








Review Approval



-  Prepare Request
-  **Search Requests**
-  Generate Reports
-  Approvals
-  Help
-  Wizard

Search Detail

-  Search Requests
- [New Search](#)
- [Refine Search](#)
- [Search Results](#)
-
- [Clone Request](#)
- [Edit Request](#)
- [Cancel Request](#)

Submittal Details

Document Info		
Title : Red Storm Overview		
Document Number : 5231256	SAND Number : 2005-2126 P	
Review Type : Electronic	Status : Approved	
Sandia Contact : DEBENEDICTIS,ERIK P.	Submittal Type : Viewgraph/Presentation	
Requestor : DEBENEDICTIS,ERIK P.	Submit Date : 04/01/2005	
Comments : Presentation as part of a contract meeting and collaboration between Russian Federal Nuclear Center VNIIEF in Sarov, Russia (a Russian Nuclear Weapons Laboratory) and Sandia.		
Peer Reviewed? : N		
Author(s)		
TOMKINS,JAMES L.		
Event (Conference/Journal/Book) Info		
Name : Meeting at RFNC VNIIEF for collaboration of contracts under Gordon-Ryabev accord		
City : Sarov	State :	Country : Russia
Start Date : 04/12/2005	End Date : 04/13/2005	
Partnership Info		
Partnership Involved : No		
Partner Approval :	Agreement Number :	
Patent Info		
Scientific or Technical in Content : Yes		
Technical Advance : No	TA Form Filed : No	
SD Number :		
Classification and Sensitivity Info		
Title : Unclassified-Unlimited	Abstract :	Document : Unclassified-Unlimited
Additional Limited Release Info : None.		
DUSA : None.		

Routing Details

Role	Routed To	Approved By	Approval Date
Derivative Classifier Approver	SUMMERS,RANDALL M.	SUMMERS,RANDALL M.	04/01/2005
Conditions:			
Classification Approver	WILLIAMS,RONALD L.	WILLIAMS,RONALD L.	04/04/2005
Conditions: Pictures of Red Storm look fine for unclassified/unlimited release. Many of the slides where not viewable, it is up to the cognizant DC to ensure slides are unclassified/unlimited release.			
Manager Approver	PUNDIT,NEIL D.	PUNDIT,NEIL D.	04/04/2005
Conditions: PLEASE UPDATE THE STATUS OF RED STORM AS CLOSE TO YOUR DEPARTURE AS POSSIBLE. Sue Kelly or Paul Iwanchuk can provide the update.			
Sandia Contact	DEBENEDICTIS,ERIK P.	DEBENEDICTIS,ERIK P.	04/04/2005
Agreement: Sandia Contact has agreed to incorporate above listed conditions prior to release.			
Comments: I sent the non-picture slides to Ron Williams in a different format and he responded as below: Erik, These came through and they look fine for Unclassified/unlimited release. So, go forth and do great things with the Russian's. Ron Also, I discussed implementation of updates with Neil and we agreed on a course of action.			
Administrator Approver	LUCERO,ARLENE M.	KRAMER,SAMUEL	05/24/2007
Please add the funding statement: Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.			

Created by WebCo Problems? Contact CCHD: by email or at 845-CCHD (2243).

For Review and Approval process questions please contact the **Application Process Owner**

SAND2005-2126P

Обзор системы Red Storm

Erik P. DeBenedictis
(для James L. Tomkins)

Сандийские национальные лаборатории
Albuquerque, NM

Задачи архитектуры системы

- Сбалансированная производительность системы: центральный процессор, память, межкомпонентные связи и ввод/вывод.
- Эксплуатационная пригодность: функциональность аппаратного и программного обеспечения отвечают потребностям пользователей в вычислениях с массовым параллелизмом.
- Масштабируемость: аппаратное и программное обеспечение системы масштабируется от системы в одном шкафу до системы с 32000 процессоров.
- Надежность: машина находится в рабочем состоянии между прерываниями достаточно долго, чтобы обеспечить реальный прогресс в выполнении этапа прикладной задачи (среднее время между сбоями не менее 50 часов), требуется полная системная поддержка RAS (сервиса удаленного доступа).
- Возможность расширения: система может быть оснащена свопом процессора и дополнительными шкафами до 100T или более.
- Переключение красный/черный: возможность переключать основные блоки машины между защищенным и незащищенным режимом вычислений.
- Пространство, питание, охлаждение: системы высокой плотности с низким потреблением энергии.
- Цена/производительность: отличное соотношение цены и производительности, во всех возможных случаях использованы детали массового выпуска.

Архитектура системы **Red Storm**

- Настоящий массовый параллелизм вычислений (MPP), разработана как единая система.
- Параллельный суперкомпьютер MIMD (с многими потоками команд и данных) с распределенной памятью.
- Полное подключение 3-мерной ячеистой сети межкомпонентных соединений. Процессор каждого вычислительного узла имеет двустороннее соединение с основной сетью связи.
- 108 шкафов вычислительных узлов и 10 368 процессоров вычислительных узлов (AMD Sledgehammer @ 2.0 GHz).
- ~30 ТВ памяти DDR.
- Переключение режимов красный/черный: ~1/4, ~1/2, ~1/4.
- На каждой стороне 8 сервисных шкафов и шкафов ввода/вывода (256 процессоров на каждый цвет).
- > 240 ТВ дискового пространства (> 120 ТВ на каждый цвет).

Архитектура системы **Red Storm**

- Функциональное разделение аппаратных средств: сервисные узлы и узлы ввода/вывода, вычислительные узлы, а также узлы RAS.
- Расчлененная операционная система (OS): LINUX на сервисных узлах и узлах ввода/вывода, LWK (Catamount) на вычислительных узлах, облегченный LINUX на узлах RAS.
- Раздельные сети RAS (надежность, готовность, обслуживаемость) и системного администрирования (Ethernet).
- Маршрутизация в межкомпонентных соединениях при помощи настольного маршрутизатора.
- Общая мощность и охлаждение менее 2 МВт.
- Каждая машина занимает менее 3 000 футов² (279 м²) площади.

Топология системы **Red Storm**

- Топология вычислительных узлов:
 - ◆ $27 \times 16 \times 24$ (x, y, z) – разделение красный/черный:
 $2,688 - 4,992 - 2,688$
- Топология сервисных узлов и узлов ввода/вывода:
 - ◆ $2 \times 8 \times 16$ (x, y, z) на каждой стороне (сеть $2 \times 16 \times 16$)
 - ◆ 256 соединений с полной шириной полосы с ячеистой сетью вычислительных узлов (всего 384)

