

МОДЕЛИ СУПЕРКОМПЬЮТЕРОВ СКИФ РЯДА-2

С.В. Абламейко¹, С.М. Абрамов², В.В. Анищенко¹, Н.Н. Парамонов¹, О.П. Чиж¹

¹Объединенный институт проблем информатики
Национальной академии наук Беларуси, Минск

²Институт программных систем Российской академии наук,
Переславль-Залесский

Рассматриваются конфигурации старших моделей суперкомпьютеров семейства СКИФ. Приведены их основные технические характеристики и состав программного обеспечения.

Введение

Суперкомпьютерные конфигурации СКИФ создаются в рамках программы Союзного государства СКИФ. С учетом продления срока реализации программа рассчитана на период в пять лет и включает два этапа.

На этапе 1 (2000-2002 гг.) отработаны основные концептуальные принципы и созданы суперкомпьютерные конфигурации, обеспечивающие возможность создания семейства моделей суперкомпьютеров среднего класса (модели Ряда 1) с пиковой производительностью до 300-400 миллиардов операций в секунду.

На этапе 2 (2003-2004 гг.) отрабатываются основные концептуальные принципы и создаются суперкомпьютерные конфигурации с массовым параллелизмом сверхвысокой производительности - модели суперкомпьютеров СКИФ Ряда 2 (СКИФ К-500 и СКИФ К-1000).

1. Суперкомпьютер СКИФ К-500

Кластер СКИФ К-500 (рис. 1) создан в 2003 году для отработки принципов построения моделей суперкомпьютеров СКИФ сверхвысокой производительности (триллионы операций в секунду). Создание этого кластера является качественно новым результатом, позволяющим вплотную приблизиться к терафлопному диапазону.

Предельная пиковая (реальная на задаче Linpack) производительность:	716,8(475,3)Gflops
Тип процессора:	Intel Xeon 2.8 Ghz
Число вычислительных узлов/процессоров:	64/128
Оперативная память узла/установки:	64*2=128 GB
Дисковая память установки:	64*60=3840 GB
Тип системной сети:	3D-топ, SCI, D336
Тип вспомогательной сети:	GB Ethernet
Конструктив узла (форм-фактор):	1U



Рис. 1. Суперкомпьютерная конфигурация СКИФ К-500 и ее основные технические характеристики

Суперкомпьютер СКИФ К-500 представляет собой кластер из 128 процессоров Хеон с частотой 2,8 ГГц. Реальная производительность кластера в 475,3 миллиарда операций в секунду, достигнутая на тесте LinPack, стала основанием для его включения под номером 407 в 22-ой выпуск списка 500 самых производительных компьютерных систем в мире top-500 (<http://www.top500.org/list/2003/11/>).

Структурная схема СКИФ К-500 представлена на рис.2.

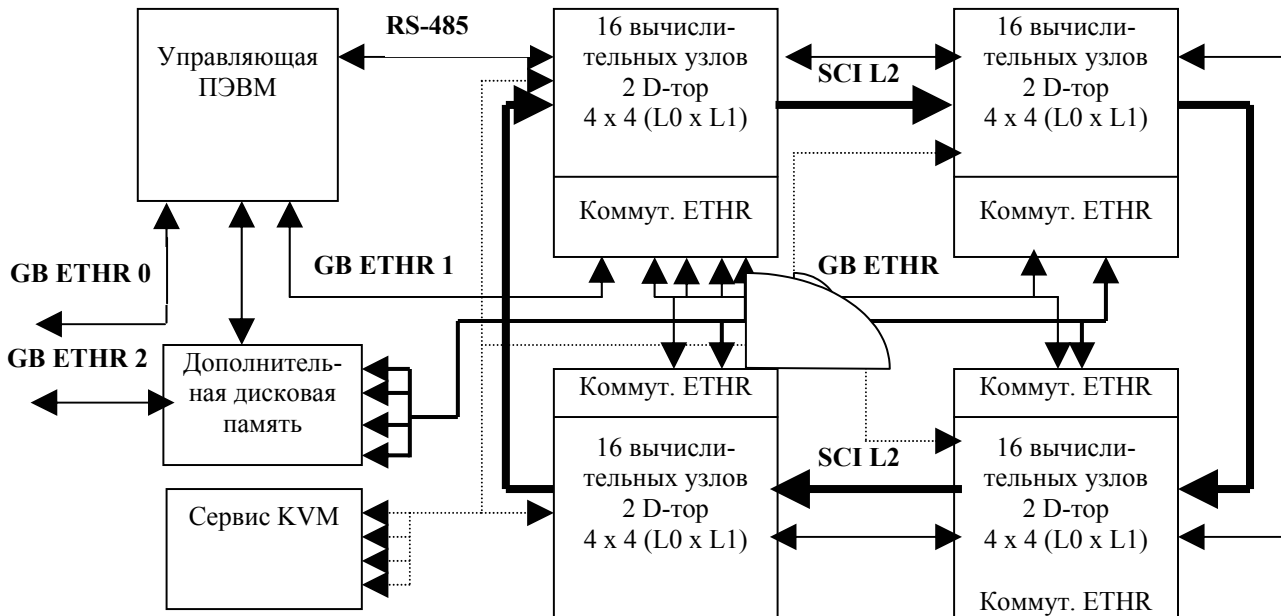


Рис.2. Структурная схема СКИФ К-500

Пиковая производительность каждого узла кластера СКИФ К-500 – 11,2 Гфлопс. SCI интерфейс обеспечивает задержку при передаче сообщений в соответствии со стандартом MPI – 3 мкс, скорость обмена узел-узел – 263 Мбайт/с.

