ПЕРВЫЕ ТЕХНИЧЕСКИЕ РЕШЕНИЯ СУПЕРКОМПЬЮТЕРНОГО НАПРАВЛЕНИЯ «СКИФ» РЯДА 4

С.М. Абрамов, В.В. Анищенко, А.М. Криштофик, Н.Н. Парамонов

Создание и освоение суперкомпьютерных средств и технологий было главной задачей программы Союзного государства «СКИФ» (2000–2004 гг.). Результаты комплексной реализации программы «СКИФ» являются существенным научно-техническим и организационным заделом для дальнейшего развития суперкомпьютерного направления в рамках новых программ Союзного государства «Триада» (2005–2008 г.г.) и «СКИФ-ГРИД» (2007–2010 гг.).

Важнейший результат выполнения программы «СКИФ» — выпуск образцов кластерных конфигураций «СКИФ» Ряда 1 и Ряда 2 производительностью в диапазоне от десятков миллиардов до нескольких триллионов операций в секунду, которые использовались как для отработки программного обеспечения, так и для реальных вычислений в интересах предприятий и учреждений России и Беларуси. В 2004 году была создана старшая модель семейства «СКИФ» — суперЭВМ «СКИФ К-1000» с пиковой производительностью 2,5 триллиона операций в секунду, который 9 ноября 2004 года включен в очередной 24-й выпуск списка top-500 под номером 98. Списки Top500 самых мощных вычислительных систем в мире публикуются два раза в год, начиная с 1993 года. Суперкомпьютер «СКИФ К-1000» входил в четыре подряд редакции списка Top500, что убедительно свидетельствует о перспективности принятых технических решений в рамках программы «СКИФ».

В процессе выполнения программы «Триада» созданы модели типовых персональных кластеров семейства «СКИФ-Триада» с широким спектром производительности — от 50 до 500 миллиардов операций в секунду, базирующиеся на основополагающих концептуальных принципах создания суперкомпьютеров направления «СКИФ». Тенденции развития информационных технологий диктуют необходимость широкого внедрения принципов параллельных вычислений для решения высокопроизводительных задач. Для этого требуется создание недорогих малогабаритных вычислительных комплексов, которые возможно устанавливать в обычных рабочих помещениях или офисах. Модели персональных кластеров — это небольшие экономичные полнофункциональные вычислительные комплексы с модульной кластерной архитектурой. Персональные кластеры «СКИФ-Триада» заполняют нишу, разделяющую обычные персональные компьютеры и суперкомпьютеры, обеспечивая возможность использования суперкомпьютерных технологий в отдельной организации, подразделении для персональных вычислений.

Одним из основных направлений программы «СКИФ-ГРИД» является суперкомпьютерное направление, главной целью которого является создание опытных образцов суперкомпьютеров «СКИФ» следующего поколения (Ряд 3 и Ряд 4), ориентированных на использование в грид-системах.

Основные усилия по разработке отечественных технических решений и выпуску опытных моделей суперкомпьютеров семейства «СКИФ» в рамках программы «СКИФ-ГРИД» можно разделить на следующие этапы:

- 2007–2008 годы: создание кластеров Ряда 3 — «СКИФ МГУ» (новый флагманский суперкомпьютер направления «СКИФ») с пиковой производительностью 60 Tflops, «СКИФ УРАЛ» с пиковой производительностью 16 Tflops и «СКИФ К-1000М (модернизация суперкомпьютера «СКИФ К-1000») с пиковой производительностью 5 Tflops;
- 2007–2010 годы: разработка отечественных технических решений для перспективных суперкомпьютеров семейства «СКИФ» (blade-серверные решения, отечественный интерконнект, решения в области ускорителей и т.п.);
- 2009–2012 годы: создание модулей и опытных образцов базовых конфигураций суперкомпьютерных систем семейства «СКИФ» Ряда 4 для отработки решений на перспективу (радикальное улучшение показателя «производительность на ватт», использование гибридных вычислительных узлов и новых архитектурных решений).
- Развитие суперкомпьютерного направления «СКИФ» отражено на рисунке 1.