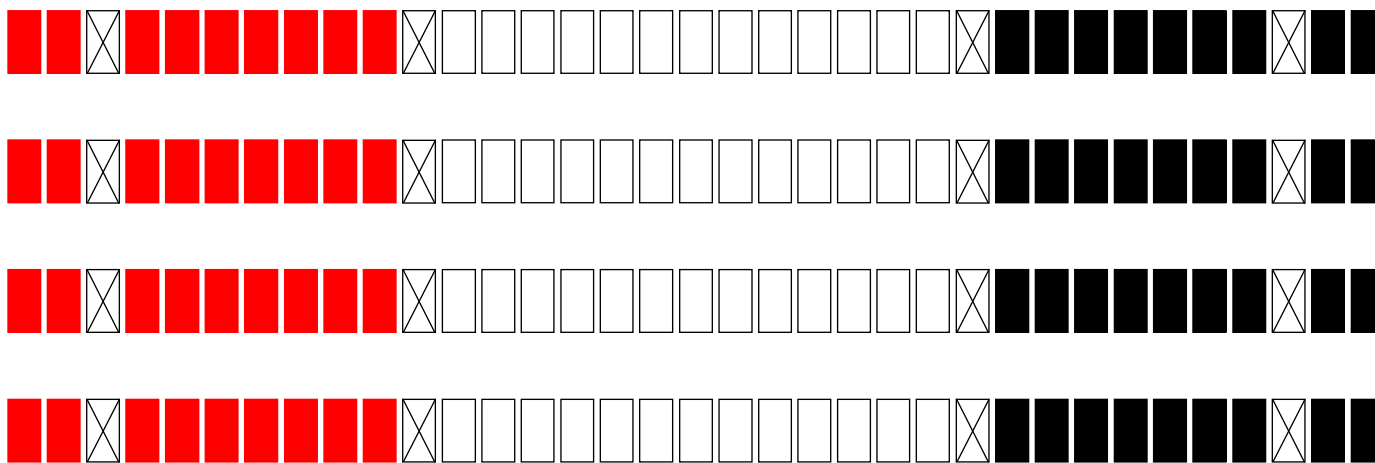
Компоновка системы Red Storm

(ячеистая сеть $27 \times 16 \times 24$)

Обычно
открытый
режим

Обычно
секретный
режим

Переключаемые узлы



Узлы ввода/вывода
и сервисные узлы

Узлы ввода/вывода
и сервисные узлы

Отключающие шкафы

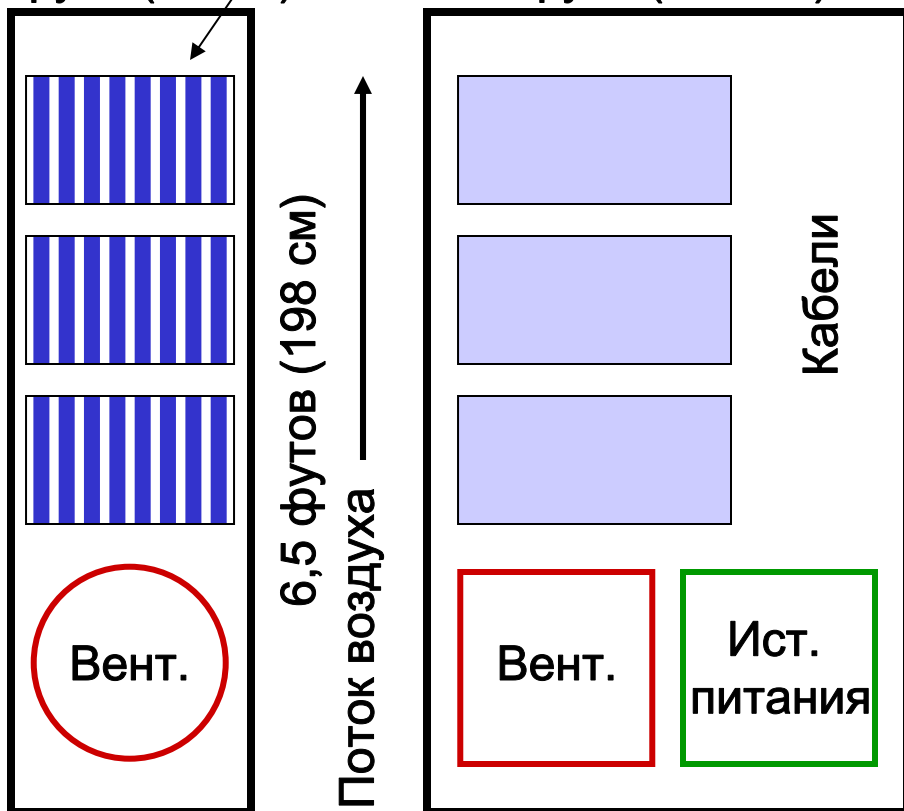
Система дисковых
накопителей не показана

Компоновка шкафа системы **Red Storm**

Платы центральных процессоров
в шкафу вычислительного узла

2 фута (61 см)

4 фута (122 см)



Лицевая
панель

Боковая панель

- Шкаф вычислительного узла
 - ♦ 3 кассеты для карт на шкаф.
 - ♦ 8 плат на кассету.
 - ♦ 4 процессора на плату.
 - ♦ 4 микросхемы сетевых интерфейсов (NIC)/маршрутизаторов на плату.
 - ♦ N + 1 источников питания.
 - ♦ Пассивная объединительная панель.
- Шкаф сервисного узла и узла ввода/вывода
 - ♦ 2 кассеты для карт на шкаф.
 - ♦ 8 план на кассету.
 - ♦ 2 процессора на плату.
 - ♦ 4 микросхемы сетевых интерфейсов (NIC)/маршрутизаторов на плату.
 - ♦ Шина PCI-X на каждый процессор.
 - ♦ N + 1 источников питания.
 - ♦ Пассивная объединительная панель.

Архитектура системы **Red Storm**

- Рабочие станции RAS (надежность, готовность, обслуживаемость)
 - ◆ Отдельные и резервные рабочие станции RAS для красной и черной сторон машины.
 - ◆ Интерфейс администрирования и контроля системы.
 - ◆ Журнал ошибок и контроль для основных компонентов системы, включая процессоры, память, сетевой интерфейс (NIC)/маршрутизатор, источники питания, вентиляторы, дисковые контроллеры и диски.
- Сеть RAS: выделенная сеть Ethernet для соединения узлов RAS с рабочими станциями RAS.
- Узлы RAS.
 - ◆ Один на каждую вычислительную плату.
 - ◆ Один на каждый шкаф.

Системное программное обеспечение

Red Storm

- **Операционные системы:**
 - ◆ LINUX на сервисных узлах и узлах ввода/вывода
 - ◆ LWK (Catamount) на вычислительных узлах
 - ◆ LINUX на узлах RAS
- **Система поддержки исполнения программ:**
 - ◆ логарифмический загрузчик
 - ◆ распределитель узлов
 - ◆ пакетная система – PBS
 - ◆ библиотеки – MPI, ввод/вывод, математика
- **Файловые системы - Lustre для систем UFS и параллельных систем**

Системное программное обеспечение

Red Storm

- Инструменты:
 - ◆ стандартные компиляторы ANSI – Fortran, C, C++
 - ◆ отладчик – *TotalView*
 - ◆ монитор производительности – PAPI
- Управление и администрирование системы:
 - ◆ учет
 - ◆ графический интерфейс пользователя RAS

Производительность системы **Red Storm**

- Максимум ~ 40 TF на основании 2 подач команд с плавающей запятой на такт. Ожидаемая производительность в ~ 10 раз выше чем у **ASCI Red**.
- Производительность MP-Linpack: > 14 TF (ожидается повышение до ~ 30 TF).
- Общая ширина полосы системной памяти: ~ 55 TB/s.
- Общая постоянная ширина полосы межкомпонентных соединений: > 100 TB/s.

