

Высокопроизводительные решения от компании IBM

Андрей Сысоев

Ведущий специалист по продвижению
высокопроизводительных вычислений

**МОСКОВСКИЙ
СУПЕРКОМПЬЮТЕРНЫЙ
ФОРУМ**
2014





АЭРОКОСМИЧЕСКАЯ ИНДУСТРИЯ

Вычислительная
гидродинамика,
строительная
механика



АВТОМОБИЛЕСТ РЮЕНИЕ

Безопасность,
гидродинамика,
акустический анализ

2



БИОТЕХНОЛОГИИ

Цепочки генов,
молекулярное
моделирование,
химические расчеты



ПРОИЗВОДСТВО

Моделирование
электроники,
оптическая коррекция,
тепловая механика

ФИНАНСЫ

Анализ рисков



ИССЛЕДОВАНИЯ

Материаловедение
биотехнологии,
климат



ГОССТРУКТУРЫ

Оборонные
исследования,
перспективная
энергетика,
погода, экология

ЭНЕРГЕТИКА

Обработка
сейсмических
данных, миграция

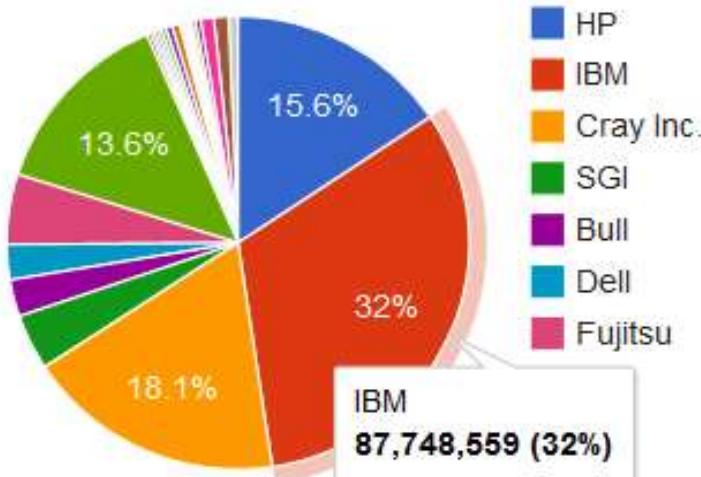
Актуальность суперкомпьютерных технологий

Использование суперкомпьютеров позволяет:

- Частично или полностью заменить дорогостоящие натурные эксперименты компьютерным моделированием
- Наглядно представить те или иные процессы путем их визуализации.
- Сократить сроки разработки опытных образцов и производственные затраты на натурные эксперименты
- Существенно сократить сроки разработки новых технологий, расширить ассортимент и повысить качество готовой продукции
- Сократить себестоимость новых разработок (за счет сокращения сроков реализации проектов и уменьшения доли более дорогого физического прототипирования)
- Достичь существенно большей точности расчетов за прежнее время



Лидирующее положение IBM в суперкомпьютерах



Полугодовые независимые рейтинги 500 самых мощных и энергоэффективных суперкомпьютеров мира



TOP500 энергоэффективных систем

TOP500 совокупная производительность

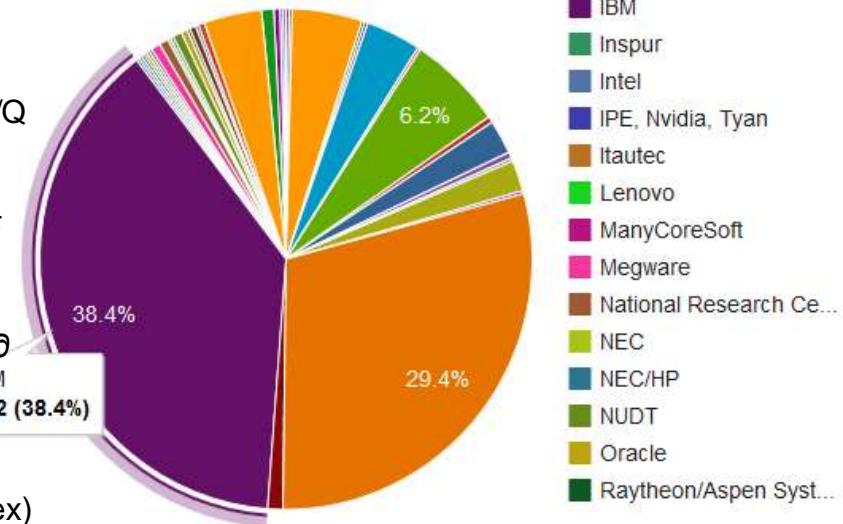
IBM лидирует по ключевым позициям

#3: LLNL Sequoia Blue Gene/Q 20.13 PFlops

- ✓ 4 в TOP 10 (#3 Sequoia BlueGene/Q Livermore, #5 Mira BlueGene/Q Argonne, #8 JUQUEEN BlueGene/Q Juelich, #9 Vulcan BlueGene/Q Livermore)
- ✓ Наибольшая совокупная производительность 58.6PF из 123.4PF (47.5%) (HP 13.5/11%, Fujitsu 12.2 /9.9%, Cray: 10.9/ 8.8%)

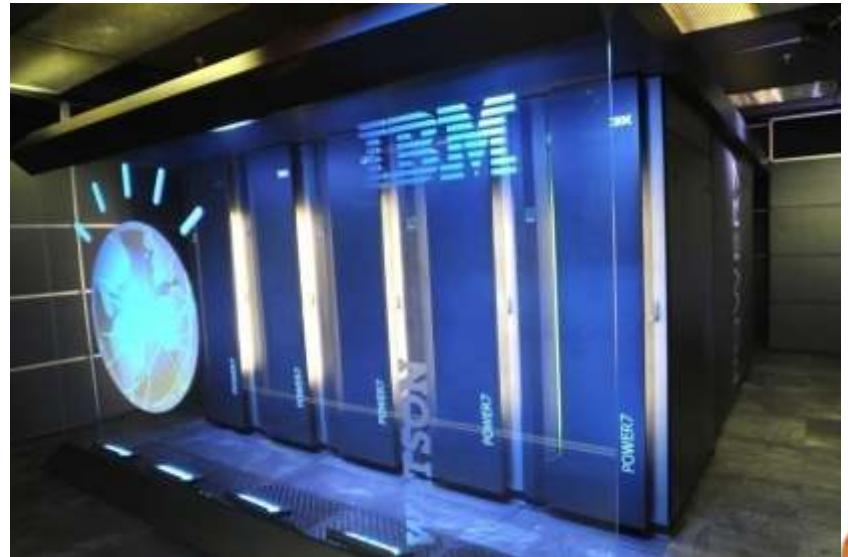
IBM лидирует по этому критерию в последних 27 списках подряд

- ✓ Самый быстрый суперкомпьютер в Европе (LRZ-SuperMUC iDataPlex) 3PF
- ✓ Самый быстрый x86 суперкомпьютер (LRZ-SuperMUC iDataPlex)
- ✓ **20 Наиболее энергоэффективных систем все IBM BG/Q's**



История

- 2012 Суперкомпьютер Blue Gene/Q
- 2011 Watson
- 2008 Первый Petaflop суперкомпьютер Roadrunner
- 2007 Web-scale mining
- 2006 Services Science (SSME)
- 2004 Суперкомпьютер Blue Gene/L
- 2003 Carbon Nanotubes
- 1998 Silicon-on-Insulator
- 1997 Copper Interconnect Wiring
- 1997 Secure Internet Communication
- 1997 Шахматный суперкомпьютер Deep Blue
- 1994 Шаблоны проектирования Design Patterns
- 1994 Silicon Germanium (SiGe)
- 1990 Statistical Machine Translation
- 1987 High-Temperature Superconductivity
- 1986 Scanning Tunneling Microscope
- 1980 Архитектура RISC
- 1971 Распознавание речи Speech Recognition
- 1970 Relational Database
- 1967 Fractals
- 1966 Память One-Device Memory Cell
- 1957 FORTRAN
- 1956 Первый жесткий диск RAMAC
- 1948 Эл.механический компьютер SSEC
- 1944 Первый компьютер Mark 1



