СУПЕРКОМПЮТРИ

ТЕХНОЛОГИИ И АРХИТЕКТУРИ

ПРОФ. ПЛАМЕНКА БОРОВСКА

СУПЕРКОМПЮТРИ

- Суперкомпютрите са най-бързите компютърни системи в света при дадено ниво на развитието на компютърните технологии
- През 60-те години на 20ти век водещ проектант на суперкомпютри е *Seymour Cray в Control Data Corp. (CDC),* като неговите суперкомпютри са доминиращи на пазара
- През 70-те години Cray напуска компанията и основава собствена компания, *Cray Research*
- Seymour Cray превзема пазара на суперкомпютри с новите си архитектурни и технологични решения през периода 1985–1990 г.
- През 80-те години голям брой по-малки производители навлизат в пазара на суперкомпютри, повечето от които фалират в средата на 90-те години по време на колапса на пазара на суперкомпютри (supercomputer market crash)

Терминът "суперкомпютър"

- По своята същност терминът суперкомпютър е доста обтекаем, поради тенденцията съвременният суперкомпютър в близко бъдеще да се превръща в обикновен компютър
- Ранните машини на CDC просто са представлявали много бързи скаларни процесори, няколко десетки пъти по-бързи от най-добрите машини, предлагани от другите производители
- През 70-те години повечето суперкомпютри разчитат на векторната обработка, и много от новите играчи на пазара разработват векторни процесори с по-ниска цена за да се приобщат към пазара на суперкомпютри
- В средата на 80-те години суперкомпютрите разполагат с ограничен брой векторни процесори, работещи паралелно. Типичният брой на процесорите е в обхвата 4 до 16.
- През 80-те и 90-те години векторните процесори са заменени с масивно паралелни системи, обхващащи стотици обикновени процесори

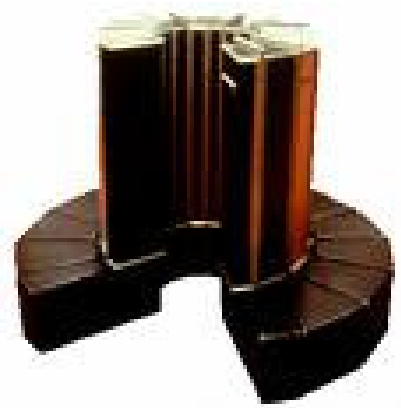
Производителност на суперкомпютрите

- Производителността на компютърните системи отразява скоростта на информационната обработка и се измерва в брой изчислителни операции, изпълнени за единица време
- Най- популярната мярка при суперкомпютрите е брой операции с плаваща точка за секунда FLOPS (Floating Point Operations Per Second)
  - Производителността на компютърните системи зависи основно от технологията и архитектурата, но съществено се влияе и от други фактори, като В/И система, пропускателната способност на системната комуникационна мрежа, и не на последно място от приложението

Развитие на суперкомпютрите

- CDC 6600 (1964, 3MFlops), CDC 7600 (1969, 36 MFlops), CDC STAR-100 (1974, 100 MFlops) – Lawrence Livermore National Lab., California, USA
- Burroughs ILLIAC IV (1975, 150 MFlops) – NASA Ames Research Center, California, USA
- Cray-1 (1976, 250 MFlops) – Energy Research and Development Administration, Los Alamos National Laboratory, New Mexico, USA
- CDC Cyber 205 (1981, 400 MFlops) – numerous sites worldwide

Развитие на суперкомпютрите



Cray 1, 1976
250 MFlops

Развитие на суперкомпютрите

- Cray X-MP/4 (1983, 941 MFlops) – U.S. Department of Energy (DoE), Lawrence Livermore National Lab., Los Alamos National Lab., Boeing
- Cray-2/8 (1985, 3.9 Gflops) – DoE - Lawrence Livermore National Laboratory
- ETA10-G/8 (1989, 10.3 GFlops) – Florida State University, USA
- Thinking Machines CM-5/1024 (1993, 65.5 GFlops) – DoE – Los Alamos National Lab., National Security Agency

Cray-2/8, най-бързият суперкомпютър от 1985 до 1989

Развитие на суперкомпютрите

- Intel Paragon XP/S 140 (1992, 143.40 GFlops) – DoE – Sandia National Laboratories, New Mexico, USA
- Fujitsu Numerical Wind Tunnel (1994, 170.40 GFlops) – National Aerospace Lab., Tokyo, Japan
- Hitachi SR2201/1024 (1996, 220.4 GFlops) – University of Tokyo, Japan

Развитие на суперкомпютрите

- Intel ASCI Red/9152 (1997, 1.338 TFlops) - DoE – Sandia National Laboratories, New Mexico, USA – в колаборация с Intel Corp.
- Intel ASCI Red/9632 (1999, 2.3796 TFlops) - DoE – Sandia National Laboratories, New Mexico, USA
- IBM ASCI White (2000, 7.226 TFlops) – DoE, Lawrence Livermore

IBM ASCI WHITE

ASCI RED най-бързият суперкомпютър през 1997-2002г.