Производительность системы **Red Storm**

Процессоры и память

- Процессоры.
 - ◆ AMD Sledgehammer (Opteron).
 - ◆ 2,0 GHz.
 - ◆ 64-битное расширение набора инструкций IA32.
 - ◆ 64 KB кэш уровня L1 для инструкций и данных на чипе.
 - ◆ 1 MB кэш уровня L2, совместно используемого для данных и инструкций, на чипе.
 - ◆ Интегрированные двойные контроллеры памяти DDR @ 333 MHz.
 - ◆ 3 интегрированных интерфейса Hyper Transport Interface @ 3,2 GB/s в каждом направлении.
- Система памяти узлов.
 - ◆ Время ожидания памяти локального процессора при пропуске страниц составляет ~80 ns.
 - ◆ Максимальная ширина полосы памяти составляет ~5,3 GB/s на каждый процессор.

Производительность системы **Red Storm**

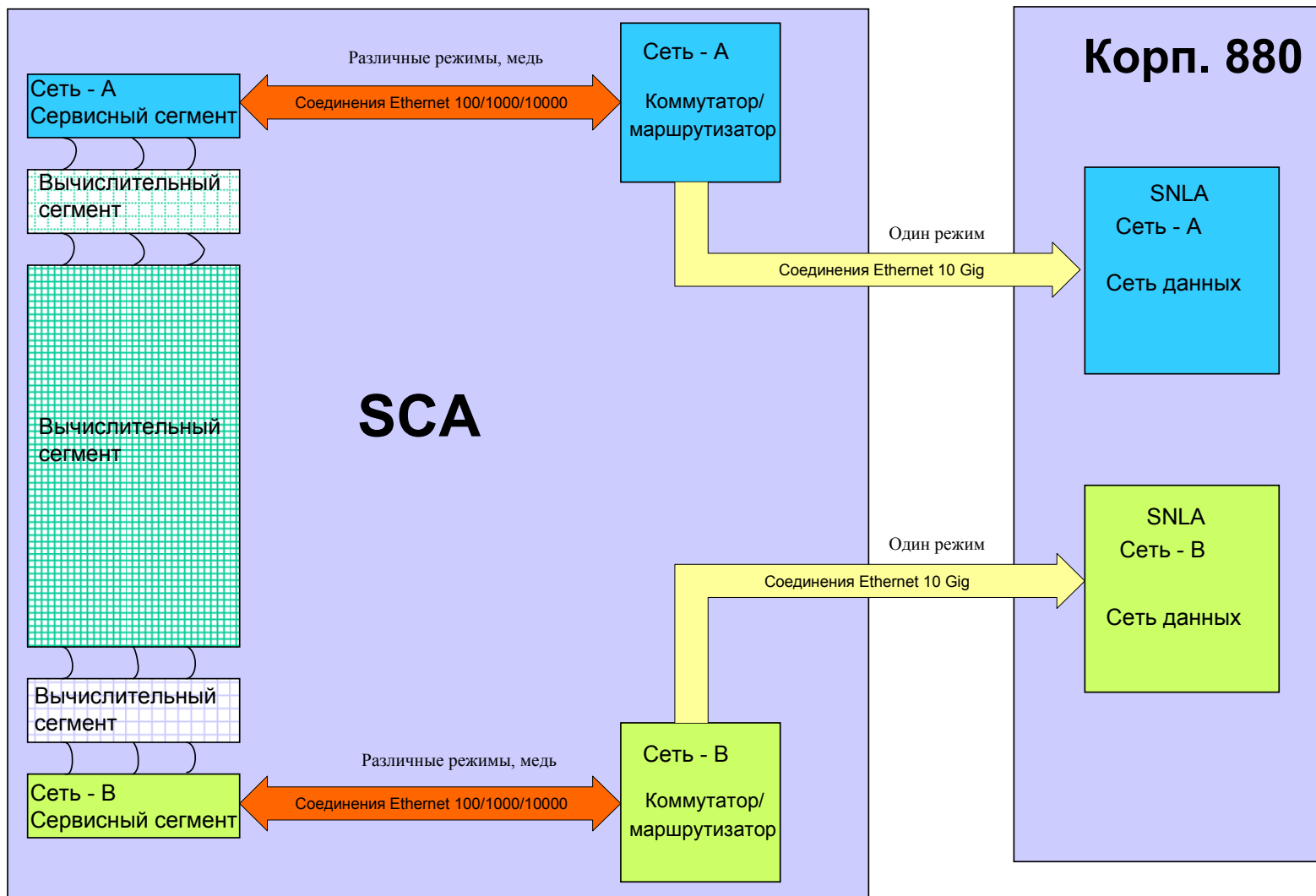
Межкомпонентные соединения и ввод/вывод

- Программный интерфейс Sandia/UNM Portals 3.3.
- Производительность межкомпонентных соединений.
 - ◆ Требования к времени ожидания MPI $< 2 \mu\text{s}$ (сосед), $< 5 \mu\text{s}$ (вся машина).
 - ◆ Максимальная ширина полосы соединения 3,84 GB/s в каждом направлении.
 - ◆ Двухсекционная ширина полосы $\sim 2,95 \text{ TB/s}$ Y-Z, $\sim 4,98 \text{ TB/s}$ X-Z, $\sim 6,64 \text{ TB/s}$ X-Y.
- Производительность системы ввода/вывода.
 - ◆ Постоянная ширина полосы файловой системы составляет 50 GB/s на каждый цвет.
 - ◆ Постоянная ширина полосы внешней сети составляет 25 GB/s на каждый цвет.

