

КЛАСТЕРНЫЕ КОНФИГУРАЦИИ СКИФ РЯДА-1

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 кластерные конфигурации, составляющие Ряд-1 младших и средних моделей суперкомпьютеров в рамках программы СКИФ. Приведены их основные технические характеристики и состав программного обеспечения.

Введение

На этапе 1 (2000-2003 гг.) реализации программы «СКИФ» были отработаны основные концептуальные принципы и созданы кластерные конфигурации, обеспечивающие возможность создания семейства моделей суперкомпьютеров младшего и среднего классов (модели Ряда 1) с пиковой производительностью до 300-400 миллиардов операций в секунду (в том числе, суперсерверов СКИФ). Параметры суперсерверов СКИФ позволяют использовать их как в качестве платформы для ресурсоемких приложений (например, серверы баз данных), так и в качестве основы для построения крупных вычислительных центров.

1. Первые две кластерные конфигурации СКИФ

«Первенцы» (рис. 1) были созданы в 2000 г. в условиях чрезвычайно ограниченных временных рамок. На первых образцах кластеров СКИФ, один из которых был установлен в Минске, второй - в Переславле-Залесском, была проведена установка первой прикладной системы для проектирования химических реакторов, разработанной в Санкт-Петербурге в Институте высокопроизводительных вычислений и информационных систем (ИВВ ИС). Так получилось, что уровень пиковой производительности этих образцов - 20 Гфлопс (табл. 1) соответствовал уровню эмбарго на тот момент на ввоз в Россию и Беларусь высокопроизводительной вычислительной техники. Ввоз такой техники требовал оформления особого разрешения.



Рис. 1. Внешний вид первых двух кластеров СКИФ

Таблица 1

Основные технические характеристики первых образцов семейства СКИФ

Месяц и год выпуска	декабрь 2000 год
Место расположения	НИИ ЭВМ (Минск) и ИПС РАН (Переславль-Залесский)
Число вычислительных узлов/процессоров	16/32
Тип процессора	Intel Pentium III-600
Частота процессора	600 МГц
Предельная пиковая (реальная на задаче Linpack) производительность	20 (11) Гфлопс
Оперативная память установки	16 x 0.5 = 8 Гбайт
Дисковая память установки	16 x 10 = 160 Гбайт
Тип системной сети	2D-top 4 x 4, SCI, D311/312
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	3U

2. Кластерная конфигурация ВМ-5100 СКИФ

Второй опытный образец семейства СКИФ - ВМ-5100 - был выпущен в 2001 г. (рис. 2). Его отличительными чертами являются более высокопроизводительные (по сравнению с первыми образцами) процессоры Intel Pentium III-1400 МГц и более компактные конструктивы: 2U вместо 3U. Благодаря процессорам Intel Pentium III-1400 МГц на этом образце были достигнуты более высокие пиковая и реальная производительность: 44,8 Гфлопс и 30,89 Гфлопс, соответственно (табл. 2). Именно этот образец вместе с образцом «Первенец» являлись базой, на которой проводились государственные испытания в феврале 2002 г.

Таблица 2

ВМ-5100: основные технические характеристики

Месяц и год выпуска	декабрь 2001 год
Место расположения	НИИ ЭВМ (Минск)
Число вычислительных узлов/процессоров	16/32
Тип процессора	Intel Pentium III-1400 МГц
Частота процессора	1400 МГц
Предельная пиковая (реальная на задаче Linpack) производительность	48 (29) Гфлопс
Оперативная память установки	16 x 1 = 16 Гбайт
Дисковая память установки	16 x 18 = 288 Гбайт
Тип системной сети	2D-top 4 x 4, SCI, D335
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	1U
Дополнительно	Установлена лицензионная версия пакета LS-DANYA, срок лицензии: 16.01.2004 г.