Развитие на суперкомпютрите

- NEC Earth Simulator (2002, 35.86 TFlops) – Earth Simulator Center, Yokohama, Japan
- IBM Blue Gene/L (2004-2007, 478.2 TFlops)- DoE/ U.S. National Nuclear Security Administration, Lawrence Livermore Lab.
- IBM Roadrunner (2008, 1.026-1.105 PFlops) – DoE – Los Alamos National Lab., New Mexico, USA

Earth Simulator 2002-2004r.
NEC – **NEC Corporation**
日本電気株式会社

Суперкомпютър "Синият ген"

Фирмата IBM е наградена с Национален медал за технологии и иновации през 2008г. за фамилията суперкомпютри IBM Blue Gene

Суперкомпютър "Синият ген"

- *Фирмата IBM е наградена с Национален медал за технологии и иновации през 2008 г., за фамилията суперкомпютри IBM Blue Gene*

Съвременни суперкомпютри

- "Традиционни производители" – компании като Cray, IBM и Hewllet-Packard, които са изкупили много от суперкомпютърните компании през 80-те години за да използват техния опит
- През май 2008, суперкомпютърът IMB Roadrunner, в Los Alamos National Laboratory, е обявен за най-бързият суперкомпютър в света – 1.026 petaflop/s
- Паралелните архитектури на съвременните суперкомпютри се базират на микропроцесори за комерсиалния клас сървъри (*server-class microprocessors)*, като PowerPC, Opteron, или Xeon, и по същество мнозинството от съвременните суперкомпютри представляват *компютърни клъстери* с популярни процесори комбинирани със специализирана и специално оектирана за целта висикоскоростна системна мрежа

Суперкомпютърът Roadrunner

- Първата универсална компютърна система, която достига производителност от порядъка на petaflop milestone (10^{15} FLOPS измерени при еталона Linpack)
- Проектирана, произведена, и тествана в подразделението на IBM в Rochester, Minnesota
- Финална дестинация – Los Alamos National Lab., New Mexico
- Roadrunner е най-новото средство, което се използва от National Nuclear Security Administration (NNSA) за осигуряването надеждността и безопасността на ядрения арсенал на САЩ

Roadrunner е клъстер от клъстери

- Основният градивен блок е Connected Unit (CU) – общ брой 18
- Roadrunner е изграден от приблизително 6500 AMD двуядрени процесори, комбинирани с 12 240 Cell Broadband Engine (Cell) процесори
- Общата върхова (теоретична) производителност надхвърля 1.3 PFlops
- Капацитетът на паметта е 98 TB равномерно разпределени между Opteron-базираните и Cell-базираните възли
- Всеки блок CU съдържа 180 изчислителни възела и 12 В/И възела

Roadrunner

- Цялата система обхваща 278 шкафа
- Общо тегло – 250 tons
- Системните връзки използват 55 мили Infiniband кабели
- Системата консумира 2.4 MW като осигурява 437 megaflops за watt