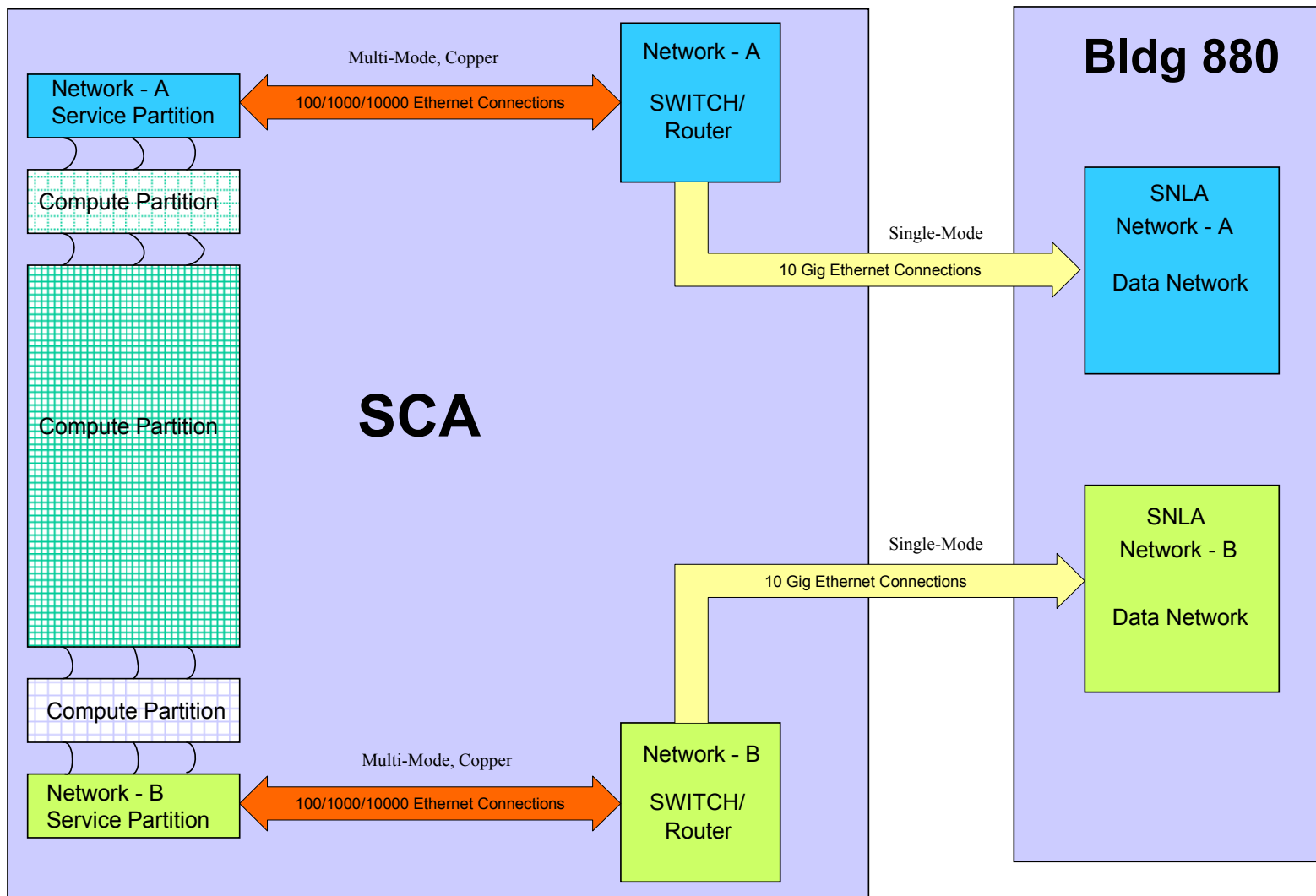
Статус сети системы Red Storm

**Установлены и готовы к
эксплуатации многочисленные
сети для поддержки системы Red
Storm.**

