



Копия текста публикации со страницы

<http://www.cio-world.ru/products/infrastructure/407719/>

Да, СКИФы – мы!

Автор: **Дмитрий Шульгин**

Опубликовано 06 марта 2009 года

Когда мы произносим слово «суперкомпьютер» (мэйнфрейм), то в голове моментально возникает соответствующий ассоциативный ряд: IBM, Cray, SGI, Hitachi, Fujitsu, Sun Microsystems, Roadrunner, Blue Gene и т.д. И мало кто из нас сразу же вспоминает о том, что в данной области у России имеются и собственные разработки, которые выполняются по программе «СКИФ». У этой линейки суперкомпьютеров на сегодня есть определенные достижения: пять машин «СКИФ» входят в так называемый TOP500, причем, самый производительный кластер, установленный в МГУ, занимает 54-ю позицию в мире (по состоянию на ноябрь 2008 года), показав в тесте Linpack производительность 47,17 TFLOPS (пиковая – 60 TFLOPS).

Ряд	Годы и пиковая производительность (расчетный диапазон)	Ядер в CPU/разрядность	Сетевые решения вспомогательной / системной сети	Форм-фактор; CPU/U	Примечание
1	2000–2003 0.020–0.5 TFlops	1 / 32	FastEthernet / SCI (2D-top). Myrinet	4U–1U; 0.5–2 CPU/U	Отечественный SCI (2D-top). Охлаждение: воздух
2	2003–2007 0.1–5 Tflops	1 / 32–64	GB Ethernet / SCI (3D-top), Infiniband	1U, Hyper-Blade; 2 CPU/U	ServNet v.1, v.2 Ускорители: FPGA, OBC. Охлаждение: воздух
3	2007–2008 5–150 Tflops	2–4 / 64	GbEthernet / Infiniband DDR	1U, лезвия (20 CPU в 5U); 2–4 CPU/U	ServNet v.3. Охлаждение: воздух—вода— фреон
4	2009–2012 500–10 000 Tflops	4–8 / 64	Infiniband QDR / отечественная системная сеть (3D-top)	Сверхкомпактные лезвия; более 10 CPU/U	Новые подходы к охлаждению. Ускорители: ПЛИС, МЦОС, GPU и др.

Первая программа «СКИФ» выполнялась в 2000-2004 гг. В 2007-м стартовала программа «СКИФ-ГРИД», которая запланировала до 2010 года. Главными исполнителями проекта являются: со стороны Беларуси – Объединенный институт проблем информатики Национальной академии наук Беларуси; со стороны России – Институт программных систем РАН. Мэйнфреймы семейства «СКИФ» разработчики делят на так называемые ряды (поколения). Суперкомпьютеры ряда 1 выпускались в 2000-2003 гг.: на них использовались еще 32-разрядные одноядерные процессоры, для системной сети применялись SCI (2D-top) и Myrinet, в качестве вспомогательной сети использовался FastEthernet. Расчетный диапазон пиковой производительности этих машин составлял

0,020-0,5 TFLOPS. Мэйнфреймы ряда 2 относятся к 2003-2007 гг.: на них уже ставили не только 32-х, но и 64-х разрядные процессоры, в качестве системной сети использовались SCI (3D-top) и Infiniband, а в качестве вспомогательной сети – Gigabit-Ethernet. Производительность второго ряда уже составляла 0,1-5 TFLOPS. Машины третьего ряда, появившиеся в 2007-2008 гг., делались на 2-4-х ядерных 64-х разрядных процессорах, в качестве системной сети использовалась Infiniband DDR, а в качестве вспомогательной сети – Gigabit-Ethernet. И хотя лучшая из машин данного класса (кластер МГУ) имеет пиковую производительность 60 TFLOPS, технология третьего ряда позволяет создавать суперкомпьютеры производительностью 5-150 TFLOPS.



Далее, в соответствии с принятой программой развития до 2012 года, запланирован выпуск машин четвертого ряда, которые будут иметь производительность от 200 TFLOPS до 10 PFLOPS. В вычислительных узлах использованы многоядерные (4-8 ядер и выше) 64-х битные процессоры архитектуры x86, применяется память объемом 6, 12 или 24 Гбайт (впаяна в плату) и SSD-диски для хранения образа операционной системы, вспомогательных файлов, файлов контрольных точек и раздела для организации виртуальной памяти. В