Roadrunner

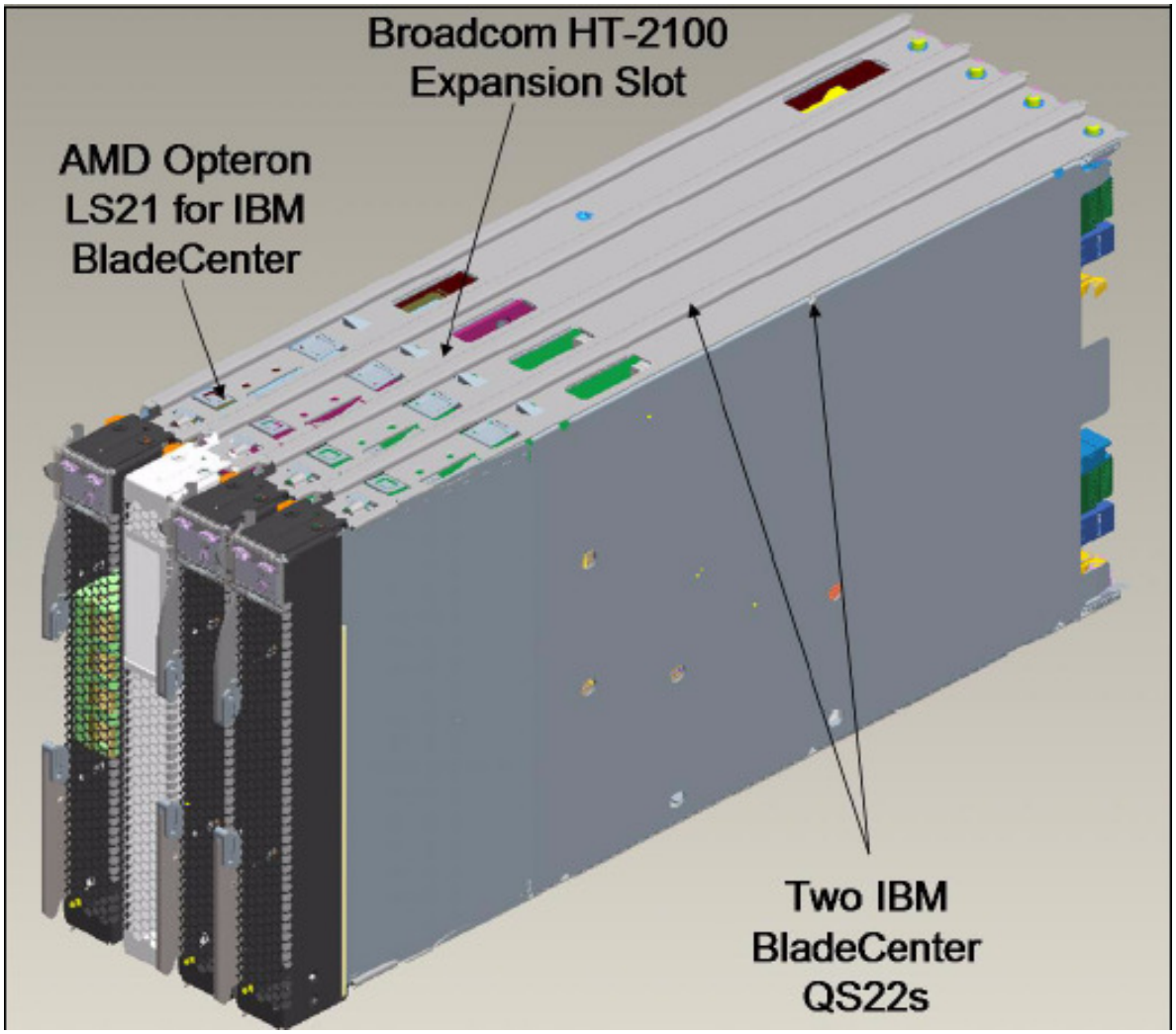
Roadrunner

Roadrunner

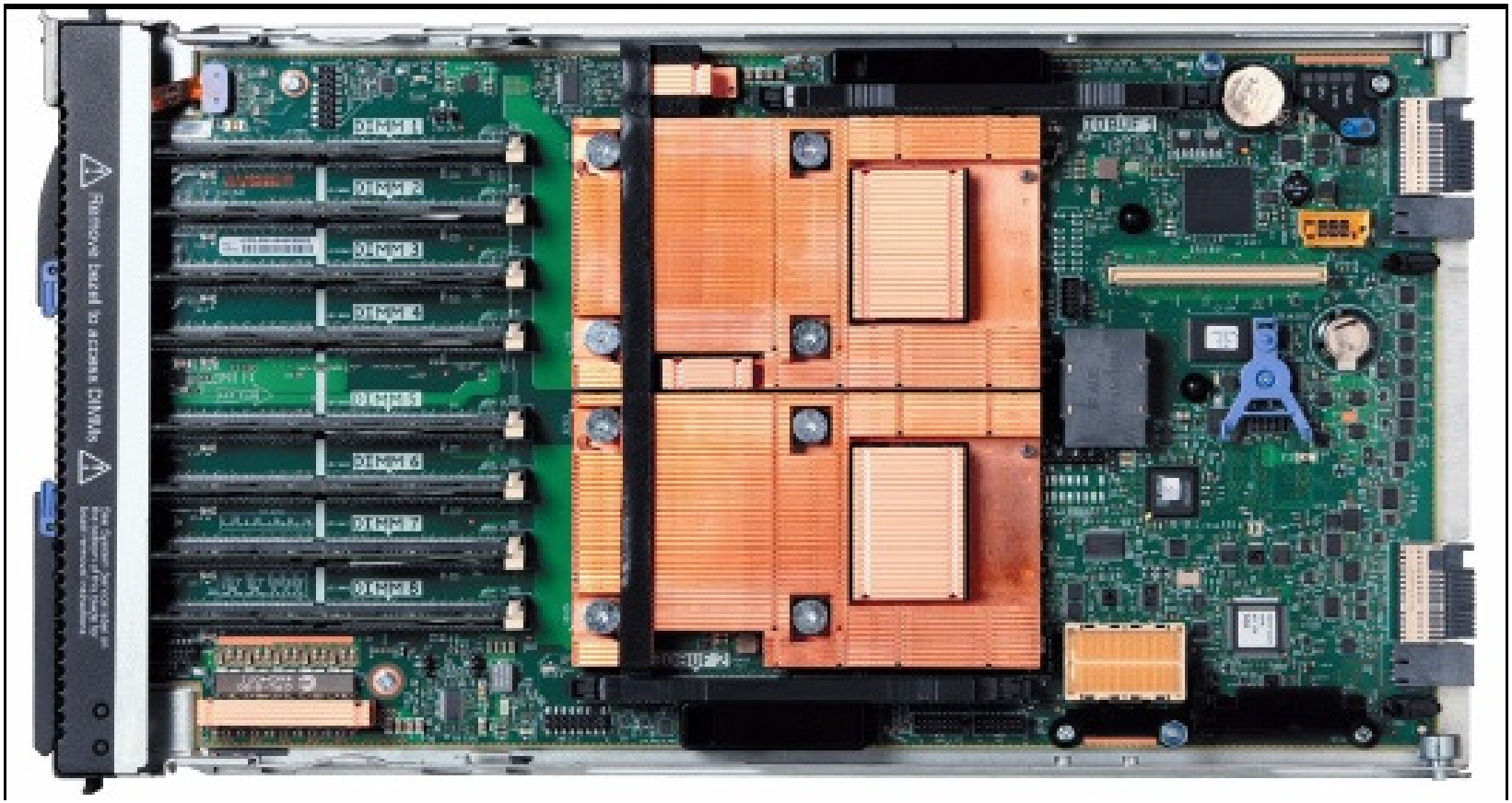
TriBlade: уникална концепция

- TriBlade е основният градивен блок на CU
- Всеки TriBlade модул съдържа AMD Opteron blade и 2 Cell IBM BladeCenter QS22 blades
- Opteron blade съдържа 2 двуядрени процесора
- Всеки от Cell blades съдържа 2 Cell eDP (с двойна точност) процесори
- Всяко Opteron ядро е свързано с чип Cell посредством специализирана PCIe връзка
- Комуникациите между възлите Opteron и процесорите Cell се осъществява посредством Infiniband комуникационна мрежа
- Операционна система – Fedora Linux
- Системно управление на клъстера от клъстери - xCAT cluster management software tools