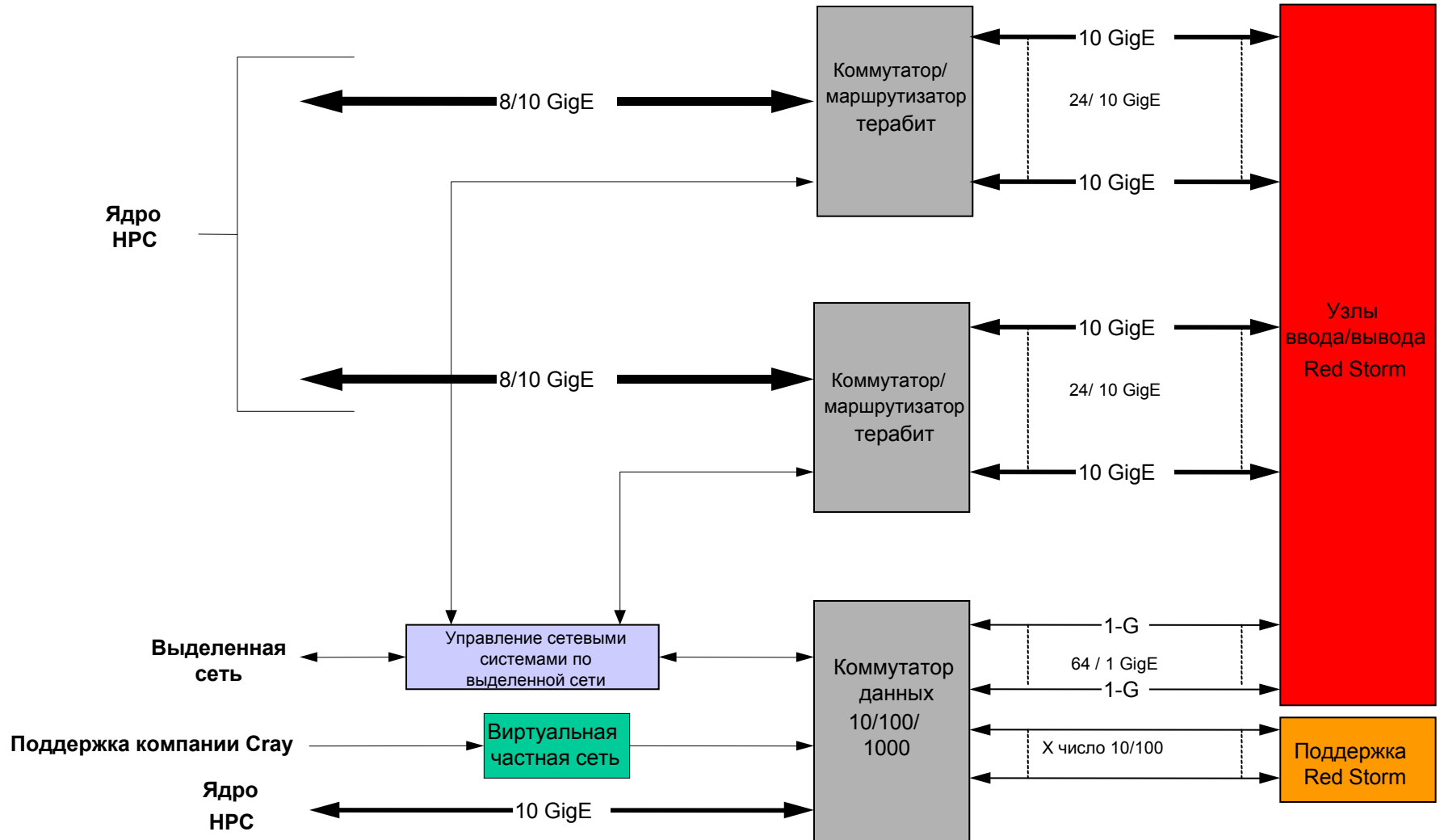
Сеть данных системы Red Storm



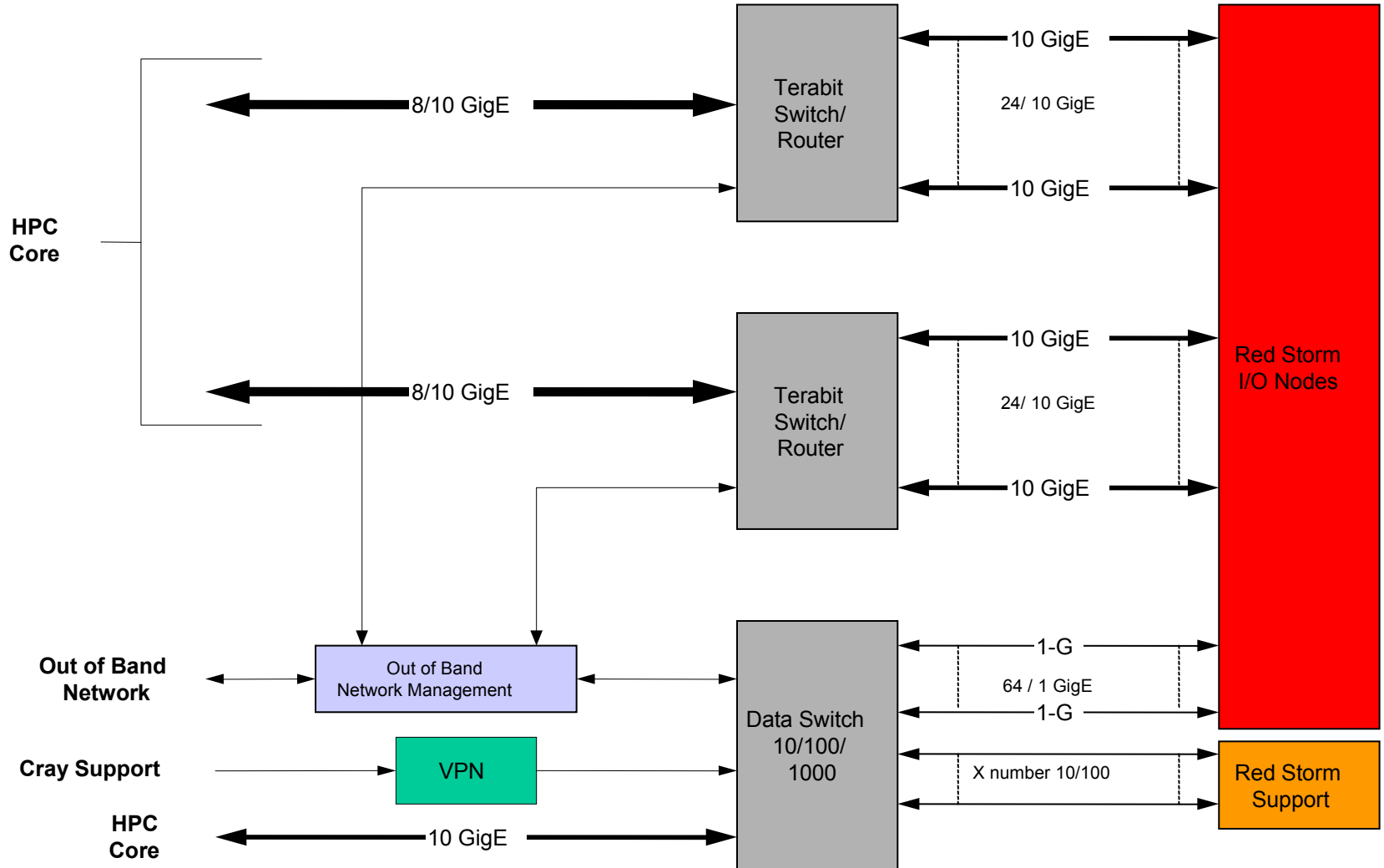
Red Storm Data Network



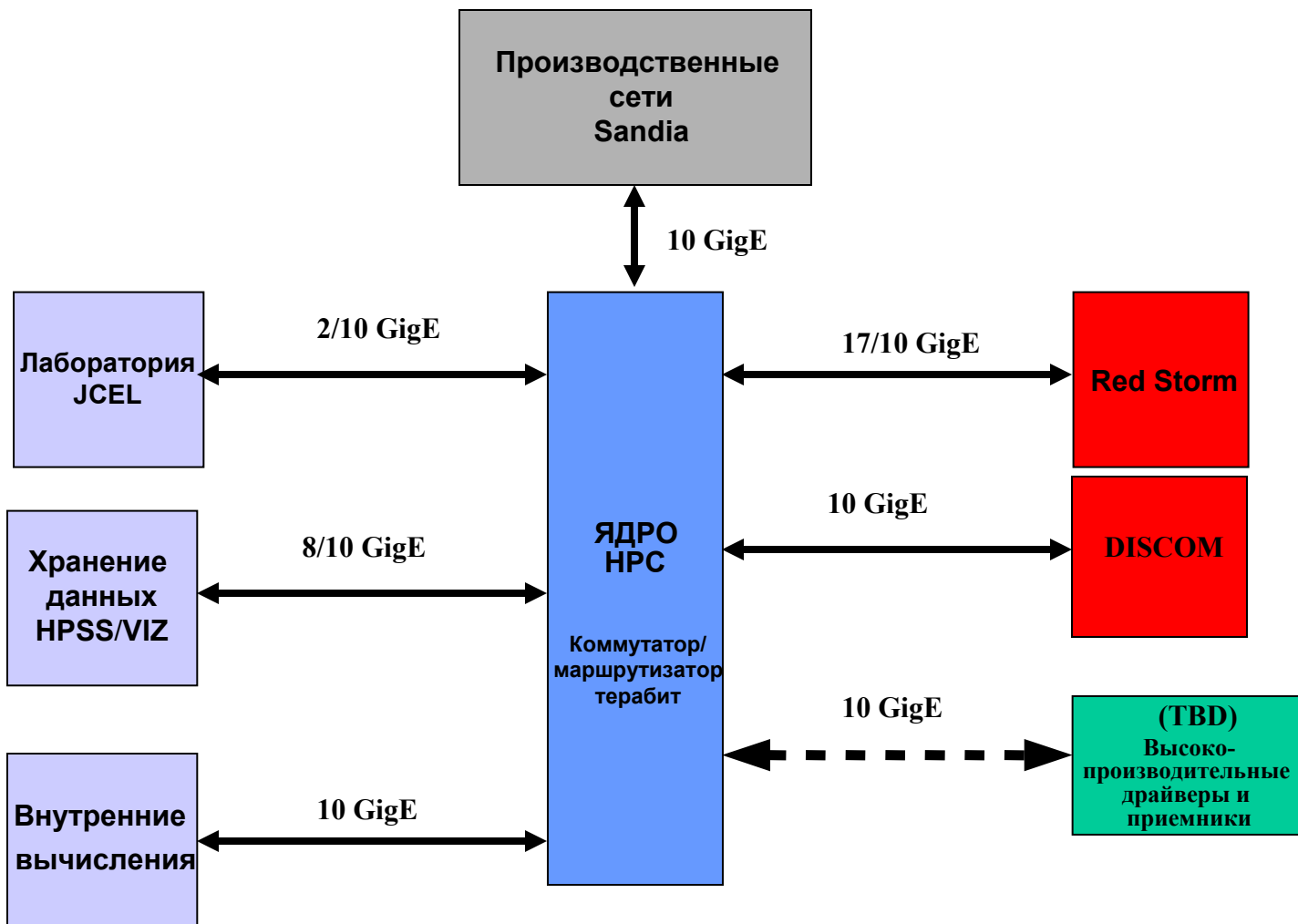
Соединения системы Red Storm