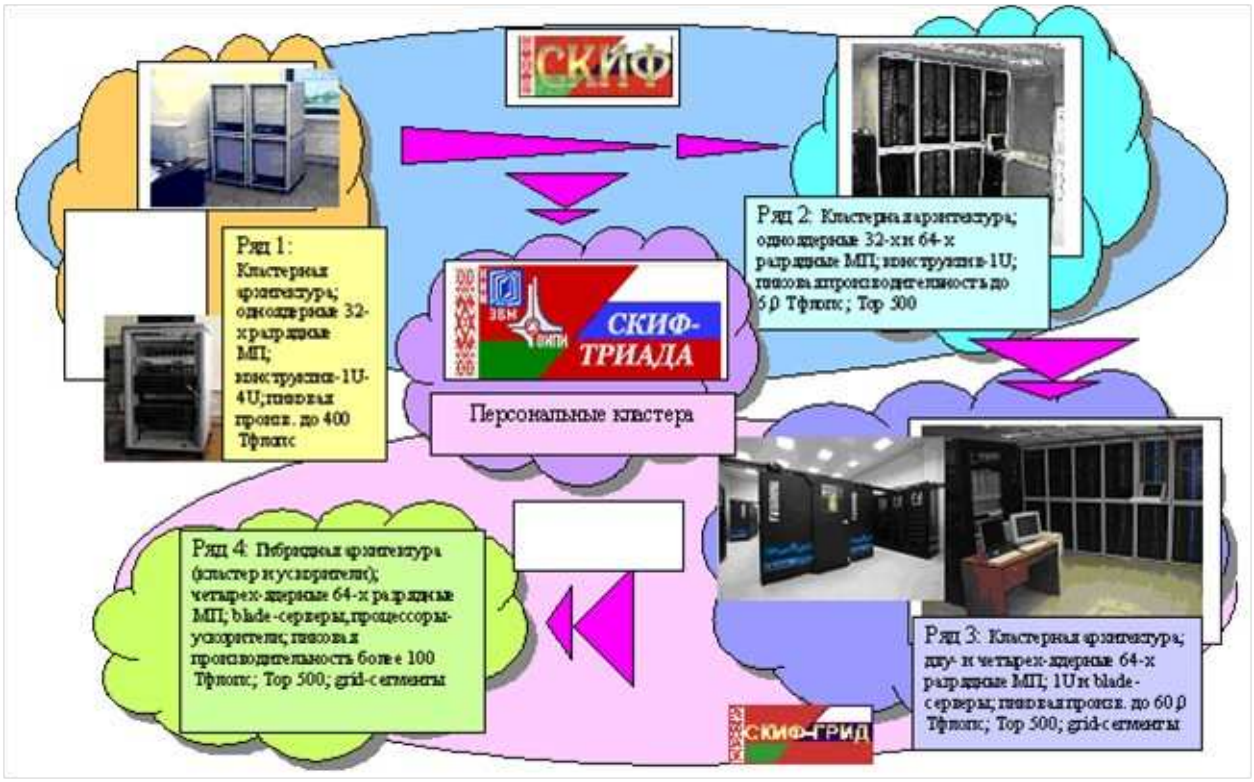

Рис. 1. Развитие суперкомпьютерного направления «СКИФ»

При создания суперкомпьютеров «СКИФ» Ряда 4 заложены технические решения, ориентированные на создание в Союзном государстве отечественных суперкомпьютерных технологий мирового уровня. Основные стратегические принципы создания перспективных конфигураций «СКИФ» Ряда 4:

- Гибридная архитектура вычислительных узлов кластера;
- Мультиядерные процессоры с архитектурой x86;
- Достижение петафлопсного диапазона (5–10 Pflops в 2012 г.);
- Водяное охлаждение на серверном конструктивно-технологическом уровне;
- Отечественные блейд-серверы;
- Отечественный высокоскоростной интерконнект.