Рис. 2. Суперкомпьютерная установка ВМ-5100 семейства СКИФ

3. Аппаратно-программный кардиологический комплекс

В 2001 г. в рамках программы были выполнены работы по созданию аппаратно-программного кардиологического комплекса (АПКК) на основе кластерных

технических решений СКИФ (рис.3, табл. 3). Комплекс предназначен для исследования микроциркуляторного звена сердечно-сосудистой системы методом биомикроскопии.

Таблица 3

Основные технические характеристики кардиокомплекса



Рис.3. Аппаратно-программный кардиологический комплекс

Месяц и год выпуска	декабрь 2001 год
Место расположения	ОИПИ НАН Беларуси (Минск)
Число вычислительных узлов/процессоров	4/3 x 2 + 1
Тип процессора	6 Intel Pentium III-1200МГц + 1 Intel Pentium IV- 1800МГц
Частота процессора	1266 МГц
Предельная пиковая (реальная на задаче Linpack) производительность	9.4 (5) Гфлопс
Оперативная память установки	4 x 1 = 4 Гбайт
Дисковая память установки	7 x 18 = 126 Гбайт
Тип системной и управляющей сети	Fast Ethernet
Конструктив узла (форм-фактор)	3 + 1-1U + 4 U

Кардиологический комплекс является ярким примером создания на базе технических решений СКИФ законченной прикладной системы. В системе была решена задача совместимости кластера и медицинской аппаратуры. Это позволяет в режиме реального времени обследовать пациента, ставить точный диагноз и, при необходимости, рекомендовать оптимальный путь лечения. В АПК используется оригинальная методика диагностики кардиологических заболеваний, разработанная Объединенным институтом проблем информатики Национальной академией наук Беларуси (ОИПИ НАН Беларуси) совместно с РНПЦ «Кардиология». Методика основана на том, что снимается и обрабатывается ряд видеоизображений капилляров на сетчатке глаза обследуемого. На такую съемку по физиологическим ограничениям отводится очень малое время, потому что глаз начинает слезиться. Кроме того, жесткое ограничение времени на обследование и постановку диагноза определяется еще и тем, что кардиологический комплекс предполагается использовать во время массовых профилактических осмотров населения.

Российская сторона тоже участвует в этой разработке. В ИВВ ИС в Санкт-Петербурге разработано программное обеспечение кластерного уровня, которое реализует экспертную систему, позволяющую кардиологу поставить обоснованный диагноз с использованием технологии искусственного интеллекта.

4. Экспериментальный гибридный суперкомпьютер семейства «СКИФ»

Первый показ экспериментального образца гибридного суперкомпьютера семейства СКИФ с названием «Гибрид» (рис. 4) был проведен в апреле 2002 г. Этот образец имеет двухуровневую архитектуру: кластерный уровень (КУ) и уровень однородной вычислительной среды (ОВС). Установка «Гибрид» используется разработчиками для отладки аппаратных и программных средств взаимодействия КУ и ОВС.

Таблица 4

Основные технические характеристики
экспериментального образца гибридного суперкомпьютера
семейства СКИФ

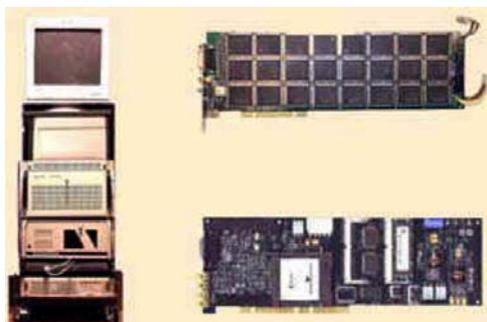


Рис. 4. Экспериментальный образец гибридного суперкомпьютера семейства СКИФ (снизу - управляющая плата ОВС, сверху - плата макета ОВС)