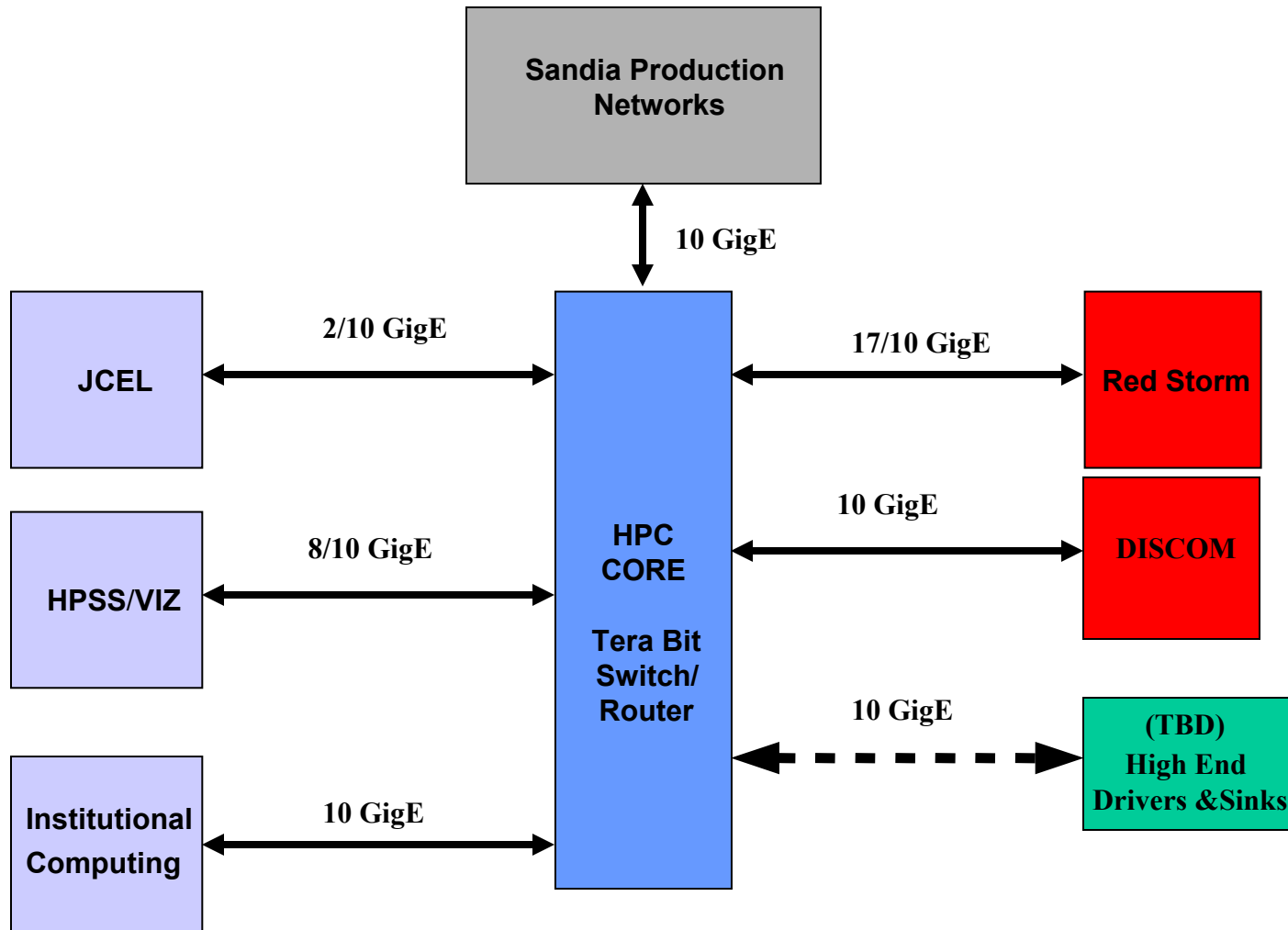
Red Storm Connectivity



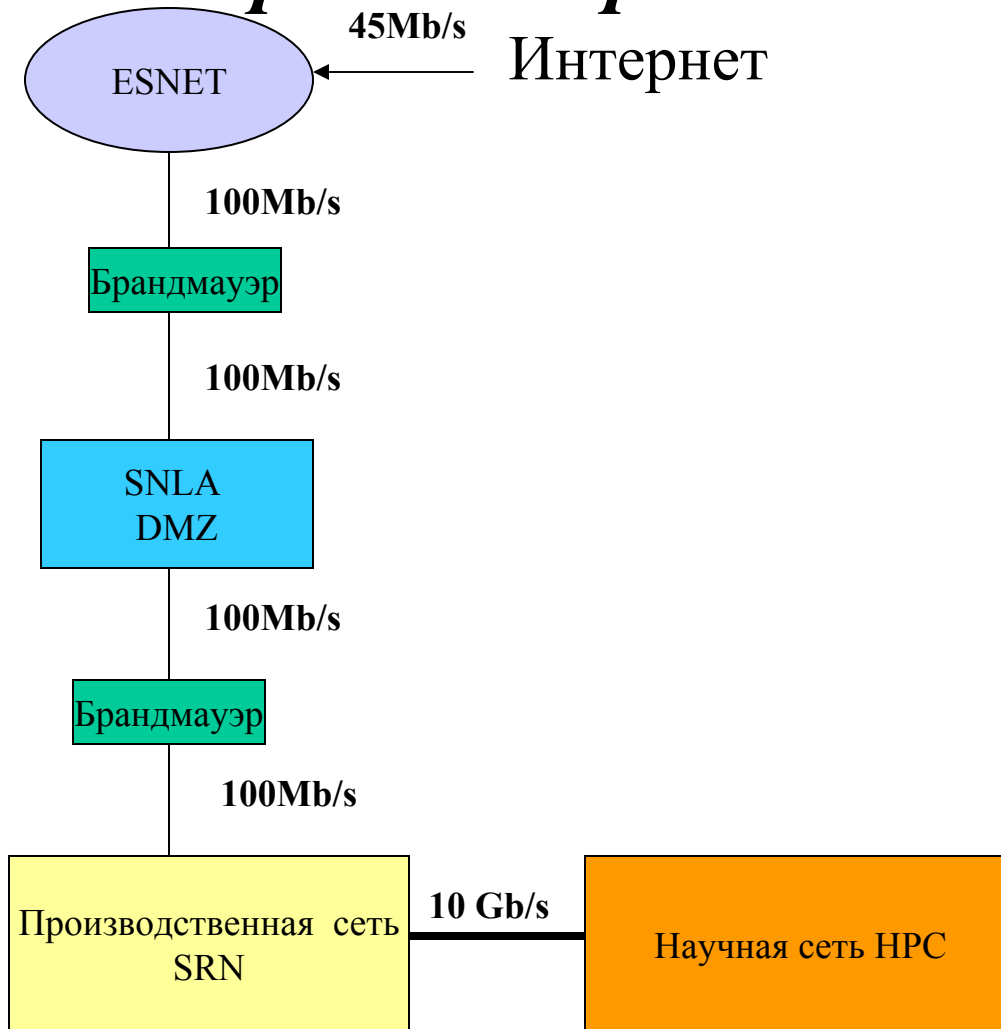
Соединения системы Red Storm с производственными сетями