IBM Research: Научно-исследовательские центры в мире



Суперкомпьютер Blue Gene/Q



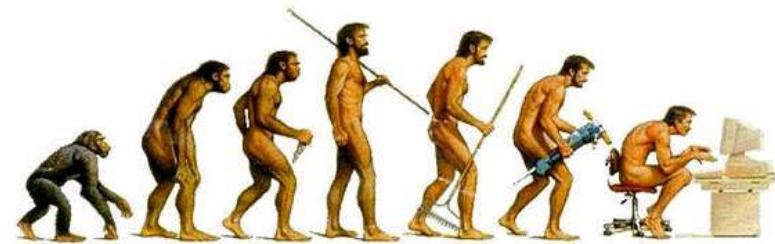
Blue Gene/L (2004)



Blue Gene/P (2007)



Blue Gene/Q (2012)



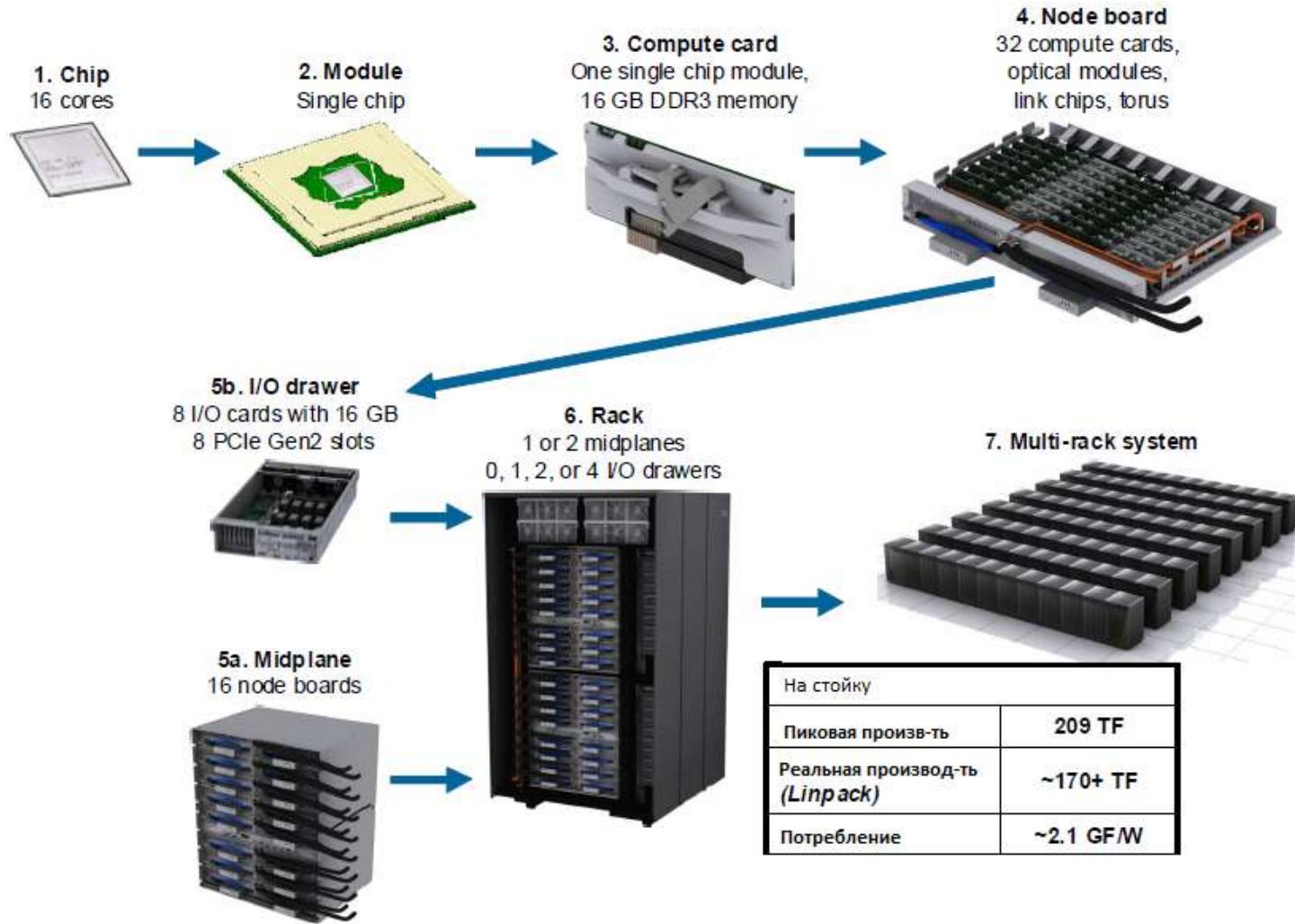
Суперкомпьютер Blue Gene/Q

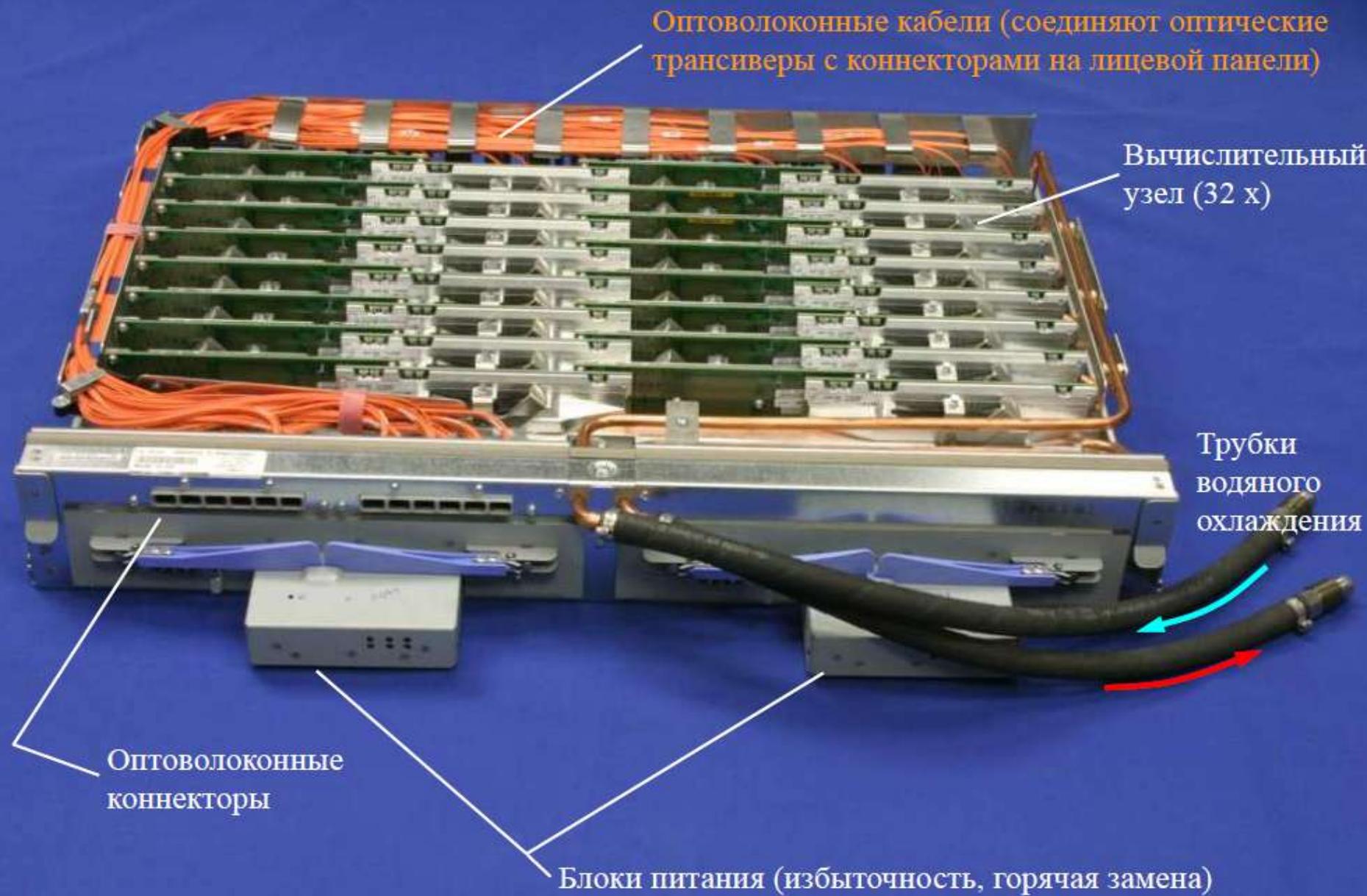
Области применения

- Квантовая химия
- Медико-биологические науки (Life Science)
- Моделирование природных явлений
- Вычислительная динамика жидкости
- Молекулярная динамика
- Астрономия и космические исследования
- Наука о материалах
- Моделирование климата



Суперкомпьютер Blue Gene/Q





IBM NeXtScale nx360 M5



NEXTSCALE

Требования к инфраструктуре:

- Модульная платформа
- Встраивается в имеющуюся инфраструктуру
- С минимальными первоначальными вложениями
- Дальнейшее развитие платформы производителем

Для решения задач:

- Блейд-система начального уровня
- Параллельных вычислений
- Аналитических расчетов
- Облака



IBM NeXtScale nx360 M5



Стандартная стойка



- Форм-фактор 6U для установки в стандартную 19" стойку
- До 12 вычислительных блейд-серверов
- Поддержка последнего поколения процессоров 12 ядерных Intel Ivy Bridge, Haswell
- Высочайшая вычислительная плотность 432 вычислительных ядер в 6U пространства.
- 6x Блоков питания hot-swap N+N или N+1, 80 PLUS Platinum
- 10x Вентиляторов охлаждения hot-swap
