже наоборот, замедляет их. Это происходит из-за того, что все больше времени требуется на служебные процедуры по организации процесса вычислений.



Рис.6. Вычислительный модуль "Саиф" (а), его плата (б), модуль "Ригель" (в) и его плата (г)

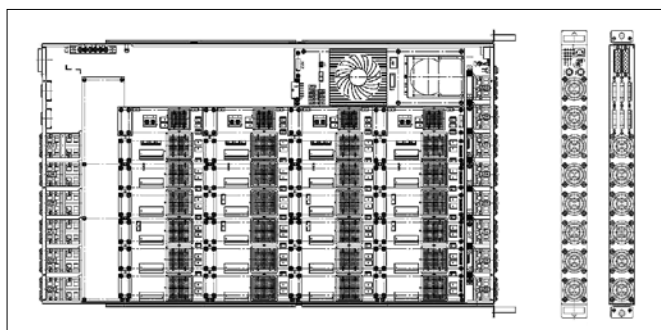


Рис.7. Компоновка модуля 24V7-750

ПЛИС обладают значительно большим вычислительным потенциалом, который в полной мере может быть реализован в системах, содержащих множество кристаллов ПЛИС, выполняющих роль основных вычислительных элементов [2, 3].

И.И.Левин (д.т.н., НИИ многопроцессорных вычислительных систем Южного федерального университета (НИИ МВС ЮФУ), Таганрог) рассказал о концепции построения распределенных вычислительных сетей (РВС) на основе программируемых логических интегральных схем (ПЛИС), разработанной в НИИ МВС ЮФУ. По производительности такие системы без особого труда конкурируют с традиционными процессорными вычислительными системами. Их особенность – возможность формирования специальной, ориентированной на конкретные задачи архитектуры, которую при необходимости можно изменить.

Идея реконфигурируемых РВС на ПЛИС – современный взгляд на структуру аналоговых вычислительных полей на дискретных элементах. Для таких систем задача представляется в виде графа, где вершины – это операции, а ребра – информационные зависимости. В отличие от процессорных систем, где большое количество процессоров усложняет создание эффективных алгоритмов обработки данных, сложность программирования РВС на ПЛИС уменьшается с ростом количества ПЛИС. "КПД" такой системы (т.е. отношение реальной производительности к пиковой) крайне высок. Открытая масштабируемая архитектура позволяет преодолеть многие ограничения, накладываемые архитектурой с ортогональной коммутацией. В качестве основы для РВС разработчики из НИИ МВС ЮФУ использовали ПЛИС высокой интеграции фирмы Xilinx семейства Virtex-2 и выше, соединенные в вычислительную сеть высокоскоростными каналами передачи данных LVDS и Rocket GTX.

Серийно производящиеся с 2010 года вычислительные модули "Саиф" и "Ригель" (рис.6) имеют производительность до 1,6 TFlops для операций одинарной точности и 0,5 TFlops – для операций двойной точности. Производительность вычислительной стойки (от 24 до 36 модулей) составляет до 51,8 или 18 TFlops, соответственно.

Нужно отметить, что применение ПЛИС семейства Virtex-6 позволяет увеличить производительность в 1,5-2 раза по сравнению с аналогичным решением на основе ПЛИС семейства Virtex-5, стоимость модуля при этом остается прежней. Использование же ПЛИС Virtex-7 улучшает производительность по сравнению с изделиями на основе Virtex-6 примерно в 1,7 раза. Таким образом, созданные вычислительные модули нового поколения – очень перспективны и конкурентоспособны для построения РВС различных архитектур и конфигураций.

Разработанный по государственному контракту перспективный вычислительный модуль 24V7-750 на основе ПЛИС Virtex-7 предназначен для построения реконфигурируемой вычислительной системы РВС-7 с пиковой производительностью до 1015 операций с фиксированной точкой. Сегодня уже изготовлены опытные экземпляры, которые прошли испытания, и идет подготовка к серийному производству. Производительность модуля – до 2,6 TFlops при выполнении операций одинарной точности и 0,82 TFlops – при операциях двойной точности, а вычислительный комплекс с числом модулей от 24 до 36 будет иметь производительность до 95 и 29,4 TFlops, соответственно.