Основные архитектурные и конструктивно-технологические принципы создания перспективных модулей, программного обеспечения и опытных образцов суперкомпьютеров «СКИФ» Ряда 4 изложены в работе [1]. В настоящем материале приведены основные характеристики суперкомпьютера «СКИФ ОИПИ» с гибридной архитектурой — «первенца» направления «СКИФ» Ряда 4.

Опытный образец суперкомпьютера «СКИФ ОИПИ» предназначен для установки и эксплуатации в суперкомпьютерном центре ОИПИ НАН Беларуси. Система кондиционирования этого вычислительного центра выполнялась по классической схеме. Кондиционеры, установленные в вычислительном зале, рассчитывались исходя из общей потребляемой всем планируемым оборудованием мощности. С учетом этого конструктивно-технологические решения, принятые для суперкомпьютера «СКИФ ОИПИ», рассчитаны на систему воздушного охлаждения, предусмотренную в суперкомпьютерном центре. Поэтому в каждой стойке этого изделия устанавливается не более двух шасси с блейд-серверами.

Структурно опытный образец суперкомпьютера «СКИФ ОИПИ» представляет собой метакластер, состоящий из двух кластеров:

- кластер на основе blade-технологий — blade-кластер;
- кластер на основе Cell-технологий — Cell-кластер.

Архитектурно blade-кластер и Cell-кластер объединяются в иерархическую гибридную кластерную конфигурацию — кластер высшего уровня или метакластер (metacluster). Метакластерная гибридная архитектура опытного образца суперкомпьютера «СКИФ ОИПИ» приведена на рисунке.2.