- Модуль управления

Все лучшее из iDataPlex стало еще лучше!



Сохраняя хорошее

- Ориентация на производительность
- Фронтальный доступ
- Поддержка акселераторов, сопроцессоров, систем хранения
- Оптимизация за счет внешней коммутации
- Выбор между водой и воздухом



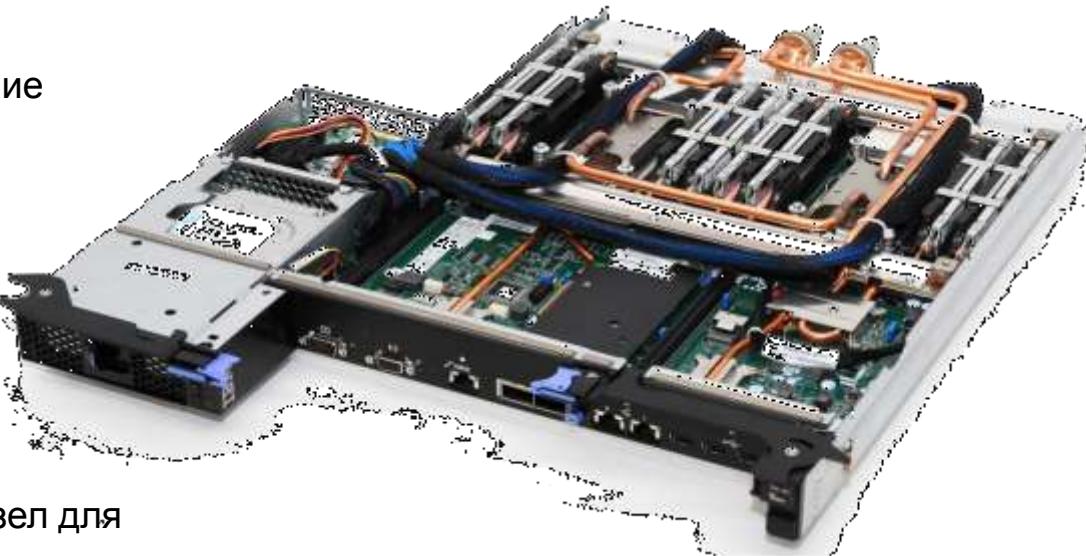
И производя еще лучше

- Более низкая стоимость, большая гибкость шасси
- Стандартные стойки
- Меньшая занимаемая площадь
- Оптимизирован для 36 портового свитча
- Создан для больших рабочих нагрузок, чем просто HPC

IBM NeXtScale nx360 M4 Direct Water Cooling (DWC)

Основные характеристики:

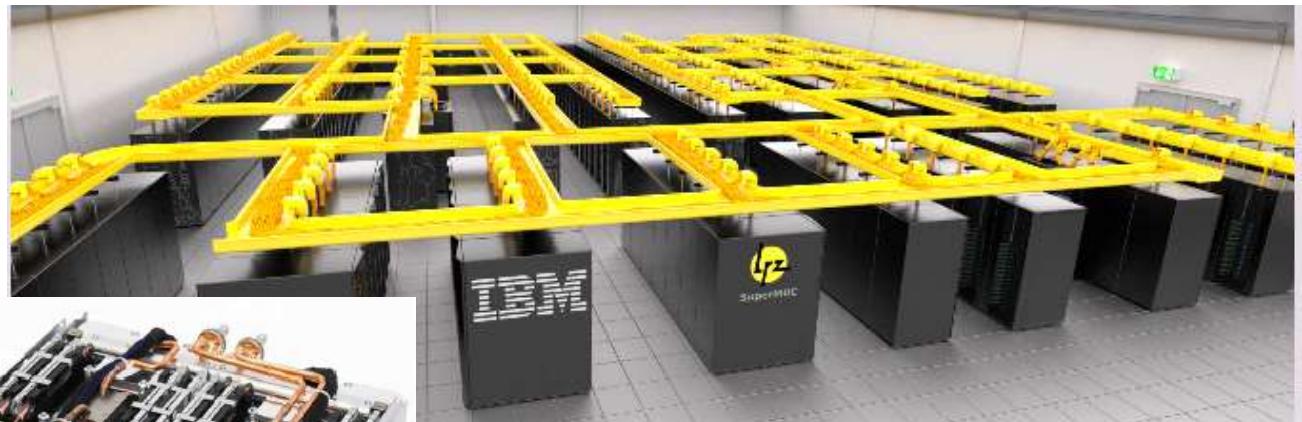
- До 40% более эффективное охлаждение системы по сравнению с воздушным, устраняя 90% тепла из узла
- Обеспечивает до 10% преимущества в энергоэффективности над воздушным охлаждением
- Две водяные охлаждающие петли на узел для охлаждения процессоров и памяти DIMM, остальные компоненты остаются с воздушным охлаждением
- 90% тепла могут быть использованы для экономических целей, таких как отопление других зданий и сооружений
- Рекомендованные условия эксплуатации:
Температура окружающей среды воздуха @ 27-35C и подаваемой воды в узел при 40-45C



3 петафлоп Суперкомпьютер на базе iDataPlex dx360 M4 с прямым водяным охлаждением развертывается в Германии в Суперкомпьютерном центре Лейбница (LRZ)