Модуль 24V7-750 (рис.7) содержит четыре вычислительные платы, на каждой из которых установлено по шесть ПЛИС XC7V585T-1FFG1761, формирующих вычислительное поле. На платах также имеются 12 микросхем памяти частотой 400 МГц общим объемом 3 Гб (к каждой ПЛИС подключено по две микросхемы памяти), контроллер интерфейса LVDS, ПЛИС контроллера вычислительной платы, компоненты подсистем синхронизации, загрузки ПЛИС и питания.

ПЛИС вычислительного поля соединены между собой 144 каналами 800-МГц интерфейса LVDS, он же используется и для последовательного соединения вычислительных плат. Для связи с другими вычислительными модулями используется 12 каналов LVDS (по 25 дифференциальных пар каждый), работающих на той же частоте. Два канала LVDS (по 20 дифференциальных пар каждый) используются для связи с модулем ввода-вывода

(связь с персональным компьютером) и внешней аппаратурой.

Кроме плат, в корпусе модуля расположены жесткий диск, блок питания, управляющий модуль и плата индикации режимов работы. Для охлаждения вычислительных плат используются семь вытяжных вентиляторов. Конструктивное исполнение вычислительного модуля 24V7-750 соответствует международному стандарту МЭК 60297-2. Модуль потребляет не более 300 Вт.

Понятно, что принципы программирования РВС на ПЛИС отличаются от таковых для вычислительных систем на микропроцессорах. Если для традиционной системы достаточно описать лишь процесс вычислений (т.е. создать программу на языке высокого или низкого уровня), то для РВС на ПЛИС необходимо задать саму структуру, в которой будут выполняться вычисления. Этот этап часто вызывает трудности, так как "традиционное" программирование подразумевает неизменную аппаратную структуру вычислительной системы.

Один из основных инструментов программирования РВС на основе ПЛИС – специально созданный для этого язык высокого уровня COLAMO, который описывает реализацию параллельного алгоритма. Для представления такой организации вычислений в языке COLAMO используется понятие "кадр" – неразрывная совокупность вычислительной структуры фрагмента задачи и множества операций чтения-записи входных и выходных потоков данных. Применение языка Colamo позволило более чем в три раза сократить время разработки ПО для РВС.

Написанная на COLAMO программа транслируется в "привычный" для ПЛИС код VHDL, при этом для каждой ПЛИС в системе может создаваться свой код. Далее стандартными средствами разработки VHDL преобразуется в битовый массив для загрузки в ПЛИС. У этого метода есть один достаточно серьезный недостаток. В современных ПЛИС, например, серии Virtex-7, применяемых в 24V7-750, содержится 58 млн. эквивалентных вентиляей. Трансляция VHDL-кода в битовый массив для таких ПЛИС существующими средствами может занять очень много времени (иногда до нескольких суток), что очень сильно замедляет процесс отладки кода. Разрабатываемое ПО нового поколения позволит не создавать каждый раз битовый массив для прошивки ПЛИС. Программирование РВС в этом случае будет осуществляться на двух уровнях – создание

архитектуры системы, которая однократно будет транслирована и загружена в ПЛИС, и программы на языке COLAMO ("софт-архитектуры"), которая будет работать "поверх" существующей архитектуры. Плата за сокращение времени разработки и отладки – небольшое уменьшение производительности готовой системы.