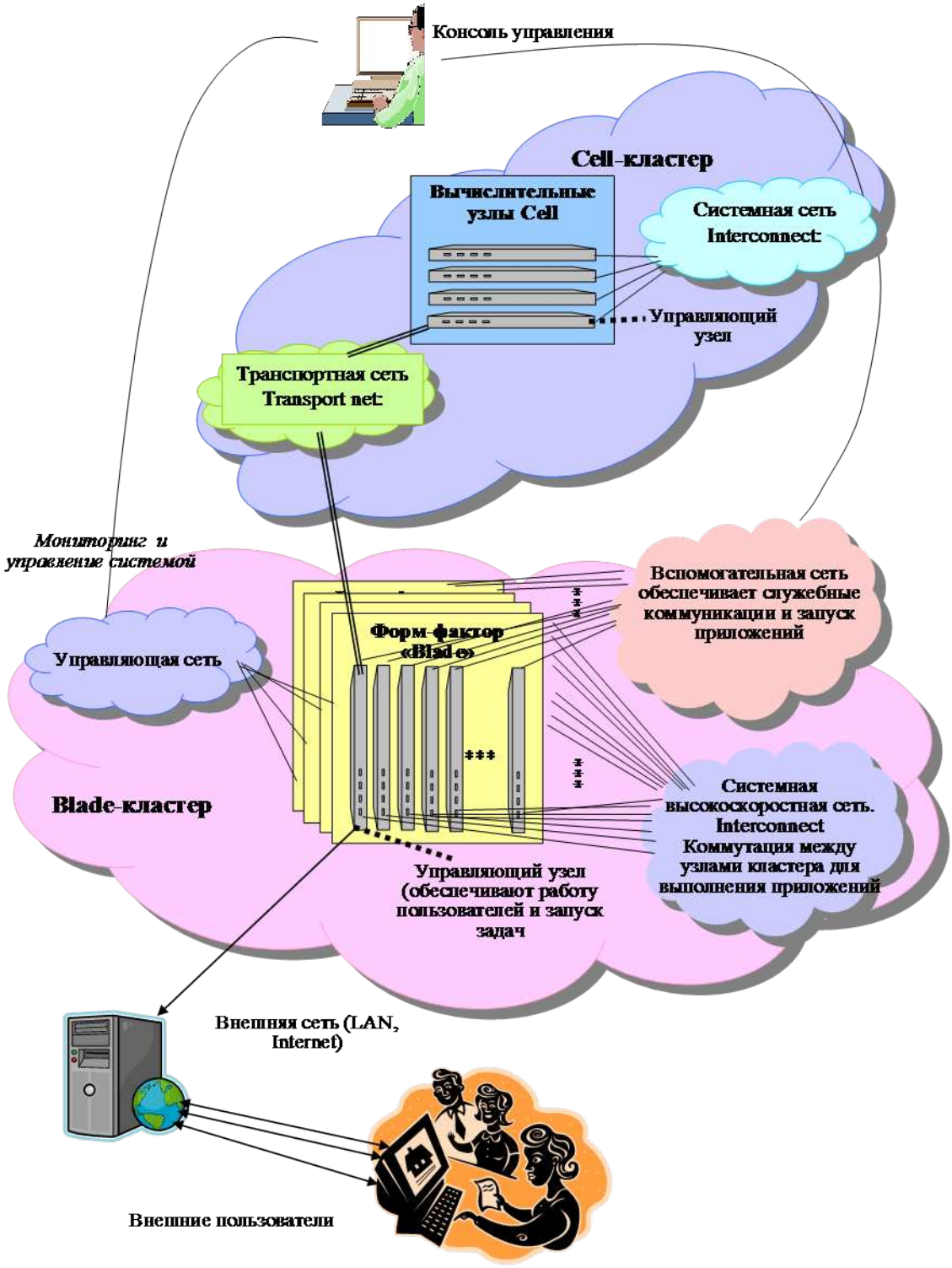

Рис. 2. Метакластерная гибридная архитектура опытного образца суперкомпьютера «СКИФ ОИПИ»

Метакластерный принцип позволяет создавать распределенные метакластерные конфигурации на базе локальных или глобальных сетей передачи данных. При этом, естественно, уменьшается степень связности подкластеров метакластерной конфигурации.

Системное программное обеспечение метакластера должно обеспечивать возможность реализации гетерогенных систем, включающих подкластеры различной архитектуры на различных программно-аппаратных платформах.

При реализации blade-кластера приняты следующие базовые технические решения:

- кластерная архитектура;
- операционная система Linux;

- системная сеть на базе сетевой технологии InfiniBand;
- вспомогательная сеть на базе сетевой технологии Gigabit Ethernet;
- управляющая (сервисная) сеть ServNet версии 2;
- основные вычислительные узлы на базе blade-серверов T-blades, ранее созданных в рамках разработки «СКИФ МГУ».

Основные вычислительные узлы имеют следующие характеристики:

- архитектура процессора: x86-64;
- модель процессора: Intel XEON E5472 3.0GHz;
- количество основных вычислительных узлов: 50;
- количество процессоров в основных вычислительных и в управляющих узлах: 2;
- общее количество процессоров в основных вычислительных узлах: 100;
- количество ядер в процессоре: 4;
- общее количество вычислительных ядер: 400;
- объем оперативной памяти каждого из 50 основных вычислительных узлов: 8,0 GB (8 x 1GB FB-DIMM ECC reg 667MHz);
- объем дисковой памяти (HDD) каждого основного вычислительного узла: — 1 x 160GB SATA 8MB 7200rpm;
- SUSE Linux Enterprise Server 10.0.

Объем оперативной памяти дополнительного вычислительного узла в конструктиве 2U: 64,0 GB (16x4 GB FB — DIMM ECC reg 667 MHz); дисковая память (HDD) — RAID: 5 x HDD 500GB SATA 16MB 7200rpm.

Управляющие узлы (конструктив 2U, 2 шт.) имеют следующие характеристики:

- CPU: 2 x Intel Xeon E5472 (8 x 3.0Ghz cores, cache L2: 12MB; FSB: 1600 MHz);
- RAM: 32GB (8 x 4GB FB-DIMM ECC reg 667MHz);
- RAID: 5 x HDD 500GB SATA 16MB 7200rpm.
- В состав опытного образца суперкомпьютера «СКИФ ОИПИ» также входят:
- комплект системной сети Infiniband 4x DDR (коммутатор Infiniband 4x DDR на 48 портов, с возможностью расширения до 144; комплект кабелей);
- комплект вспомогательной сети Gigabit Ethernet (стекируемый коммутатор Gigabit Ethernet,48 портов; комплект кабелей Patch cord3);
- система бесперебойного электропитания;
- комплект распределителей питания в стойках.