LEIBNIZ SUPERCOMPUTING CENTRE



iDataPlex dx360 M4

С прямым (горячим)
водяным охлаждением

TOP500 Июнь 2014

R_{max} = 2.897 PFlops

Самая мощная в Европе

#12

* Turbo mode can be sustained because of liquid cooling

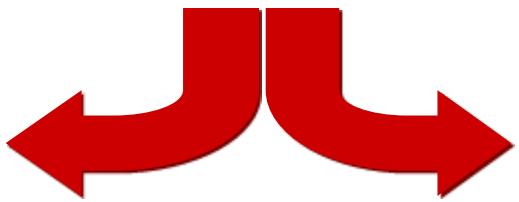
Система	IBM iDataPlex Sandy Bridge
Количество групп	18
Узлов на группу	512
Кол-во сокетов на узле	2
Ядер на сокет	8
Ядер в узле	16
Общее кол-во ядер	147,456
Пиковая производительность (PFlops)	3.185
Linpack производительность (Nominal)	2.582
Linpack производительность (Turbo*)	2.897
Общий объем памяти (TBytes)	288
Памяти на ядро (GBytes)	2
Общей памяти на узел (GBytes)	32

IBM NeXtScale nx360 M5

Воздух



Вода

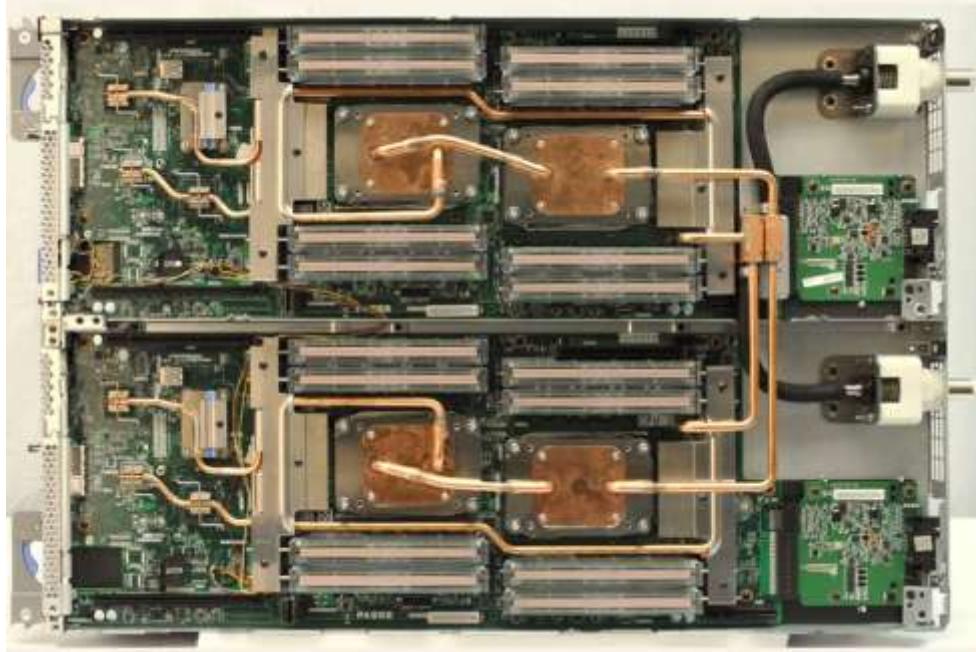


Ваш выбор

- Воздушное охлаждение вентиляторами
- Установка в любые ЦОДы
- Гибкость
- Варианты инсталляции
 - Storage NeX
 - PCI NeX (GPU, Phi)
- Инновационное водяное охлаждение
- Отсутствие вентиляторов
- Экстремально энергоэффективность
- Экстремально тихо
- Низкое энергопотребление
- Плотность, занимаемое пространство
- Низкая стоимость владения

IBM NeXtScale nx360 M5 Direct Water Cooling (DWC)

Вид сверху



Вид спереди

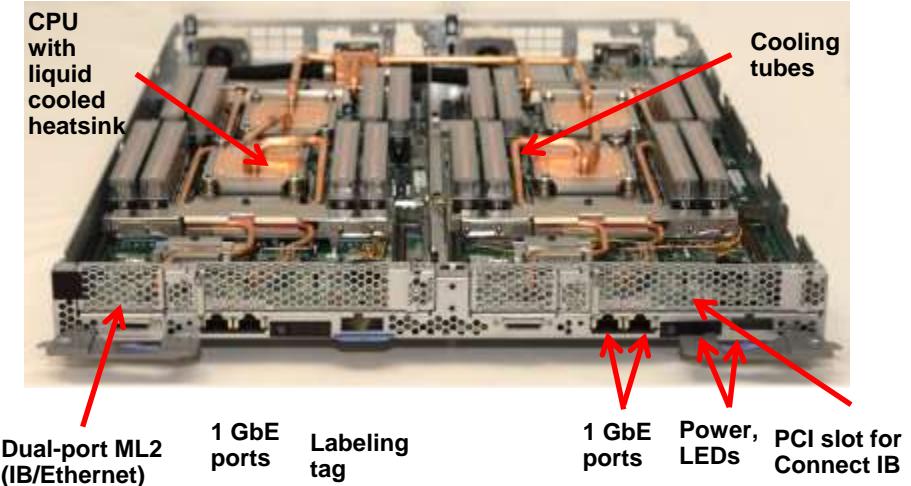


Вид сзади

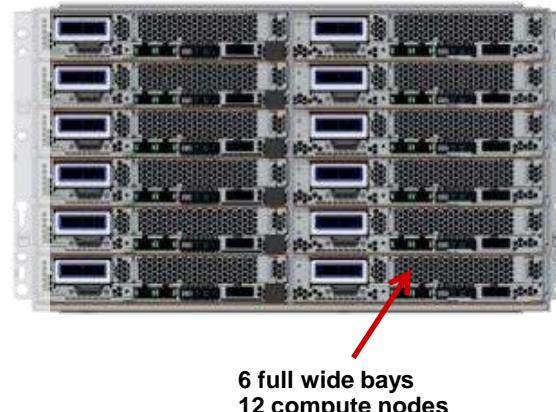
IBM NeXtScale nx360 M5 Direct Water Cooling (DWC)

Water Cool Node & Chassis

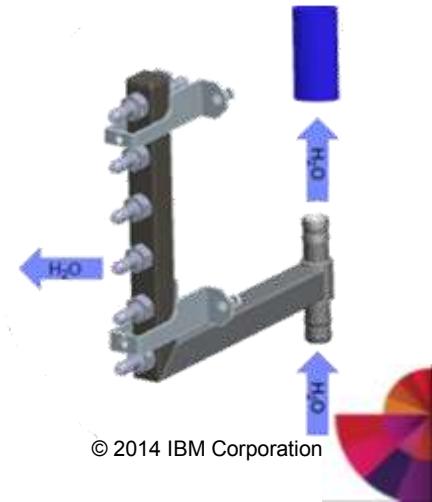
- Полной ширины, 2-узловой вычислительный трэй
- 6U Шасси, 6 слотов (12 узлов/шасси)
- Трубопровод доставляет воду непосредственно в узел
- Вода циркулирует через охлаждающих трубок для охлаждения уровня компонента
- Intel E5-2600 v3 CPUs
- DDR4 DIMM slots
- Infiniband FDR (ML2 or PCIe)
- 6x 900W или 1300W PSU
- Без вентиляторов за исключением БП
- Сенсор протечки / LED ошибок



n1200 WCT Шасси



n1200 WCT Трубопровод



IBM Platform HPC + NeXtScale nx360 M5

Ключевые возможности

- Широкие возможности настройки и простой в использовании
- Удобный в планировании нагрузки
- Web-интерфейс для доступа из любой точки
- Поддержание нагрузки, система мониторинга, отчетность
- Коммерческие библиотеки MPI

Преимущества для клиентов

- Быстрый выход на готовности системы
- Время получения результатов
- Короткое время обучения пользователей
- Всесторонний контроль над кластером