РАЗВИТИЕ НАЦИОНАЛЬНЫХ ГРИД-СЕТЕЙ В БЕЛАРУСИ

Модель распределенных вычислений – мощнейший инструмент, позволяющий практически неограниченно наращивать вычислительную мощность суперкомпьютерной системы. Для объединения удаленных гетерогенных вычислительных узлов используются так называемые грид-сети. В России и странах СНГ ведутся достаточно масштабные работы по созданию национальных грид-сетей. Об этом рассказали гости форума из Татарстана и Беларуси.

Состояние и планы развития национальной грид-сети Республики Беларусь были темой выступлений представителей Объединенного института проблем информатики (ОИПИ) НАН Беларуси (Минск). Грид-сеть Беларуси создана в рамках программы Союзного государства "СКИФ-ГРИД" [4] и представлена двумя сегментами с программным обеспечением промежуточного уровня gLite (объединяет ресурсные центры ОИПИ НАН Беларуси, ОИЭЯИ-Сосны НАН Беларуси, БГУ и БНТУ) и Unicore (ОИПИ НАН Беларуси, БГУИР и ГрГУ им. Я.Купалы). Сеть интегрирована в опытный участок общего информационно-вычислительного пространства Союзного государства и в Европейскую грид-инфраструктуру EGI.

Управляющие функции осуществляют Удостоверяющий центр (регистрационный центр, сертификационный центр, база данных программного комплекса удостоверяющего центра, онлайн-репозиторий) и Операционный центр, в который входят единый реестр грид-ресурсов и пользователей грид-сети, служба мониторинга ресурсов, служба учета ресурсов, служба единого файлового пространства, а также веб-сайт Операционного центра. Сетевая инфраструктура сети включает коммуникационные узлы доступа, высокоскоростные волоконно-оптические каналы передачи информации, сетевое, коммуникационное и серверное оборудование академической сети BASNET. Общая координация и организация работ по управлению Национальной грид-сетью возложены на

Национальный центр грид-технологий на базе ОИПИ НАН Беларуси.

Создатели белорусской грид-сети столкнулись с проблемой недостаточного финансирования. Недостаток денежных средств закономерно приводит к нехватке специалистов, обслуживающего персонала, а также к дефициту оборудования и запасных частей.

В перспективе планируется создание центров обработки данных с производительностью 30-50 TFlops. Предполагается развитие корпоративных грид-сетей и включение их в общую национальную сеть. Наряду с наращиванием вычислительных мощностей предусматривается увеличение объемов систем хранения данных. Программа модернизации сети BASNET предусматривает подключение к общеевропейской научной сети GEANT на скоростях, которые обеспечили бы Беларуси полноценное участие в международных проектах [5], а также увеличение пропускной способности внешнего шлюза до 10 Гбит/с и создание каналов до 1 Гбит/с для связи с регионами.

В рамках межгосударственной программы инновационного развития стран СНГ планируется разработка технологий объединения разнородных грид-сетей в единую инфраструктуру, наращивание вычислительных мощностей (проект UHP-INSPUR) и внедрение суперкомпьютерных и грид-технологий в других странах, в первую очередь в СНГ.

РАЗВИТИЕ НАЦИОНАЛЬНЫХ ГРИД-СЕТЕЙ В ТАТАРСТАНЕ

Об истории, структуре и работе национальной грид-сети Республики Татарстан (РТ) рассказал директор казанского филиала МСЦ РАН, д.ф.-м.н. А.М.Елизаров.

Гражданская телекоммуникационная сеть (ГС) Республики Татарстан, созданная в 1994 году на базе Казанского государственного университета (КГУ, ныне – Казанский федеральный университет, КФУ) совместно с Казанским научным центром РАН (КазНЦ РАН) и рядом вузов Казани, была крупной телекоммуникационной системой регионального масштаба, объединявшей более 130 организаций.

Организационной и финансовой основой ГС был проект федеральной программы [6], реализованный на средства Министерства науки РФ, Российского фонда фундаментальных исследований (РФФИ) при финансовой поддержке Правительства РТ. Фактически ГС объединила

на общей технической базе компьютерные сети почти всех вузов Казани и Набережных Челнов, институтов КазНЦ РАН и АН РТ, республиканских больниц и клиник, организаций государственного управления, культуры и многих других некоммерческих организаций. До появления ГС в республике практически не было школ, библиотек, организаций здравоохранения, имевших доступ в глобальные сети.