Месяц и год выпуска	апрель 2002 год
Место расположения	Суперкомпьютерные системы (СКС), Москва
Число вычислительных узлов/процессоров	2/3
Тип процессора	Intel Pentium III-800 МГц
Частота процессора	800 МГц
Предельная пиковая (реальная на задаче Linpack) производительность	2.4 (1.2) Гфлопс
Оперативная память установки	2 x 0.5 = 1 Гбайт
Дисковая память установки	2 x 80 = 160 Гбайт
Тип системной сети	1 x 2 1D-top, SCI, D320
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	4U + 6 U
Дополнительно	Подключена экспериментальная реализация модуля ОВС на Xilinx, 216 эл. процессоров, 10 МГц

Во время первой презентации установки «Гибрид» было показано решение задачи обработки изображений. На Т-системе была реализована программа, которая получала цепочку кадров видеоизображений, содержавших шумы. Т-система разбивала видеоряд на кадры, каждый из которых отдавался для обработки в ОВС. На уровне ОВС производилась фильтрация шумов в кадре, и результат возвращался в Т-систему. Пользователю показывались два окна в системе X-Windows. В режиме реального времени в одном окне Т-система предъявляла видеоизображение до обработки, а в другом - изображение после очистки от шумов.

В рамках создания данной установки программные мероприятия, связанные с разработкой и производством средств ОВС, как аппаратной, так и программной части, выполняло предприятие «Суперкомпьютерные системы» (СКС, Москва). Комплекс работ по обеспечению взаимодействия КУ и ОВС выполнялся СКС совместно с Институтом программных систем Российской академии наук (ИПС РАН). В настоящее время установка «Гибрид» (табл. 4) используется для отладки опытных образцов ОВС, выпускаемых в Минске (НИИ ЭВМ и УП «Белмикросистемы» НПО «Интеграл»).

5. «Первенец-М»: модернизация первого образца суперкомпьютера

Модернизация кластера «Первенец» (рис.5, табл. 5) была выполнена в июле 2002 г. При минимальных затратах удалось улучшить основные технические характеристики системы в два-три раза, а производительность - почти в пять раз. В настоящее время установка «Первенец-М» является основной вычислительной

мощностью в Переславле-Залесском. Она доступна по сети, и свыше ста пользователей – потребителей высокопроизводительных вычислений – имеют доступ к ней и решают свои задачи на этой установке.

Таблица 5

Основные технические характеристики установки
«Первенец-М»



Рис. 5. Кластерная установка
«Первенец-М», Переславль-Залесский,
ИПС РАН, 2002 г.

Месяц и год выпуска	июль 2002 год
Место расположения	ИПС РАН (Переславль-Залесский)
Число вычислительных узлов/процессоров	16/32
Тип процессора	AMD AthlonMP 1800 +
Частота процессора	1533 МГц
Предельная пиковая (реальная на задаче Linpack) производительность	98 (57) Гфлопс
Оперативная память установки	2 x 1 = 16 Гбайт
Дисковая память установки	16 x 40 = 640 Гбайт
Тип системной сети	4 x 4 2D-top, SCI, D335
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	3U
Дополнительно	Установлена лицензионная версия пакета STAR-CD, срок лицензии: 07.07.2004 г.

6. «Студент»: вспомогательный кластер семейства СКИФ

Это девятиузловой вспомогательный кластер (рис. 6, табл. 6), выполненный в конструктивах MiniTower на стойке, которая была сделана в мастерских ИПС РАН. На установке работают в основном студенты из Московского государственного университета (МГУ) и из университета города Переславля.

7. Установка семейства СКИФ в НИИ механики МГУ и установка «Mugin»

По аппаратным решениям этот кластер является близким аналогом «Первенца-М». Отличительной особенностью этого образца является то, что его реализация была выполнена не за счет средств программы «СКИФ», а целиком за счет средств НИИ механики МГУ (рис. 7, табл. 7).