качестве системной сети используется отечественная 3D-top на базе FPGA, а в качестве вспомогательной сети – Infiniband QDR или 10-гигабитный Ethernet. Разработчики достигли более высокой плотности упаковки вычислительной мощности: до 10 процессоров в модуль форм-фактора 1Unit. С повышением плотности упаковки уменьшаются физические размеры установки, длины соединительных линий между узлами, сокращаются задержки, происходит существенное снижение количества соединительных кабелей и кабельных разъемов в системе, за счет чего улучшается надежность. Однако все это, в свою очередь, потребовало применения новых подходов к проблеме охлаждения вычислительной установки: в машинах ряда 4 применяется уникальная система непосредственного водяного охлаждения печатных плат и микросхем. Данная система требует в 2 раза меньше энергозатрат. В случае аварийной остановки циркуляции теплоносителя, вода в течение некоторого времени сохраняет способность охлаждать микросхемы. Кроме того, система водяного охлаждения не содержит ни одной механической подвижной части, что также повышает общую надежность установки и заметно улучшает ее эргономические качества – бесшумность работы узлов.

Производители суперкомпьютеров «СКИФ» ряда 4 предусмотрели трех варианта своих изделий, предназначенные для различной целевой аудитории. Первый из них представляет собой, как называют его авторы, «мобильный суперкомпьютер». Это одно шасси, которое можно расположить на рабочем месте сотрудника, поскольку, как уже упоминалось выше, водяная система охлаждения делает подобное изделие, практически, бесшумным, к тому же, шасси имеет вполне приемлемое электропотребление. Пиковая производительность такого вычислителя не будет превышать 3 TFLOPS. Собственно говоря, данное шасси является строительным блоком для более крупных систем: шкаф, система из нескольких шкафов. Все шасси закрываются сенсорным жидкокристаллическим экраном, с помощью которого можно управлять системой.

Для крупных лабораторий, конструкторских отделов и региональных ВУЗов можно использовать один шкаф или конструкцию из нескольких шкафов, каждый из которых содержит от двух до восьми шасси и всю необходимую соединительную инфраструктуру: соединения системной сети, вспомогательной сети, сервисной сети, подсистем электропитания и охлаждения. Пиковая производительность такого решения может

составлять от шести до 350 TFLOPS. И наконец, для крупных суперкомпьютерных национальных центров предлагается система из 15 и более шкафов, объединенных общей инфраструктурой. Пиковая производительность самого мощного вычислителя может достигать 10 PFLOPS. А это уже тот уровень производительности, который позволит войти отечественному суперкомпьютеру в первую десятку.

Все строительные блоки, все решения суперкомпьютеров «СКИФ» ряда 4, программное обеспечение – все это будет доступно отечественным компаниям по партнерской программе. Вводится два статуса: «партнер технологии «СКИФ-ГРИД» – это компании, которые занимаются постройкой своих собственных программно-аппаратных комплексов на базе новых решений, предлагаемых объединением «СКИФ»; «провайдер товаров и услуг» – это фирмы, получающие уже готовые изделия. Разработчики отечественных суперкомпьютеров ряда 4 оценивают общий объем российского рынка в сумму около 500 млн. долларов. Постройка одной машины нового класса с пиковой производительностью 0,5 PFLOPS обойдется примерно в 50-70 млн. долларов. Что же касается совокупных затрат России и Белоруссии на разработку «СКИФ-ГРИД» в 2007-2010 гг., то они оцениваются в сумму, превышающую 1 млрд. рублей.

Однако по отношению к отечественным суперкомпьютерам можно задать примерно тот же каверзный вопрос, который обычно задают относительно иномарок российской сборки: «Чем вы хвастаетесь? Что здесь российское, кроме самой сборки?» На этот тяжелый вопрос со стороны разработчиков машин «СКИФ» ответил директор Института программных систем РАН Сергей Михайлович Абрамов: «Есть такое понятие – внешняя экспертиза, которая никогда не подвергала сомнению тот факт, что суперкомпьютеры серии «СКИФ» являются оригинальной разработкой российских ученых и инженеров. Вся конструкторская документация на изделие в целом принадлежит России и Белоруссии. Машины разрабатываются и собираются нами, собственность подтверждена патентами. КПД машин «СКИФ», как правило, выше, чем у систем зарубежного происхождения. И еще один важный момент. Когда вы создаете машину сами, нет никаких ограничений ни по области ее применения, ни по мощности создаваемых кластеров. Если в машинах «СКИФ» ряда 1 отечественными были, по сути дела, только сами шкафы, то в новых суперкомпьютерах ряда 4 отечественным является все, за исключением микросхем на плате».



Что же, можно было бы искренне порадоваться за отечественного производителя, если бы не это небольшое исключение, ведь, под невзрачным термином «микросхемы на плате» господин Абрамов скрыл одну важную деталь – процессоры. Что бы там ни говорили разработчики «СКИФ», сейчас отечественные мэйнфреймы, например, известный кластер в МГУ, работают на четырехъядерных процессорах Intel Xeon E5472 3,0 ГГц. И без этой американской начинки российские машины никогда не