В 2000 году была создана сеть SENet-Tatarstan, которая объединила институты КазНЦ РАН, АН РТ и вузы Казани. При финансовой поддержке РАН, РФФИ и Федеральной целевой программы (ФЦП) "Интеграция" в 2001 году был реализован проект перевода коммуникационной основы SENet на оптоволоконные соединения. Функционирование ГС и SENet было скоординировано.

В этом же году в КазНЦ РАН в рамках ФЦП "Интеграция" был запущен первый 10-процессорный вычислительный кластер на базе рабочих станций Compaq Alpha. Это положило начало внедрению высокопроизводительных вычислений в научные и прикладные исследования институтов КазНЦ РАН и вузов Казани. Для того чтобы создать наиболее благоприятные условия для проведения научных исследований с использованием современных методов квантовой химии и подготовки соответствующих специалистов в КГУ и Казанском государственном технологическом университете (ныне Национальный исследовательский технологический университет – КНИТУ), в 2001 году были организованы вычислительные кластеры на базе учебных классов. С использованием удаленного доступа по сети SENet-Tatarstan появилась возможность проведения расчетов на кластере КазНЦ РАН с высокой скоростью обмена данными.

В 2007 году был создан Суперкомпьютерный центр коллективного пользования КазНЦ РАН, основным вычислительным ресурсом которого стал вычислительный кластер MBC-15000 с девятью базовыми блоками IBM BladeCenter и двухпроцессорными платами JS20, коммуникационной средой Myrinet и Gigabit Ethernet с пиковой производительностью 2,25 TFlops. Соответствующее системное и прикладное ПО функционировало под управлением ОС Linux и обеспечивало полный технологический цикл: организацию и управление проведением массовых расчетов посредством системы очередей в многозадачном и многопользовательском режимах; организацию распределенных вычислений с использованием грид-технологий на базе Globus ToolKit

и ССРВ; обеспечение защищенного удаленного доступа пользователей к вычислительным ресурсам по сетям общего пользования; сбор данных об использовании вычислительных ресурсов; управление и мониторинг технического состояния компьютерного оборудования и инженерной инфраструктуры; информирование пользователей о подключении к кластеру и установленном ПО, технологиях запуска расчетных задач и т.д. В течение 2007–2009 годов ежегодно выполнялось около 10 тыс. расчетных заданий общей продолжительностью 16–19 тыс. процессорных дней. Основной объем расчетов приходился на квантовохимические расчеты [7], в которых использовалось свободно распространяемое и лицензионное ПО, в том числе пакеты Priroda, GAMESS, Gaussian. В 2009 году Казанский государственный технический университет (ныне Национальный исследовательский технический университет, КНИТУ-КАИ) запустил собственный вычислительный кластер пиковой производительностью до 2 TFlops.

КазНЦ РАН тесно сотрудничает с Межведомственным суперкомпьютерным центром (МСЦ) РАН по внедрению высокопроизводительных расчетов в научные исследования и развитию аппаратного комплекса. В 2009 году на базе КазНЦ РАН был создан Казанский филиал (КазФ) МСЦ РАН, который продолжил начатые работы, значительно расширил тематику суперкомпьютерных приложений и вычислительную базу научных исследований в Татарстане. В 2010 году в IT-парке РТ установлен суперкомпьютер МВС-100КФ "Сююмбике" с пиковой производительностью 10 TFlops, ориентированный на функционирование в составе распределенной инфраструктуры суперкомпьютерных приложений РФ. В этом же году КазФ МСЦ РАН начал эксплуатацию вычислительного кластера на базе Microsoft Windows Server 2008 HPC edition, предназначенного для обеспечения вычислительными ресурсами разработок в среде Microsoft Windows. К настоящему моменту суммарная пиковая производительность вычислительных ресурсов КазФ МСЦ РАН достигла 20 TFlops.