Рис. 7. Внешний вид кластерной установки в НИИ механики МГУ

Рис. 6. Внешний вид вспомогательной кластерной установки «Студент»

Таблица 6
Основные технические характеристики кластерной установки «Студент»

Месяц и год выпуска	июль 2002 год
Место расположения	ИПС РАН (Переславль-Залесский)
Число вычислительных узлов/процессоров	9/18
Тип процессора	Intel Pentium III-600 МГц
Частота процессора	600 МГц
Предельная пиковая (реальная на задаче Linpack) производительность	11 (6) Гфлопс
Оперативная память установки	16 x 0.5 = 4.5 Гбайт
Дисковая память установки	90 Гбайт
Тип системной сети	3 x 3 2D-top SCI SCI D311/312
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	MiniTower

Таблица 7
Характеристики кластерной установки в НИИ механики МГУ

Месяц и год выпуска	июль 2002 год
Место расположения	НИИ механики МГУ (Москва)
Число вычислительных узлов/процессоров	8/16
Тип процессора	AMD AthlonMP 1800 +
Частота процессора	1533 МГц
Предельная пиковая (реальная на задаче Linpack) производительность	49 (28) Гфлопс
Оперативная память установки	8 x 1 = 8 Гбайт
Дисковая память установки	8 x 80 = 640 Гбайт
Тип системной сети	2 x 4 2D-top, SCI D335
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	4 x 4U + 4 x 5U
Дополнительно	Кластер оснащен сервисной сетью, разработанной в ИПС РАН

В том же 2002г. для изучения технологии Myrinet в Минске (ОИПИ НАН Беларуси и УП «НИИЭВМ») был разработан кластер с условным названием «Myrin» (рис. 8, табл. 8).



Рис. 8. Установка «Myrin»

Таблица 8
Основные технические характеристики установки «Myrin»

Месяц и год выпуска	ноябрь 2002 год
Место расположения	ОИПИ НАН Беларуси
Число вычислительных узлов/процессоров	8/16
Тип процессора	Intel Xeon 2.8 ГГц
Частота процессора	2.8 ГГц
Предельная пиковая (реальная на задаче Linpack) производительность	89 (59) Гфлопс
Оперативная память установки	8 x 1 = 8 Гбайт
Дисковая память установки	8 x 40 = 320 Гбайт
Тип системной сети	Myrinet
Тип управляющей (вспомогательной) сети	Fast Ethernet

Конструктив узла (форм-фактор)	1U
--------------------------------	----

8. «T-Forge 32»: первый в СНГ кластер с процессорами AMD Opteron

ИПС РАН совместно с компанией «Т-Платформы» в июне 2003 г. была завершена сборка и проведена презентация первого в России кластера на базе процессоров AMD Opteron 244 - «T-Forge 32» (рис. 9, табл. 9). Установка была реализована с системной сетью на базе плат GB Ethernet. Но даже на такой системной сети кластер показал весьма высокие технические характеристики: высокую пиковую производительность и очень хороший КПД (отношение реальной Linpack-производительности к пиковой производительности) – 64%. В рамках же одного узла процессоры Opteron показали еще лучшее соотношение между пиковой и реальной производительностью на задаче Linpack - 80%.

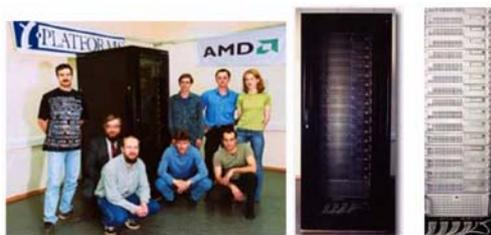


Рис. 9. «T-Forge 32» на базе процессоров AMD Opteron 244, Москва, компания «Т-Платформы», июнь 2003 г.

Таблица 9
Основные технические характеристики кластера «T-Forge 32»

Месяц и год выпуска	июнь 2003 год
Место расположения	Центр кластерных технологий (ЦКТ), Москва
Число вычислительных узлов/процессоров	16/32
Тип процессора	AMD Opteron 244
Частота процессора	1800 МГц
Предельная пиковая (реальная на задаче Linpack) производительность	115.2 (73.8) Гфлопс
Оперативная память установки	16 x 2 = 32 Гбайт
Дисковая память установки	16 x 60 = 960 Гбайт
Тип системной сети	GB Ethernet
Тип управляющей (вспомогательной) сети	GB Ethernet
Конструктив узла (форм-фактор)	2U
Дополнительно	Linpack 64% (размерность 110,000). Кластер оснащен сервисной сетью, разработанной в ИПС РАН

Впервые на кластере «T-Forge 32» была применена технологическая новинка - кластер оснащен сервисной сетью, разработанной в ИПС РАН.