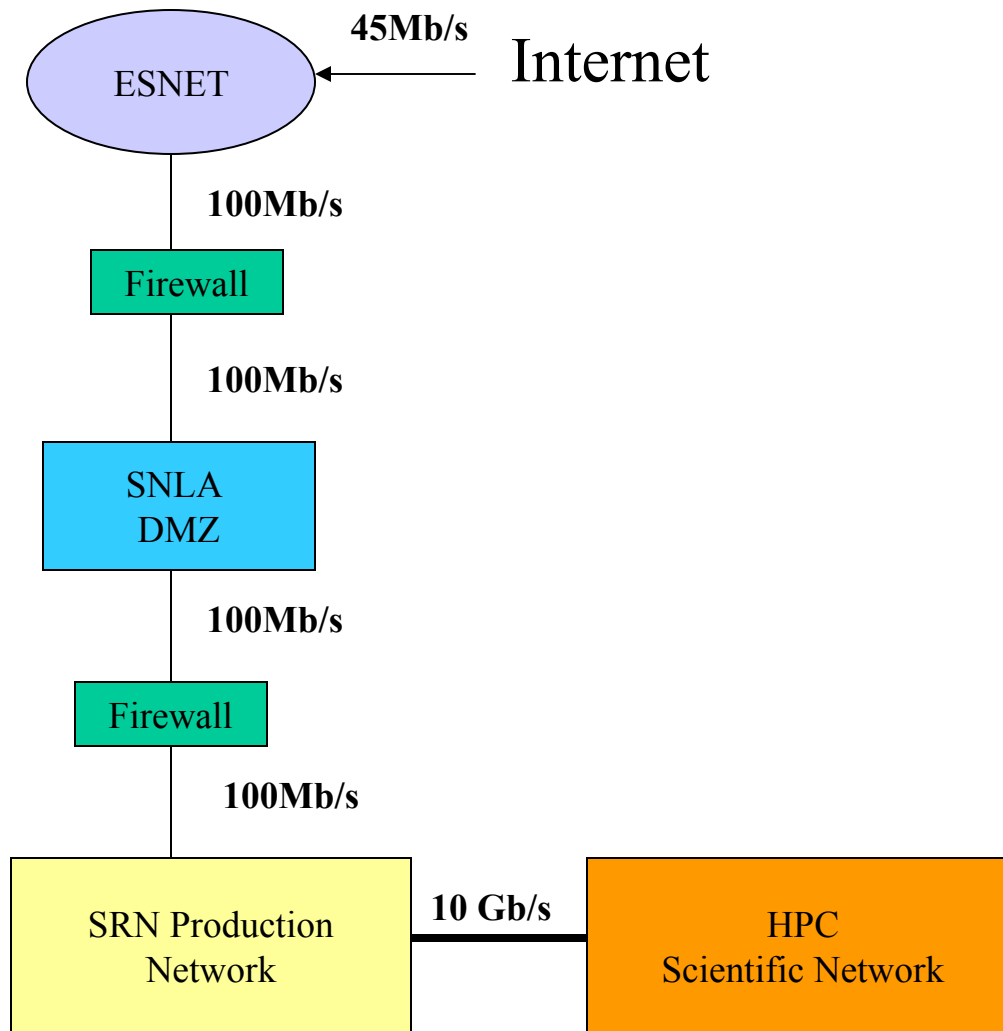
Red Storm Connections To Production Networks