В перспективе в РТ планируется построение грид-инфраструктуры, интегрированной в общероссийскую грид-сеть. Планируется также создать на базе инфраструктуры сети SENet-Tatrstan грид-сеть SEGrid-Tatrstan, с помощью которой ее участники получают доступ к распределенным удаленным вычислительным ресурсам разного рода – от уникальных суперкомпьютеров

до средних и малых суперкомпьютерных кластеров и отдельных серверов.

Проект SEGrid получил поддержку Правительства Татарстана. В крупнейшем в Поволжье IT-парке, кроме суперкомпьютера "Сююмбике", будет размещен узел общероссийской инфраструктуры суперкомпьютерных приложений, работы по созданию которой ведутся в рамках деятельности Национальной ассоциации исследовательских и научно-образовательных электронных инфраструктур "е-Арена".

Понятно, что развивать грид-системы (как и прочие проекты подобного масштаба) невыгодно без весомого экономического и практического обоснования. Часть выступления А.Елизарова была посвящена именно возможностям использования суперкомпьютерных и грид-технологий на практике.

С помощью суперкомпьютерной сети РТ выполняются квантово-химическое изучение механизмов химических реакций, разработка параллельных алгоритмов и программ решения задач в области аэрогидродинамики и теории упругости, применение вычислительных систем с массовым параллелизмом при анализа естественного микросейсмического поля и при разведке залежей природных углеводородов, а также решение задач механики набухающих систем в медицинской физике и биомеханике.

Перспективное направление развития суперкомпьютерной сети РТ – присоединение республиканских суперкомпьютерных ресурсов к научно-образовательным грид-инфраструктурам, создаваемым в РФ. Одним из первых проектов в этом направлении, реализованных КазНЦ РАН в 2008 году, стало подключение имеющегося вычислительного комплекса к распределенной инфраструктуре для суперкомпьютерных приложений РАН [8], работы по созданию которой осуществлялись МСЦ РАН. Сегодня эта инфраструктура объединяет ряд вычислительных систем РАН в Москве, Казани и Санкт-Петербурге. С ее помощью пользователи получили доступ по привычному им интерфейсу к распределенным вычислительным ресурсам РАН, суммарная производительность которых превышает 100 TFlops.

КазФ МСЦ РАН принял участие в проекте "ГридННС" [9], который реализовывался в рамках Федеральной целевой программы "Развитие инфраструктуры наноиндустрии в Российской Федерации на 2008–2010 годы". Его целью было дать возможность географически распределенным

научным и инженерным коллективам – участникам Национальной нанотехнологической сети (ННС) эффективно использовать удаленные информационные, коммуникационные и вычислительные инфраструктуры.

Распределенные вычислительные структуры позволили проводить высокоточное компьютерное моделирование сложных объектов на уникальных по своей архитектуре суперкомпьютерах. В качестве примера можно привести исследование структуры пленок ультрананокристаллического алмаза (УНКА) в КазФ МСЦ РАН [10]. Использование программного пакета квантово-химических расчетов CPMD (Car-Parinello Molecular Dynamics) фирмы IBM и суперкомпьютера факультета ВМК МГУ позволило проанализировать стабильность УНКА, содержащих более 1500 атомов, с крайне высокой для систем такого размера точностью. Работа на менее производительных универсальных вычислительных кластерах, составляющих основную часть грид-систем, увеличила бы время расчетов в несколько раз. Еще один пример – изучение механизмов химических реакций [11, 12]. За счет использования высокопроизводительных вычислительных кластеров (и при достаточной квалификации исследователя) эти исследования были проведены за несколько недель, а могли бы растянуться на несколько лет.