Опыт работы с процессорами Opteron показал, что это действительно очень мощное решение, новое слово в компьютерной отрасли, поскольку процессор Opteron предоставляет пользователям реализацию мягкого перехода от 32-разрядных приложений к 64-разрядным приложениям. Процессор поддерживает и 32-разрядный режим работы, и 64-разрядный режим работы, причем 32-разрядный режим поддержан

очень эффективно – все 32-разрядные приложения работают с отличной скоростью. Эта работа была важна тем, что впервые в кластерах в России применялись 64-разрядные архитектуры.

Кроме того, что применение 64-разрядных архитектур дает повышение производительности, существует еще такой аспект высокопроизводительных вычислений, как потребность в больших объемах памяти. Как правило, большие приложения требуют больших объемов памяти и, следовательно, больших объемов адресного пространства, которые могут обеспечивать только 64-разрядный режим адресации. Только этого обстоятельства достаточно, чтобы сказать, что переход на 64-разрядные архитектуры перспективен.

9. Суперкомпьютерная кластерная установка ЕС1710.03

В 2003 г. в рамках программы «СКИФ» в ОАО «НИЦЭВТ» создана старшая модель кластеров СКИФ Ряда-1 ЕС1710.03 (рис. 10, табл. 10). Этот образец создавался для отработки типового решения для установок с высокой производительностью. К его сильным сторонам относятся весьма высокая производительность (пиковая производительность 0,43 Тфлопс), компактность (образец собран в одной стойке, форм-фактор 1U), и, самое главное, системная сеть в установке ЕС1710.03 выполнена на платах SCI №335, которые выпускает ОАО «НИЦЭВТ» и которые являются аналогами плат SCI N335 компании Dolphin.

Таким образом, произошла частичная замена импортных комплектующих на комплектующие отечественного производства. В 2003 г. ОАО «НИЦЭВТ» была выпущена установочная партия адаптеров N335 (рис. 11) - 36 плат для 36 узлов вычислительной системы ЕС1710.03. Особенностью установки ЕС1710.03 является также то, что в ее состав включено сетевое устройство хранения (Network Attached Storage - NAS) на 480 Гбайт.

Таблица 10
ЕС1710.03: основные технические характеристики



Рис. 10. Установка ЕС1710.03, НИЦЭВТ, Москва, ноябрь 2003 г.

Месяц и год выпуска	октябрь 2003 год
Место расположения	НИЦЭВТ, Москва
Число вычислительных узлов/процессоров	36/72
Тип процессора	Intel Xeon 2.8 ГГц
Частота процессора	2800 МГц
Предельная пиковая (реальная на задаче Linpack) производительность	403 (230) Гфлопс
Оперативная память установки	36 x 2 = 72 Гбайт
Дисковая память установки	36 x 60 + 480 = 2640 Гбайт
Тип системной сети	6 x 6 2D-top, SCI N335
Тип управляющей (вспомогательной) сети	Fast Ethernet
Конструктив узла (форм-фактор)	1U
Дополнительно	В составе NAS 480 Гбайт



Рис. 11. Адаптер системной сети SCI N335, выпускаемый ОАО «НИЦЭВТ»

Заключение

Кластерные конфигурации семейства СКИФ, разработанные в 2000-2003 гг., использовались как для отладки программного обеспечения кластерного уровня, так и для реальных вычислений в интересах предприятий и учреждений России и Беларуси. Эти установки послужили материальной базой для успешного выполнения программы в 2000-2004 гг. и получения всех основных результатов по программе.