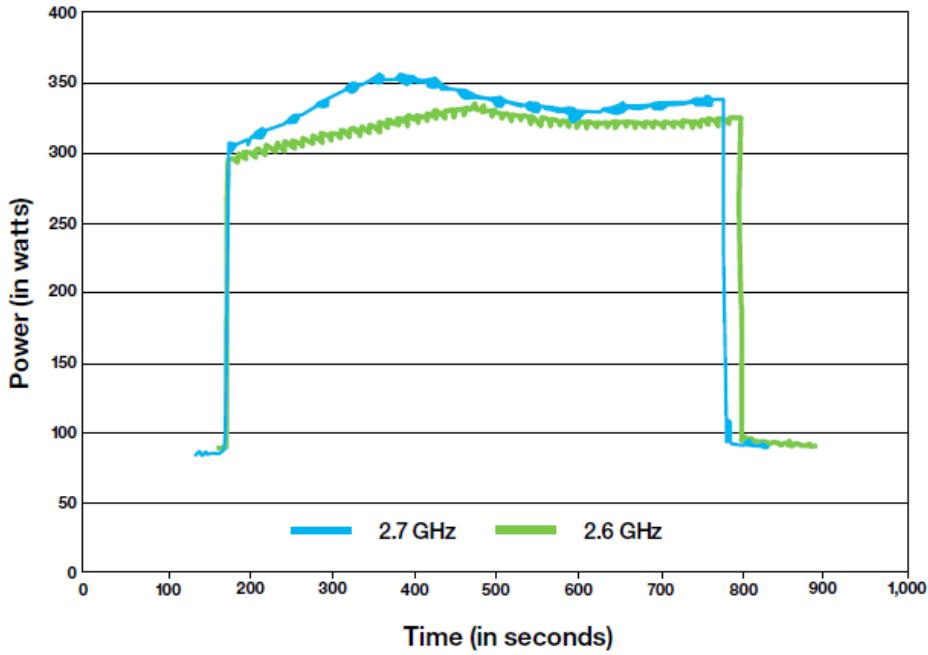
Легкий в использовании интерфейс:

Единый интерфейс для развертывания, управления и использования



Energy Aware Scheduling (EAS): изменение частоты

Quantum ChromoDynamics Application



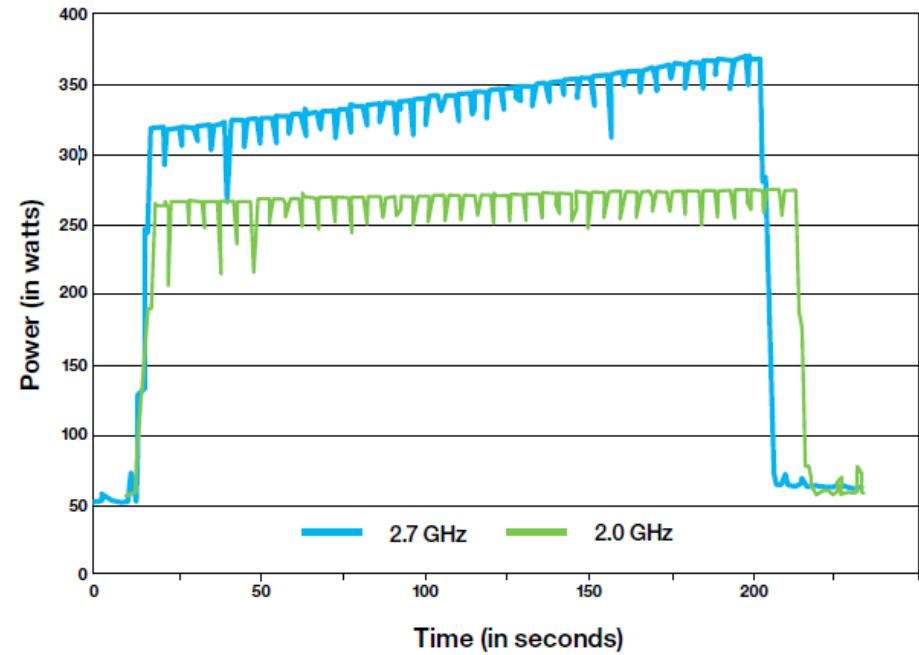
$\Delta f = 2.6\text{GHz}$ ($\Delta f = -4\%$)

$\Delta\text{Power} = -5\%$

$\Delta\text{Time} = +2\%$

$\Delta\text{Energy} = -3\%$

Astrophysics Application



$\Delta f = 2.6\text{GHz}$ ($\Delta f = -26\%$)

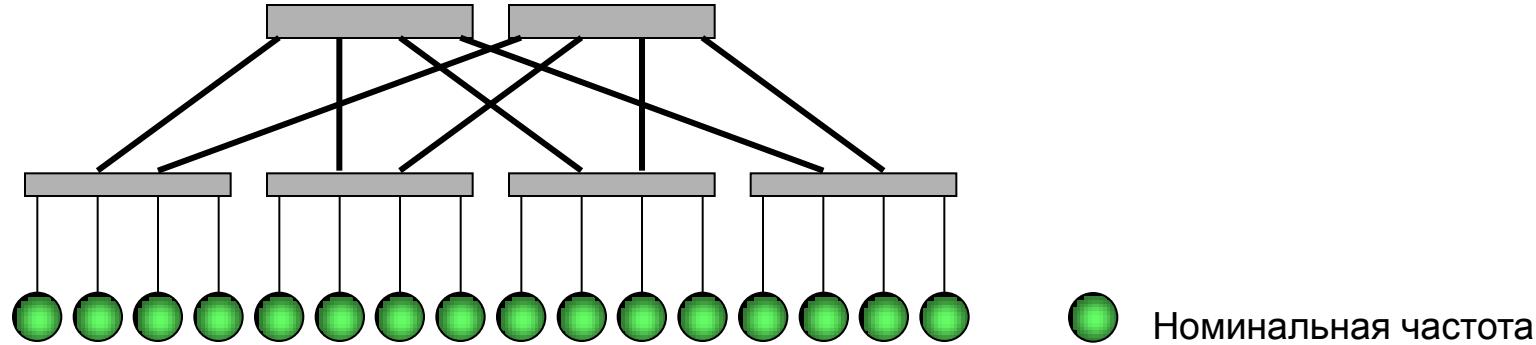
$\Delta\text{Power} = -17\%$

$\Delta\text{Time} = +5\%$

$\Delta\text{Energy} = -12\%$

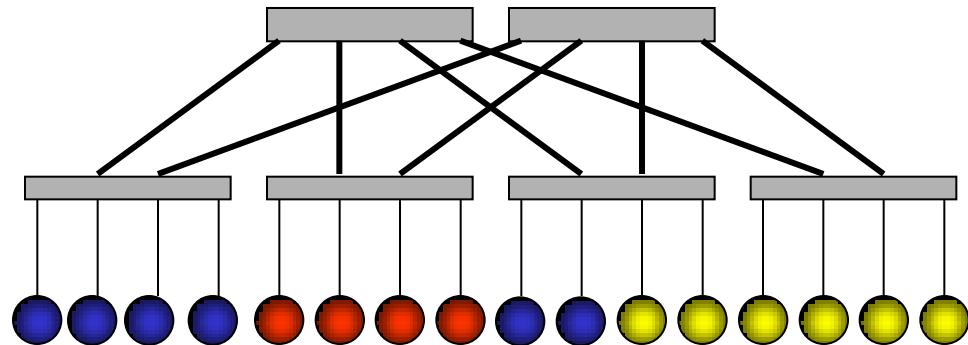
Energy Aware Scheduling (EAS)

Текущее планирование заданий



● Номинальная частота

Планирование заданий с учетом потребления электроэнергии



- Частота для задания 1
- Частота для задания 2
- Узлы в энергосберегающем состоянии

X6: Шестое поколение технологии EXA



Flex System x880 X6



System x3950 X6



System x3850 X6

Fast.

Инновационный eXFlash =
**43% ниже стоимость
решения**

3x больше оперативной памяти,
50% больше ядер процессора =
лидирующая прозво-
дительность с базами данных

Масштабируемость

Agile.

Оптимизированные конфигурации
= **28% ниже стоимость
приобретения**

Легка в конфигурации и сервисе

Поддержка на протяжении
нескольких поколений

Resilient.