Суперкомпьютеры применяются и в нефтедобывающей отрасли. Сегодня это обусловлено в первую очередь тем, что большие месторождения РТ уже длительное время находятся в эксплуатации, добыча нефти из них только по традиционным технологиям падает, а расходы на амортизацию инфраструктуры растут. Применение же новых технологий сопровождается (особенно в условиях отраслевого и бюрократического лоббирования) большой долей неопределенности и, следовательно, риска при использовании новых методов непосредственно на месторождении. Лабораторная экспертиза при этом зачастую не может быть расценена как достоверная из-за невозможности перенести результаты лабораторных экспериментов на реальный геологический объект. Положительную роль здесь может сыграть трехмерное моделирование процессов добычи нефти, опирающееся на созданные теоретические модели многофазной многокомпонентной фильтрации и реализуемое на современных высокопроизводительных компьютерных системах. Его использование позволяет сократить число дорогостоящих натурных испытаний нового оборудования и тем самым поднять рентабельность

отрасли в целом. Также суперкомпьютеры широко применяются при интерпретации данных трехмерной сейсмической разведки и других методов волнового исследования структуры земных недр. Детальное моделирование расположения и структуры продуктивных пластов с использованием натуральных данных позволит минимизировать затраты на разбуривание мелких месторождений, которые ранее относились к неперспективным, и начать их эксплуатацию.

СОЗДАНИЕ РОССИЙСКОГО НАЦИОНАЛЬНОГО СТАНДАРТА ГРИД-СЕТЕЙ

По определению, грид-сеть – это согласованная, открытая и стандартизированная компьютерная среда, обеспечивающая гибкое, безопасное, скоординированное разделение вычислительных и информационных ресурсов (частей этой сети) в рамках одной виртуальной организации. Разработкой стандартов грид-сетей за рубежом занимаются национальные организации отдельных стран. Эти организации (более 400 из 50 стран) объединяет Международная общественная организация OpenGridForum – OGF (www.ogf.org). В настоящее время OGF разработала почти 200 стандартов.

Ясно, что национальная российская грид-сеть должна строиться на национальных же стандартах, т.е. на ГОСТ. В федеральном законе "О техническом регулировании" сказано (ст.13): "К используемым на территории РФ документам относятся национальные стандарты". В списке документов далее указаны "надлежащим образом заверенные переводы на русский язык международных стандартов, принятые на учет Росстандартом". Использование же англоязычных стандартов, не прошедших хотя бы регистрации, не оговорено, а следовательно, незаконно.

Однако в "требованиях по стандартизации" конкурсной документации указано: "Применить документ WSRF BP 1.1". В более фундаментальных документах (направления работ Отделения нанотехнологии и информационных технологий РАН, программа Президиума РАН №14, программа фундаментальных исследований государственных академий, положение о Совете РАН "Высокопроизводительные вычислительные системы, научные телекоммуникации и информационная инфраструктура") сказано именно о национальном стандарте. О том, как разрабатывается национальный стандарт грид-сетей и с какими проблемами пришлось столкнуться, рассказал А.Я.Олейников, гл. научн. сотрудник Института радиотехники и электроники им. В.А.Котельникова РАН (Москва). Этот институт непосредственно занимается разработкой стандартов грид-сетей.

Понятно, что российский национальный стандарт не должен противоречить существующим. Создание же нового стандарта, даже на основе уже имеющихся, может занять несколько лет. Одно лишь обоснование того, какие стандарты и в каком порядке нужно разрабатывать – очень трудоемкий процесс.