Вычислительные узлы соединяются высокоскоростной сетью SCI и образуют трехмерный тор (3D-тор) 4 x 4 x 4 (L0 x L1 x L2) – кубической структуры. Блоки по 16 узлов в отдельных шкафах соединены в двухмерный тор (2D-тор 4 x 4). 3D-тор образуется с помощью 16 колец внешних связей SCI интерфейса между соответствующими вычислительными узлами разных шкафов по линку L2.

Вспомогательный сетевой интерфейс GB Ethernet 0 предназначен для загрузки программ, данных, управления и мониторинга. Он имеет звездообразную топологию, в которой к коммутатору соединенному с управляющей ПЭВМ подключены 3 коммутатора остальных шкафов. Вычислительные узлы одного шкафа подключаются к своему коммутатору. GB Ethernet 1 предназначен для доступа из внешней локальной сети к управляющей ПЭВМ.

Дополнительная дисковая память (файловый сервер) подключена к каждому из 4-х коммутаторов отдельным GB Ethernet каналом. С управляющей ПЭВМ она связана по GB Ethernet перекрестным кабелем. Она использует технологию хранения данных RAID5 и имеет общую емкость дисковой памяти около 800 Гбайт.

Сервисная сеть RS-485 предназначена для выполнения включения/выключения электропитания вычислительного узла, аппаратного сброса узла и выполнения взаимодействия с вычислительным узлом в консольном режиме с управляющей ПЭВМ.

KVM обеспечивает подключение любого из вычислительных узлов каждого шкафа к общей клавиатуре, видеомонитору, "мышь" для сервисного обслуживания.

Программное обеспечение кластера:

- операционная система LINUX RED HAT с поддержкой SMP;
- система управления, администрирования кластером и поддержка стандарта MPI 1.2 фирмы Scali (SSP 3.0.1);
- компиляторы C, C++;
- система программирования и управления выполнением вычислений – T-система.

2. Суперкомпьютер СКИФ К-1000

Суперкомпьютер СКИФ К-1000 создается с целью комплексной реализации принципов построения моделей суперкомпьютеров СКИФ Ряда 2 кластерного уровня с массовым параллелизмом сверхвысокой производительности (триллионы операций в секунду), включая сетевые (метакластерные) конфигурации. Суперкомпьютер СКИФ К-1000 является старшей моделью семейства СКИФ. Прогнозируется, что технические параметры суперкомпьютера СКИФ К-1000 должны обеспечивать его включение в top-500 в 2004-2005 годах.

3. Основные технические требования к суперкомпьютеру СКИФ К-1000

Реальная производительность суперкомпьютера СКИФ К-1000 - не менее 1,1 Тфлопс.

Каждый узел кластера должен содержать двухпроцессорную системную плату SMP, микропроцессоры не ниже Opteron 2.2 GHz, дисковую подсистему, средства для подключения к системной сети и вспомогательной сети Gigabit Ethernet, и иметь следующие параметры:

- емкость основной памяти – не ниже 2,0 Гбайт;
- емкость дисковой подсистемы – не менее 80 Гбайт;
- потребляемая мощность – не более 400 ВА.

Сетевые коммутаторы Infiniband должны обеспечивать работоспособность всех серверных узлов в топологии Fat Tree с обеспечением FBB 10 Gbit/sec. Система управления вычислительными узлами должна обеспечивать для любого вычислительного узла операции включение-выключение питания, аппаратного сброса (reset), возможность организации консольного доступа к любому серверному узлу с управляющего узла на этапах: загрузки BIOS, загрузки и работы операционной системы. Теоретическая пиковая производительность вычислительного узла кластера должна быть не менее 10 Гфлопс. Дополнительная дисковая память должна быть реализована с использованием современных типовых средств файловых серверов (с применением RAID-технологии). Программное обеспечение суперкомпьютера СКИФ К-1000 включает:

- ядро ОС Linux, адаптированное для работы на суперкомпьютерах семейства СКИФ;
- параллельная файловая система PVFS-SKIF;
- система очередей задач для кластерного уровня суперкомпьютеров семейства СКИФ;
- система мониторинга FLAME (Functional Active Monitoring Environment);
- распределенная программная система отладки MPI-программ (ПС TDB);

- ядро T-системы;
- компилятор TGCC языка TC;
- средство поддержки TC-режима для среды XEmacs;
- свободно распространяемые параллельные библиотеки и приложения, адаптированные для работы на суперкомпьютерах СКИФ;

Заключение

Для практического внедрения результатов программы СКИФ организован режим удаленного доступа к вычислительным ресурсам суперкомпьютерных конфигураций ОИПИ НАН Беларуси, сконцентрированных в создаваемом Национальном суперкомпьютерном центре коллективного пользования.