**Надежность корпоративного
класса**

CPU, Memory, I/O гибкость

**Интегрированное управление
Hypervisor**

X6: Шестое поколение технологии EXA x3850 X6

Compute Book

- Intel Xeon E7 Family Processor (E7-48xx, E7-88xx)
- 24 DDR3 DIMMs, всего до 6TB
- Up to 8 Flash based DIMMs
- 2 x Hot-Swap Fans (removable from module)



4x Compute Books

Storage Book

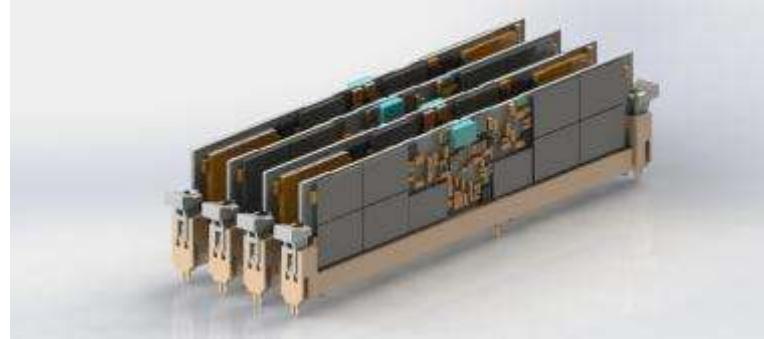
- 8 x 2.5" HDD/SSD
- 16 x 1.8" eXFlash SSD
- 2 Storage Controllers
- SAS 12Gbps



X6: Шестое поколение технологии EXA x3850 X6



- До 8 процессоров на систему
- До 12TB оперативной памяти DDR3

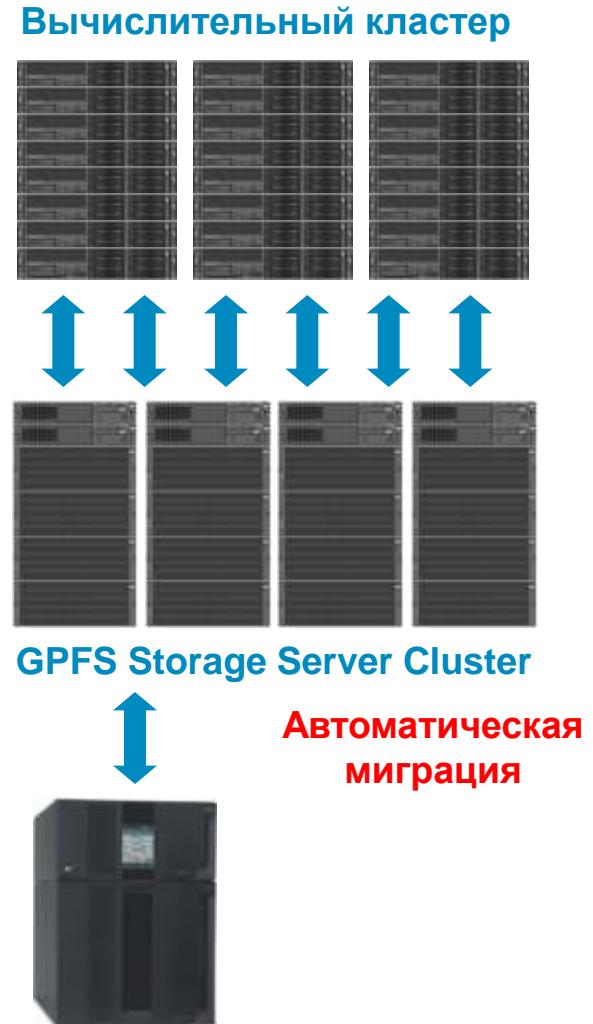


eXFlash DIMM

- До 30% ниже латентность в сравнении с PCIe Flash
- До 32 шт 200GB DDR3 Storage DIMM
- До 32 шт 400GB DDR3 Storage DIMM

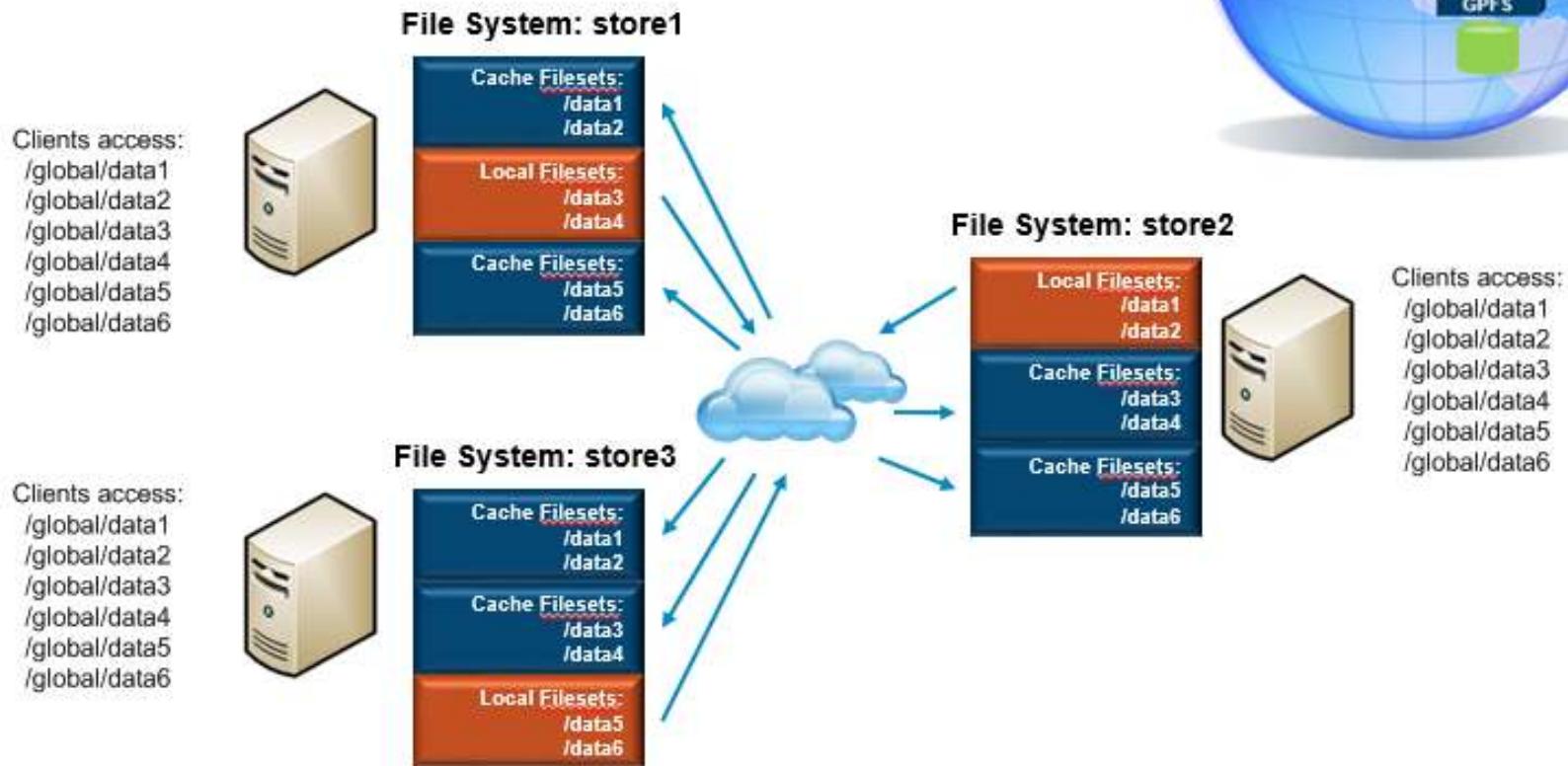
GPFS эффективно управляет всем жизненным циклом данных

- Автоматизация управления данными на основе политик
- Выделение отдельных пулов:
 - SSD диски
 - Высокоскоростные SAS диски
 - SAS диски большой емкости
 - ...
- Соотношение ценности данных к стоимости дисков
- Интеграция с ленточными системами хранения