В первую очередь было решено принять в качестве стандарта язык JSDL (Job Submission Description Language). Его использование позволило бы унифицировать создание и применение грид-систем и, таким образом, сэкономить время и ресурсы. Это предложение было распространено в Интернете, о нем рассказывалось на конференции в Дубне еще в 2010 году, однако мировое сообщество разработчиков грид-систем отнеслось к этой идее достаточно пассивно. Тем не менее, эта разработка была включена в план государственной стандартизации 2011 года. Первая редакция стандарта должна проходить общественное обсуждение, но и в этом случае реакция сообщества практически отсутствовала. "Единственной организацией, которая без всякой инициативы с нашей стороны дала отзыв на первую редакцию, был Институт прикладной математики им. М.С.Келдыша, активно работающий в области грид-технологий, – рассказывал А.Я.Олейников. – Интересной была реакция Лаборатории информационных технологий Объединенного института ядерных исследований (ОИЯИ), которая взаимодействует непосредственно с OGF, она не поддержала нашу разработку".

Несмотря на отсутствие интереса со стороны сообщества, создание национального

стандарта началось. С 2013 года вступает в действие ГОСТ Р 55022-2012 "Информационная технология. Спецификация языка описания представления задач (JSDL). Версия 1.0". Завершена первая редакция ГОСТ Р "Модель открытой грид-среды" на основе документа OGF GFD 0-80 I – его роль аналогична роли стандарта ISO 7498 на семиуровневую модель взаимосвязи открытых систем. В терминах этой модели можно строить профили стандартов. Подана заявка на третий стандарт "Грид-технологии. Термины и определения". Это тоже достаточно важный момент, так как терминология в этой области еще не устоялась, например, не существует даже общепринятого определения грид-сетей. Далее будет начата разработка базового профиля грид-среды, включающего семь стандартов. Это потребует гораздо больших ресурсов, как человеческих, так и финансовых, поэтому очень важно заинтересовать грид-сообщество, чтобы объединить усилия. Как отдельный стандарт оформлен единый подход к обеспечению интероперабельности систем разных классов, который позволит вести разработки грид-систем более последовательно и эффективно [13, 14].

Таким образом, идея национального стандарта грид-сетей начинает постепенно воплощаться в реальность. Однако остается открытым один, самый главный вопрос – стоит ли тратить силы и ресурсы на создание собственного стандарта грид-сетей, вместо того чтобы развивать сами сети?

* * *

Несмотря на все бедствия, которые переживала наша электронная промышленность и исследовательские институты с момента распада СССР, российская суперкомпьютерная отрасль развивается, предлагая порой вполне конкурентоспособные решения. Конечно, речь не идет об элементной базе, так как отставание России в этой области значительно, и для высокопроизводительных вычислений применяются и будут применяться компоненты иностранных производителей. Первой и пока что единственной "ласточкой" обновленной отечественной электронной промышленности могли бы стать микропроцессоры "Эльбрус", которые, однако, позиционируются скорее как нишевые изделия для специальных применений, а не как массовый продукт. Тем не менее, российские разработчики способны создавать высокопроизводительные вычислительные комплексы на базе иностранных компонентов

и узлов и вполне могут конкурировать с зарубежными фирмами – интеграторами и поставщиками готовых решений.

Для пользователей же суперкомпьютерных систем – а это могут быть и научные работники, и инженеры, и самые разные коммерческие и государственные предприятия и организации – грид-технологии практически уничтожают границы, позволяя удаленно использовать мощнейшие вычислительные кластеры.

Иными словами, перспективы у российских суперкомпьютеров есть. Ясно, что просто создать вычислительную систему недостаточно, нужно ее показать так, чтобы потенциальные покупатели увидели ее и захотели приобрести. Увы, умение продавать – не самая сильная сторона российского производителя. Национальный суперкомпьютерный форум собрал (и, хочется надеяться, будет собирать ежегодно) множество специалистов, коммерческих представителей и просто интересующихся суперкомпьютерами со всей страны и из-за рубежа. Именно форум может стать тем самым инструментом, с помощью которого российские предприятия будут демонстрировать и продавать свои разработки – как концепты, так и серийные образцы – и, кроме этого, обмениваться опытом и знаниями.