Пиковая производительность blade-кластера опытного образца суперкомпьютера «СКИФ ОИПИ»: 5,088 Тфлопс;

Планируемая реальная производительность blade-кластера опытного образца суперкомпьютера «СКИФ ОИПИ» на тесте Linpak: 3, 816 Тфлопс при показателе эффективности кластера Ueff = 75%.

Отличительной чертой опытного образца суперкомпьютера «СКИФ ОИПИ» является реализация перспективной гибридной кластерной архитектуры, в которой наряду с классическим кластером используются специализированные процессоры Cell. В частности, возглавляющий 31-ю редакцию списка Top 500 (июнь 2008 г.) суперкомпьютер фирмы IBM «Roadrunner», реализован на blade-технологиях (классический кластер) и процессорах Cell. Реальная производительность этой суперЭВМ — 1026 Тфлопс.

Процессор Cell использует архитектуру на базе RISC, имеет одно центральное ядро и восемь вычислительных ядер и обеспечивает высокую производительность в вычислениях с плавающей запятой. Центральное ядро представляет собой RISC-процессор, оптимизированный для высокопроизводительных вычислений. Он имеет 128 регистров общего назначения по 128 бит каждый. Содержимое одного регистра может трактоваться как набор целых чисел по 8, 16, 32 или 64 бита в каждом, четверка чисел с одинарной точностью или два числа с двойной точностью. Система команд поддерживает работу как с одним элементом, так и со всеми элементами регистра одновременно. Помимо арифметических и логических команд, имеются быстрые команды вычисления элементарных функций. RISC-процессор может выдавать по 2 команды за такт, таким образом, его производительность SPU на вычислениях с одинарной точностью на обычной для Cell частоте 3.2 ГГц составляет 25.6 Гфлопс. Скорость работы на командах двойной точности в 2 раза ниже. Все элементы процессора Cell соединены внутренней шиной (Element Interconnection Bus, EIB). Максимальная пропускная способность EIB составляет 204 ГБ/сек.

При реализации Cell-кластера в суперкомпьютере «СКИФ ОИПИ» приняты следующие базовые технические решения:

- количество вычислительных узлов — 8 (в том числе 1 узел выполняет вычислительные и управляющие функции);
- количество процессоров в вычислительных узлах: 2;
- общее количество процессоров в вычислительных узлах: 16;

- вычислительный узел: 2 x CPU PowerXCell 8i 3.2Ghz; RAM: 8GB (8x1 Gb); HDD:2 x SATA 250GB; Rails; 1U;
- пиковая производительность Cell-кластера: Rpeak= 1638 Гфлопс (double); Rpeak= 3276 Гфлопс (single);
- Interconnect: Gigabit Ethernet;
- Transport net: Gigabit Ethernet.

Суммарная пиковая производительность опытного образца суперкомпьютера «СКИФ ОИПИ» может достигать 8,364 Тфлопс.

Компактность конфигурации (3 стойки 19” высотой 42U) и воздушное охлаждение при суммарной пиковой производительности около 8 Тфлопс позволяют использовать опытный образец суперкомпьютера «СКИФ ОИПИ» в качестве замены суперкомпьютерным конфигурациям «СКИФ К-500» и «СКИФ К-1000М» для развития грид-технологий и решения широкого класса прикладных задач в рамках программы Союзного государства «СКИФ-ГРИД».

ЛИТЕРАТУРА:

1. С. М. Абрамов, В. Ф. Заднепровский, А. А. Московский, А. Б. Шмелев Суперкомпьютеры Ряда 4 семейства «СКИФ» // В сборнике трудов Международной конференции «Программные системы: теория и приложения», Переславль-Залесский, ИПС РАН, май, 2009