GPFS обеспечивает глобальный доступ к данным

Доступ к данным из любой точки мира
благодаря технологии распределенного
кэширования Active File Management (AFM)



Параллельная система хранения данных IBM System x GPFS Storage Server (GSS)

Ключевые возможности

- 2 x Управляющих серверов IBM System x3650,
- DCS3700 with Disk (2TB or 3TB NL-SAS, SSD), Software,
- InfiniBand / Ethernet
- IBM Declustered RAID (GPFS-Native RAID)
- **Отсутствие Storage Controllers!!!**



JBOD
Disk Enclosures

GSS 24: Light and Fast
2x3650 servers +
4xJBOD 20U rack
10 GB/Sec

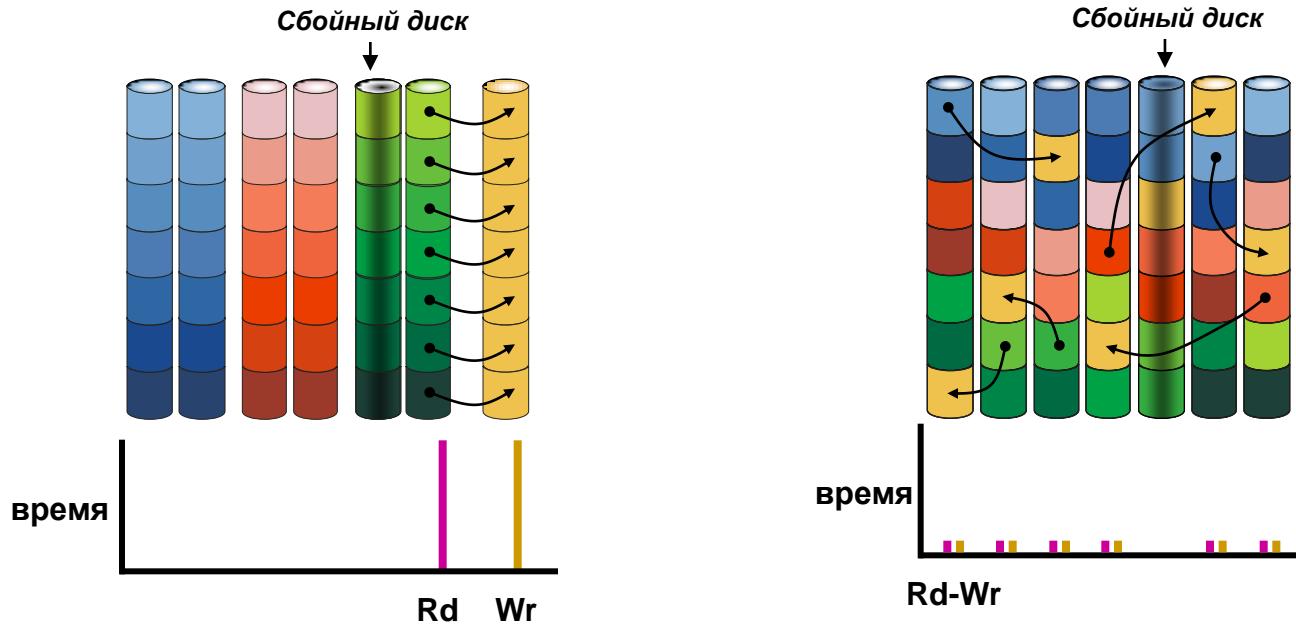


GSS 26: HPC Workhorse
2x3650 servers +
6xJBOD Enclosures, 28U
12 GB/sec



High-Density HPC Option
6x3650 servers + 18xJBOD
2x-42U Standard Racks
36 GB/sec

Параллельная система хранения данных IBM System x GPFS Storage Server (GSS)



- GSS GPFS Native RAID (De-clustered RAID)
- Обработка одновременного отказа до 3-х дисков, контрольные суммы
- Поддержка замены дисков без остановки системы хранения
- Мгновенные снимки, репликация, средства журналирования, мониторинг состояния файловой системы

Комплексное решение Intelligent Cluster

- Предварительное заводское тестирование на отказоустойчивость и совместимость компонентов.
- Доставляется на площадку в собранном и укомплектованном виде, остается лишь подключить питание и приступить к работе.
- Компоненты установлены, скаблированы, а кабели промаркованы в соответствии с кабельным журналом и схемой коммутации.

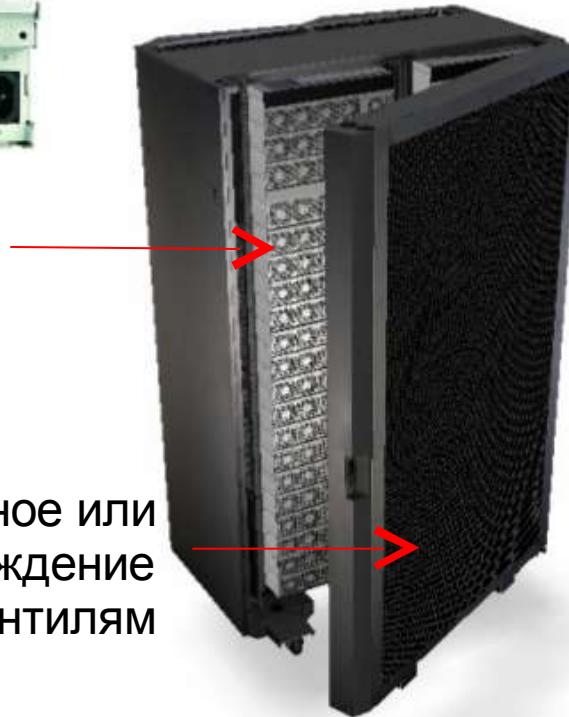


Rear Door Heat eXchanger (Теплообменник в задней двери стойки)