ЛИТЕРАТУРА

1. **А.Ким, В.Фельдман.** СуперЭВМ на основе архитектурной платформы "Эльбрус". – Электроника:НТБ, 2009, №2, с.75–80.
2. **Левин И.И.** Реконфигурируемые вычислительные системы с открытой масштабируемой архитектурой. Труды пятой международной конференции "Параллельные вычисления и задачи управления" РАСО'2010. – М.: Учреждение Российской академии наук, Институт проблем управления им. В.А.Трапезникова РАН, 2010, с.83–95.
3. **Каляев А.В., Левин И.И.** Модульно-наращиваемые многопроцессорные системы со структурно-процедурной организацией вычислений. – М.: Янус-К, 2003.
4. **Абрамов С.М., Анищенко В.В., Заднепровский В.Ф., Криштофик А.М.** Основные результаты и эффективность выполнения программы Союзного государства "СКИФ-ГРИД". Четвертая международная научная конференция "Суперкомпьютерные системы и их применение" (SSA'2012): доклады конференции (23–25 октября 2012 года, Минск). – Минск: ОИПИ НАН Беларуси, 2012, с.7–11.
5. **Анейчик С.А., Костюкевич Ю.В., Нозик В.М.** Академсеть BASNET как высокопроизводительная транспортная среда для реализации суперкомпьютерных технологий. Четвертая международная научная конференция "Суперкомпьютерные системы и их применение" (SSA'2012): доклады конференции (23–25 октября 2012 года, Минск). – Минск: ОИПИ НАН Беларуси, 2012, с.113–115.
6. Межведомственная программа создания национальной сети компьютерных телекоммуникаций для науки и высшей школы. 1995–1998. – М.: Министерство науки и технической политики России, 1995.
7. **Чачков Д.В., Астафьев М.Н., Биктимиров М.Р., Туриянский Е.А., Шамов Г.А., Шамов А.Г., Храпковский Г.М.** Квантово-химические расчеты термического разложения нитросоединений на базе вычислительного кластера Казанского научного центра РАН. Восьмая международная научная конференция "Высокопроизводительные параллельные вычисления на кластерных системах" (НРС-2008). Сборник трудов. – Казань, 2008, с.341–346.
8. **Савин Г.И., Шабанов Б.М., Корнеев В.В., Телегин П.Н., Семенов Д.В., Киселев А.В., Кузнецов А.В., Вдовикин О.И., Аладышев О.С., Овсянников А.П.** Создание распределенной инфраструктуры для суперкомпьютерных приложений. – Программные продукты и системы, 2008, №2, с.2–7.
9. <http://www.ngrid.ru>
10. **Denis Tarasov, Ekaterina Izotova, Diana Alisheva, Natalia Akberova, Robert A. Freitas Jr.** Structural stability of clean, passivated, and partially dehydrogenated cuboid and octahedral nanodiamonds up to 2 Nanometers in Size. – J. Comput. Theor. Nanosci., 2011, т.8, с.147–167.
11. **Alexander G. Shamov, Grigorii M. Khrapkovskii.** A theoretical study of the gas-phase pyrolysis of nitroethylene. – Mendeleev's Communications, 2001, №4, с.163–164.
12. **Шамов А.Г., Николаева Е.В., Храпковский Г.М.** Теоретическое изучение механизма газофазного распада нитроэтилена. – Журнал общей химии, 2004, т.74, вып.8, с.1327–1342.
13. **Гуляев Ю.В., Журавлев Е.Е., Олейников А.Я.** Методология стандартизации для обеспечения интероперабельности информационных систем широкого класса. Аналитический обзор. – Журнал радиоэлектроники, 2012, №3 (jre.cplire.ru/jre/Mar/12/2/text/pdf).
14. **Батоврин В.К., Гуляев Ю.В., Олейников А.Я.** Обеспечение интероперабельности – основная тенденция в развитии открытых систем. – Информационные технологии и вычислительные системы, 2009, №5, с.7.