Доступ к системе *Red Storm* через Интернет



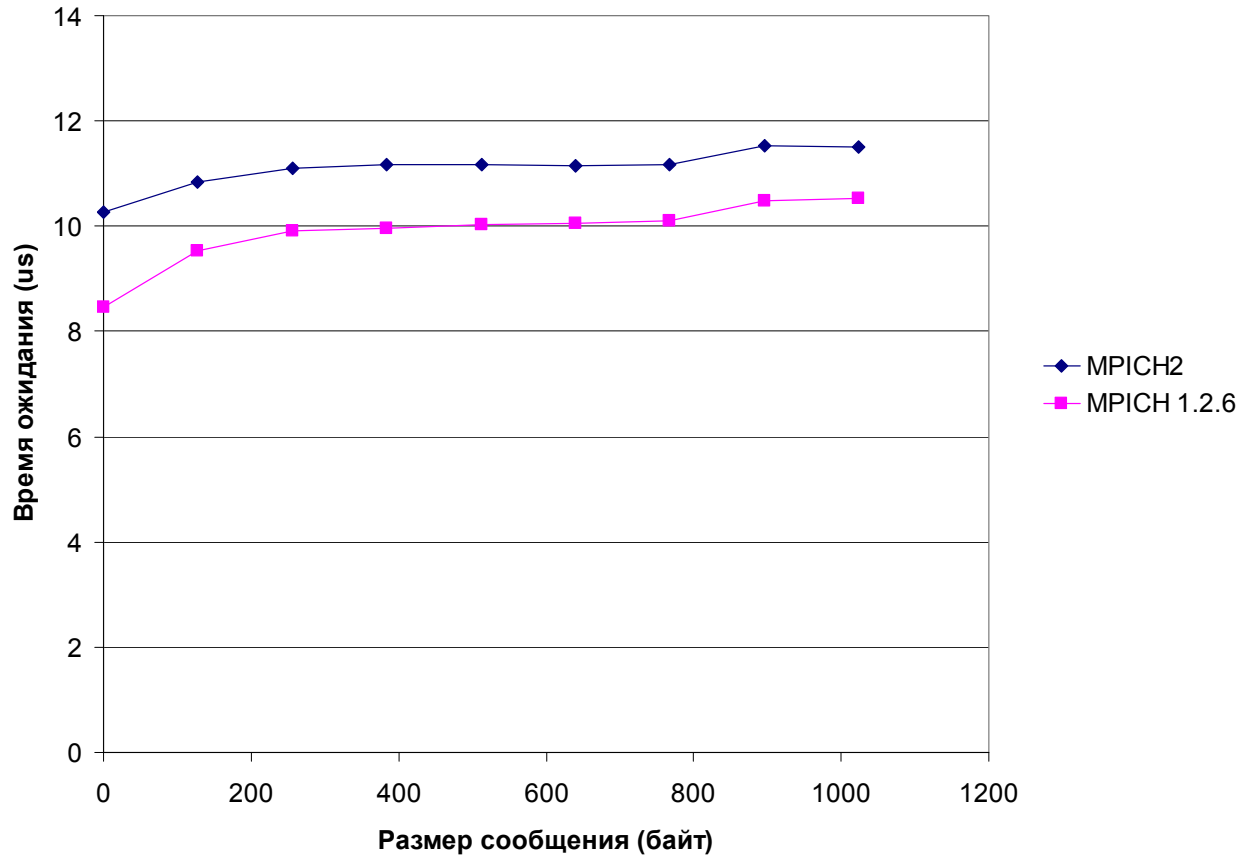
Internet Access to Red Storm



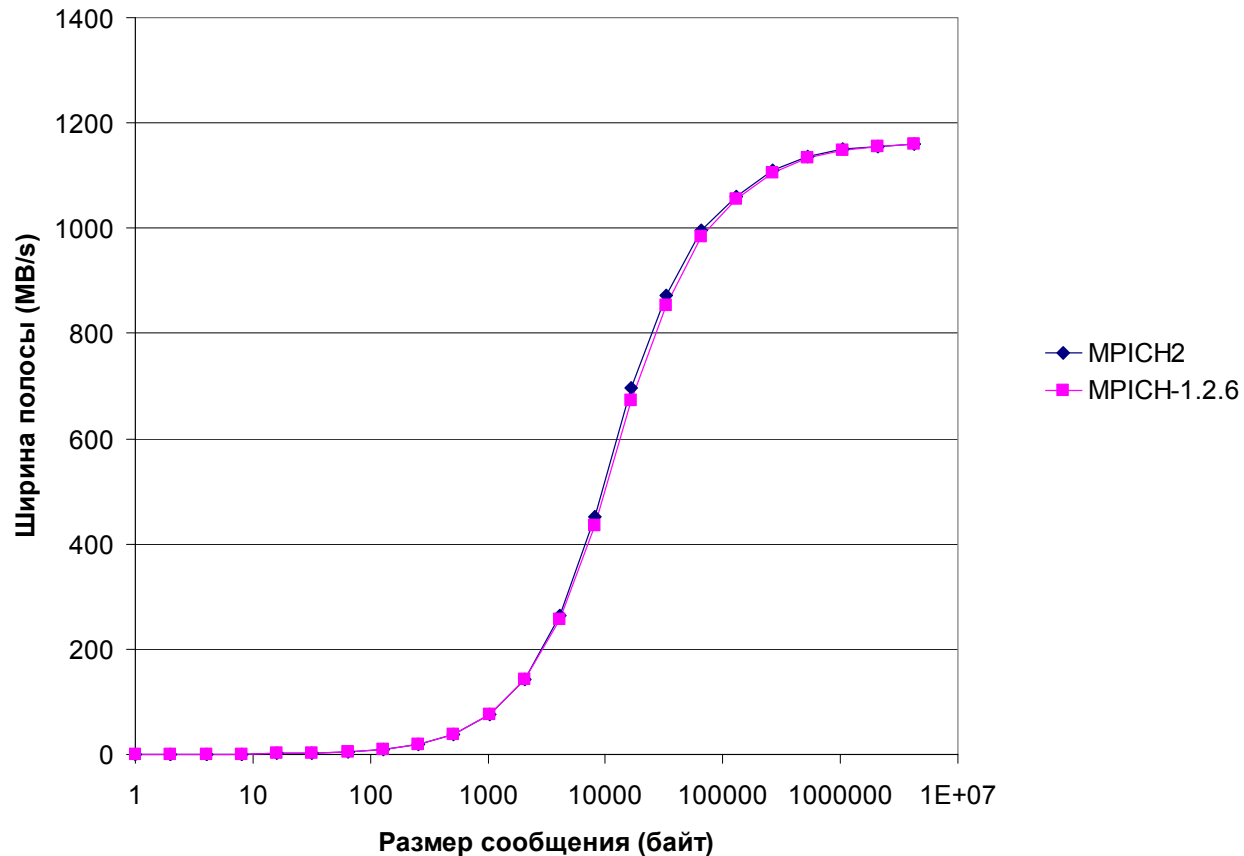
Статус проекта **Red Storm**

- Аппаратные средства.
 - ◆ Система полностью смонтирована и интегрирована.
- Системное программное обеспечение является совместным проектом компаний Cray и Sandia.
 - ◆ Программное обеспечение Sandia Catamount (поддержка выполнения программ и LWK) работоспособно и прошло масштабное тестирование.
 - ◆ В настоящий момент (3/17) возможна загрузка 2x20.
 - ◆ Идет работа над расширением до 3x20 и 3x27.
 - ◆ Ограниченная возможность ввода/вывода – система Lustre не полностью работоспособна.
- Сеть
 - ◆ Встроенные программы порталов находятся в стадии активной разработки.
 - В настоящее время принимает сигнал прерывания на каждом новом сообщении.
 - Время ожидания составляет ~8,5us.
 - Ширина диапазона составляет 1,1 – 1,6 GB/s.

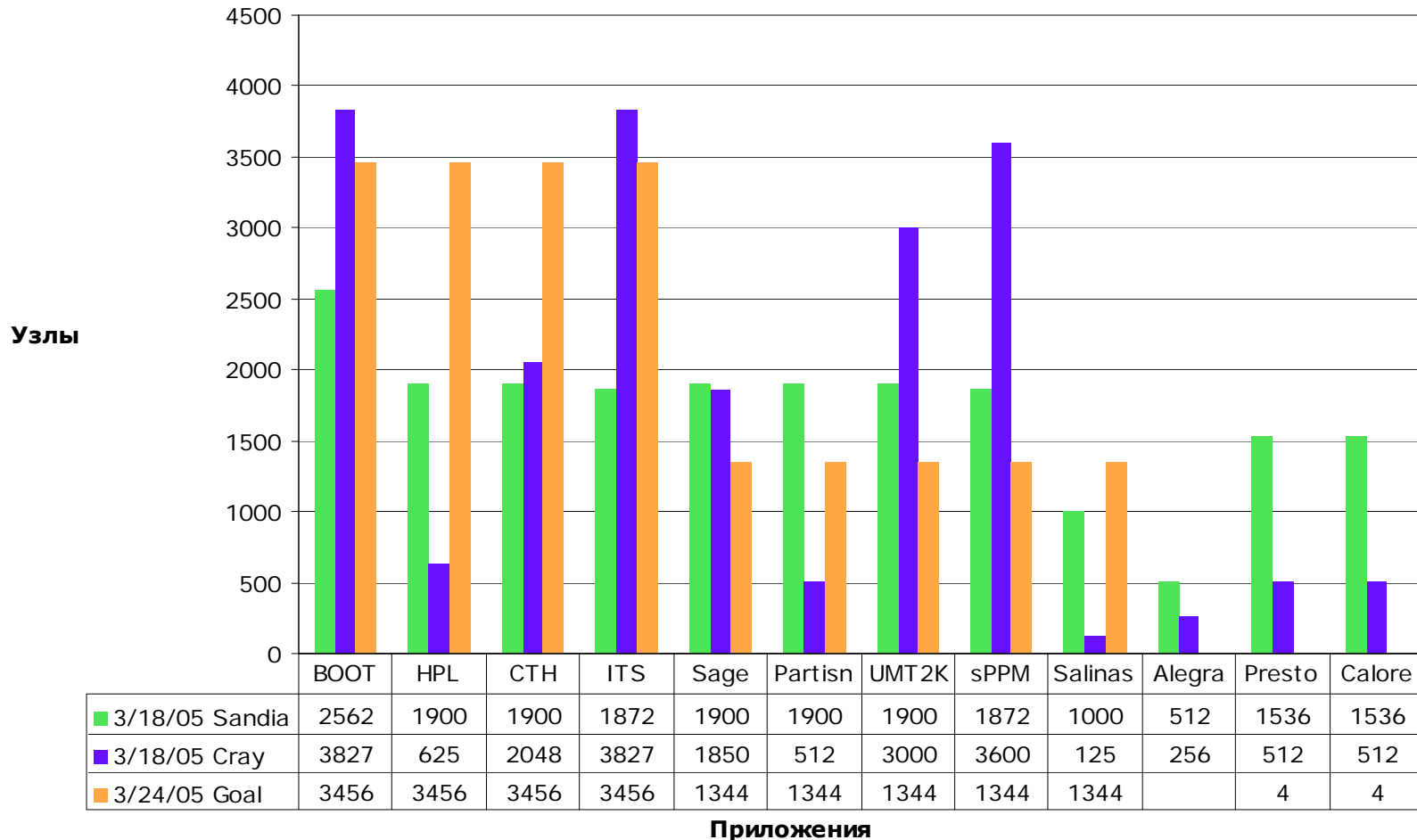
Попарное считывание интерфейса MPI



Попарное считывание интерфейса Pallas MPI



Статус приложений системы **Red Storm**



Статус приложений системы Red Storm

3/24/05 Application Scaling Goal
Sandia / Cray Status as of 4/1/05