Задняя панель



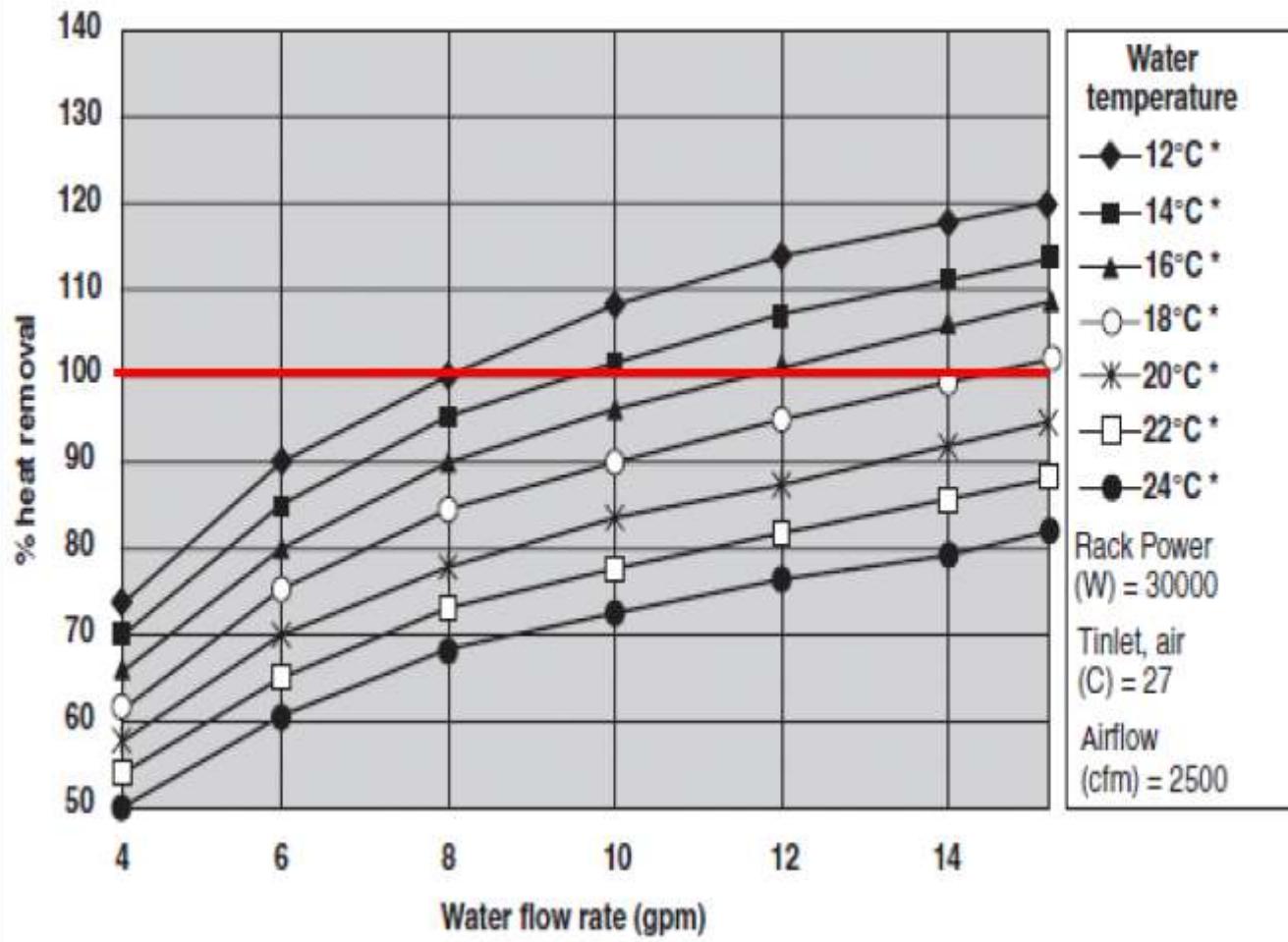
Панели вентиляторов
и кабели питания



Дополнительное воздушное или
пассивное водяное охлаждение
с подключением к стандартным вентилям



Rear Door Heat eXchanger (Теплообменник в задней двери стойки)

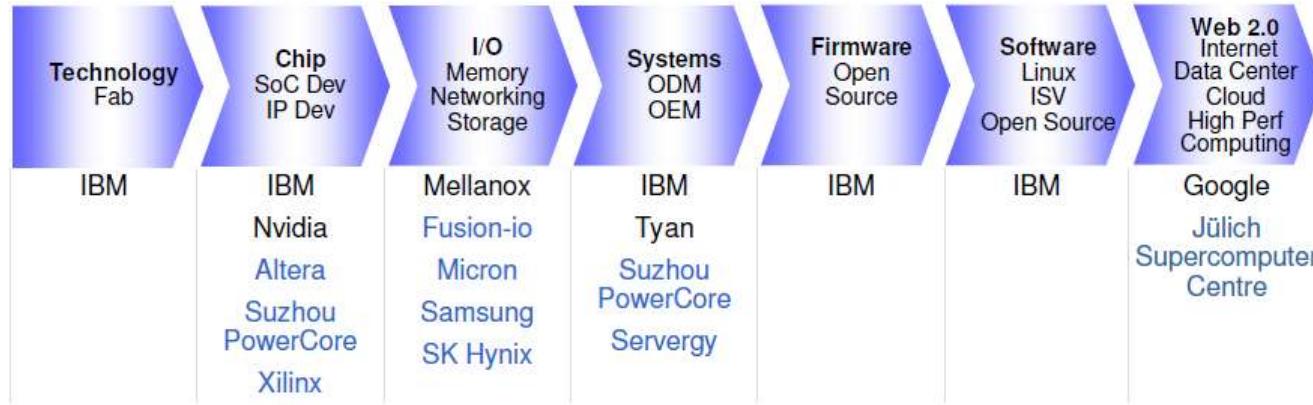


Переход System X в Lenovo

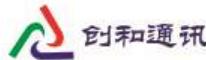


- 29 сентября 2014 одобрение регулирующих органов
- Lenovo становится поставщиком № 3 (x86 аппаратных средств, ПО, услуги)
- 6,500 сотрудников
- 60 стран мира
- Рынок ~\$225 миллиардов
- Фазы операций трансформации в Lenovo начинаются с 1-го октября.

OpenPOWER Foundation



Gold Members

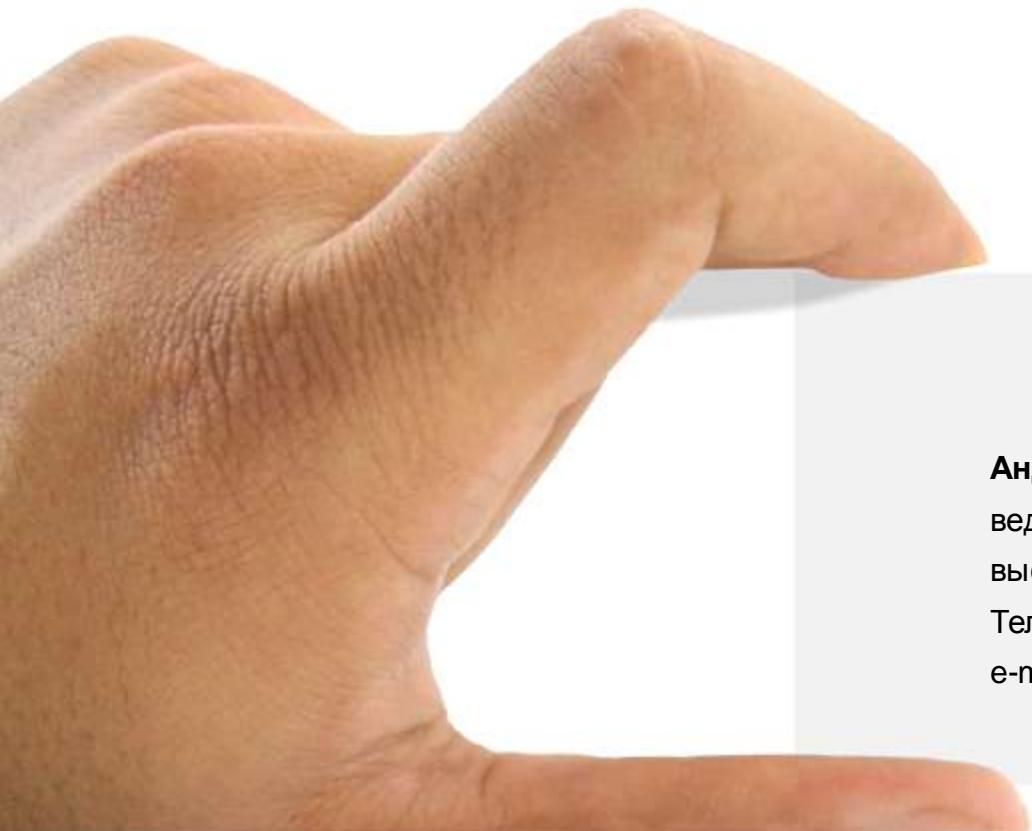
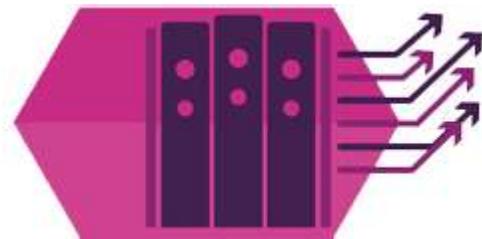


Sichuan Huaxun
Zhongxing
Technologies



Wistron





Спасибо



Андрей Сысоев

ведущий специалист по продвижению
высокопроизводительных вычислений

Тел: +7 (906) 782 22 90

e-mail: Andrey.Sysoev@ru.ibm.com

Эпохи вычислений

Эпоха Когнитивных
Систем



Эпоха
Программируемых
Систем

Эпоха
Счетно-
перфорационных
Систем