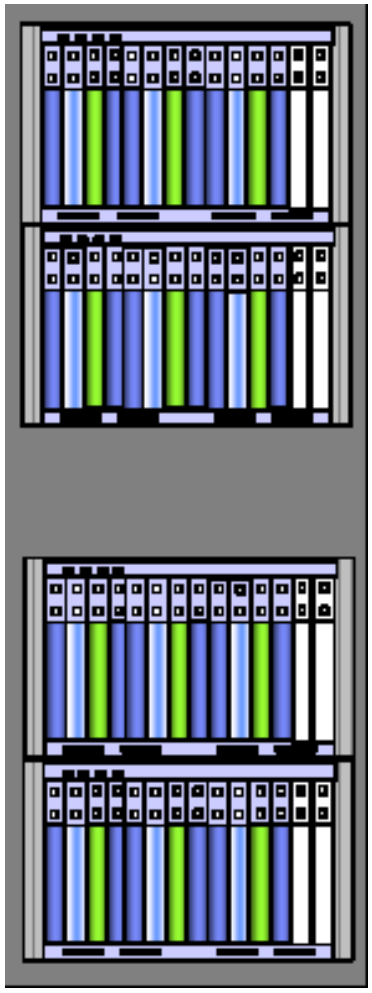
TriBlade

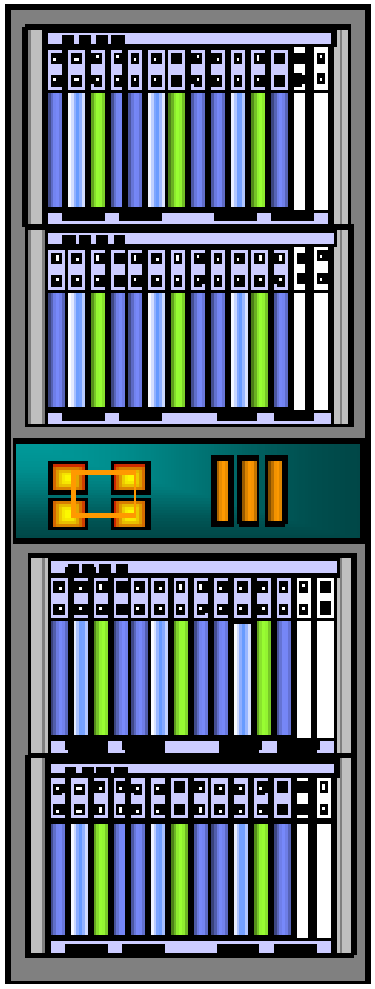


IBM BladeCenter QS22

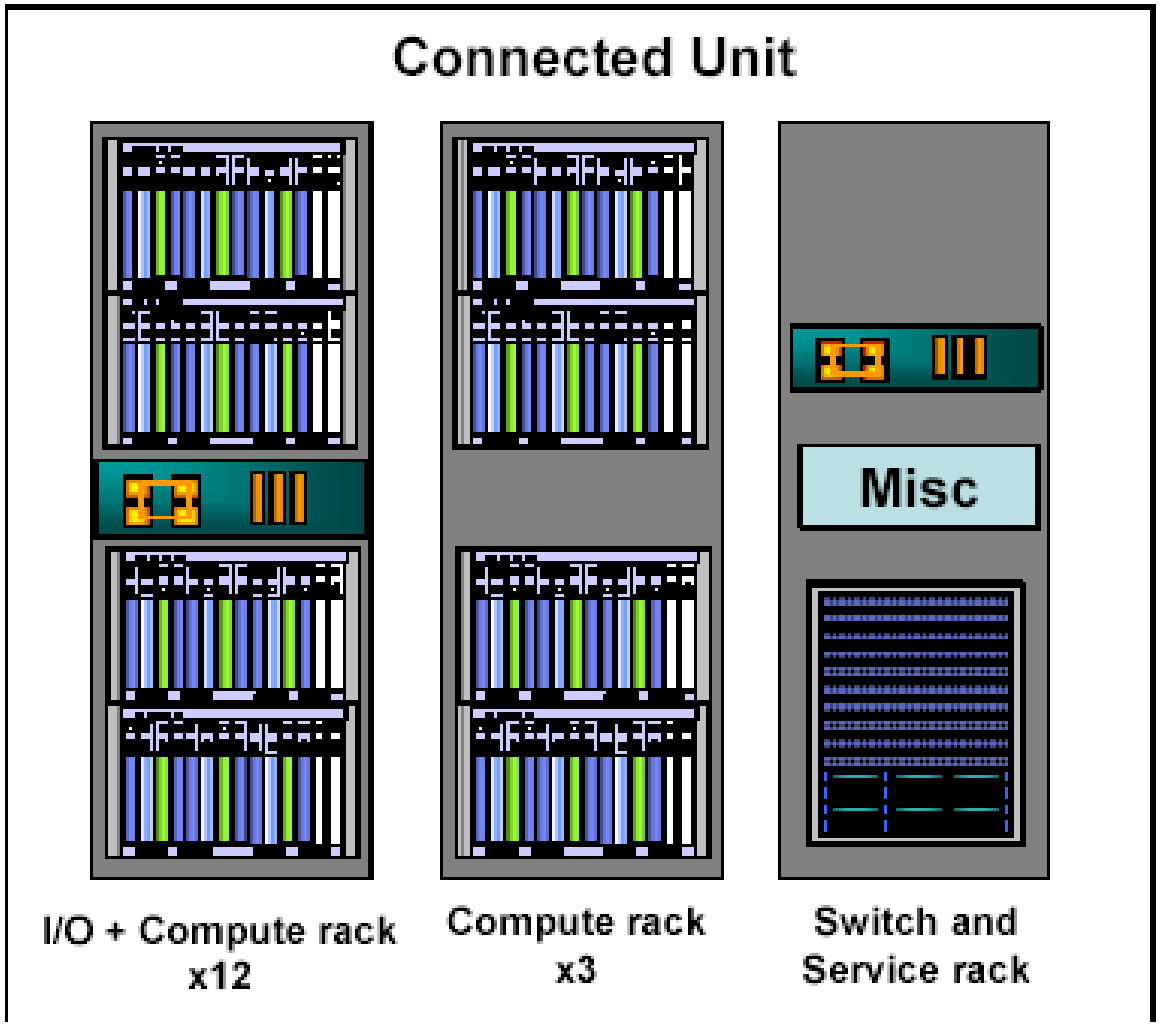


Изчислителни и В/И шкафове

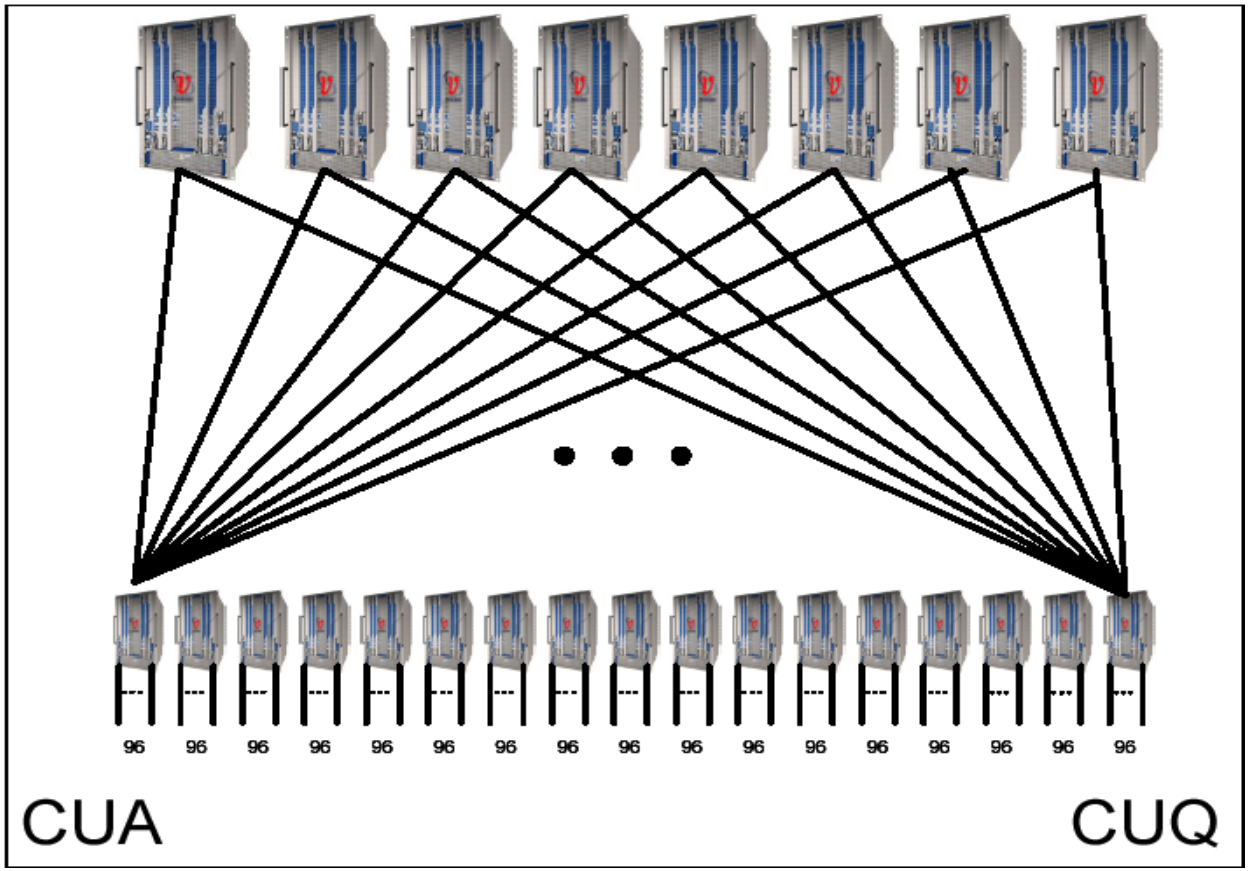




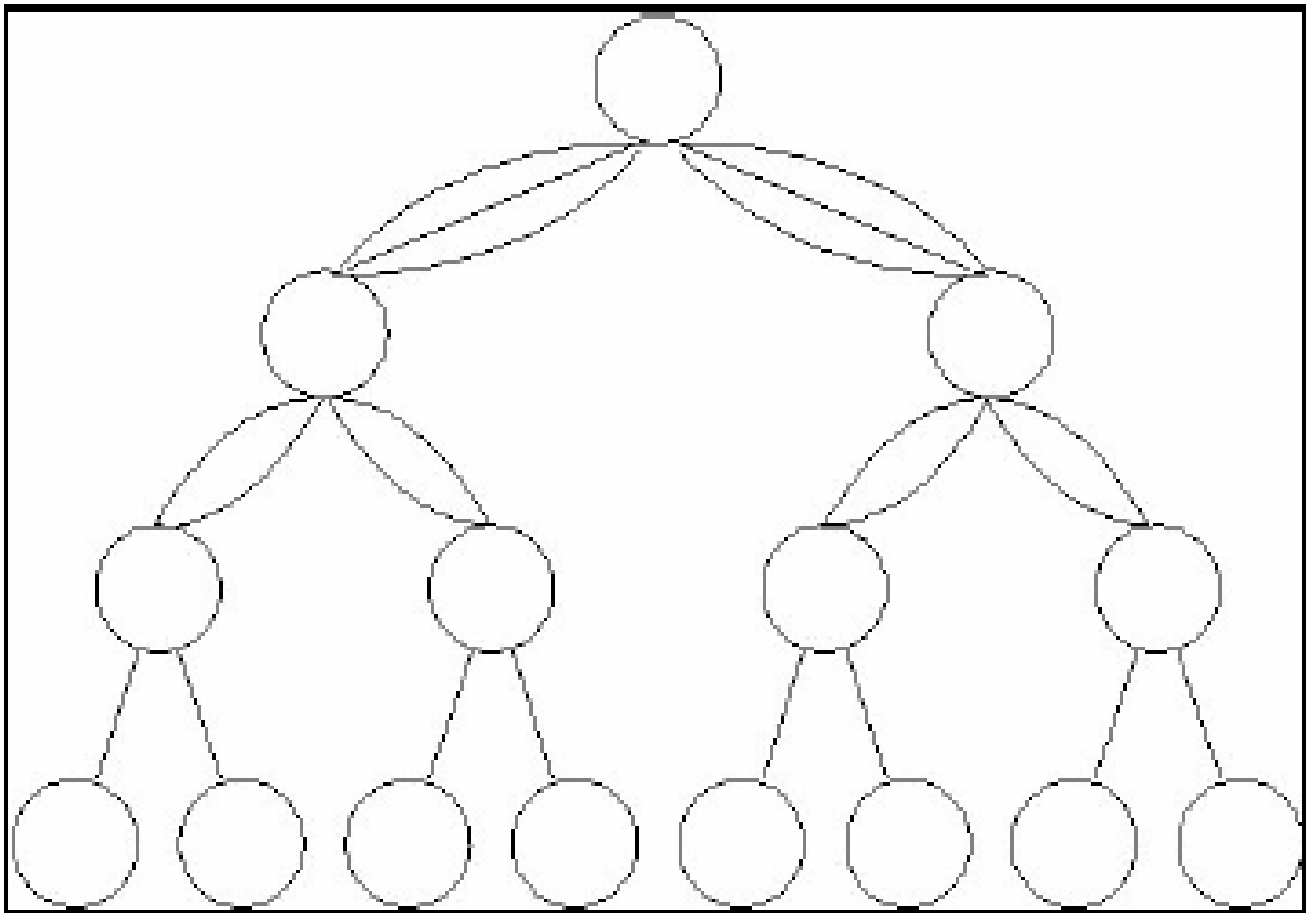
Шкафове, съдържащи Connected Unit



Свързване на клъстерите



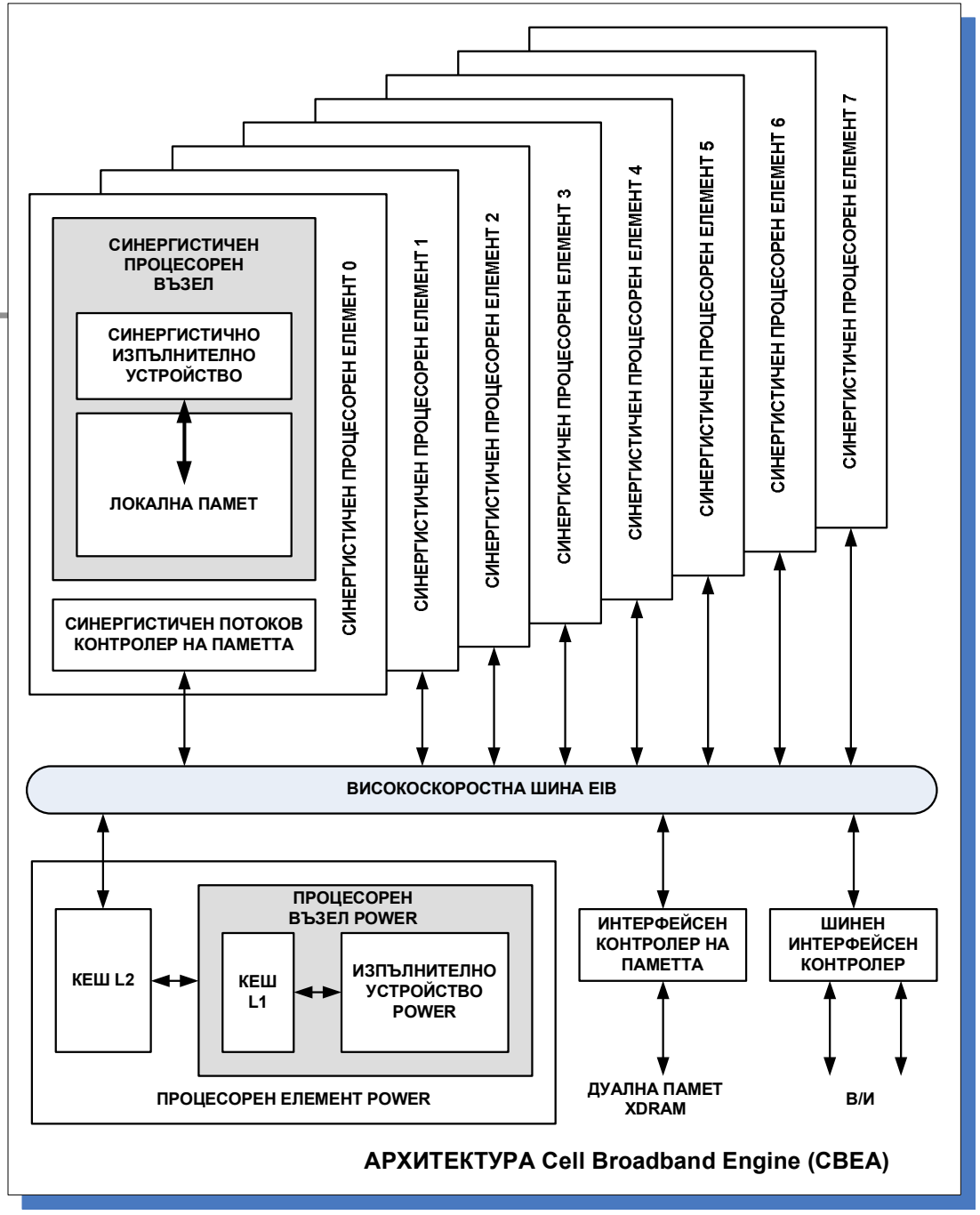
Мрежова топология Fat tree при Infiniband-базирани клъстери



Cell архитектура

- Едночипов мултироцесор с 9 процесорни елемента (PE) използващи модела с обща памет
- Процесорът Cell обхваща 1 Power Processor Element (PPE) и 8 Synergistic Processor Elements (SPE), Memory Flow Controller (MFC) и Internal Interrupt Controller (IIC)
- PPE – съдържа 64-битово ядро с PowerPC архитектура и може да изпълнява 32-битови и 64-битови приложения
- SPE – специално проектиран за интензивни изчисления - SIMD/векторни приложения, не могат да изпълняват задачите на ОС
  - SPE зависят от PPE за изпълнение на ОС и, в много случаи, управлението на нишките за потребителския код на високо ниво
- PPE зависи от SPEs за осигуряване на високите скорости на изчисляване на приложенията

Архитектура на Cell



Достъп до паметта

- Достъпът до паметта на PPE е като този на конвенционален процесор
- SPEs осъществяват достъп до главната памет с команди за директен достъп до паметта (DMA), които прехвърлят данни и инструкции между главната памет и private локалната памет (LS - local store)
  - SPE осъществяват по-скоро достъп до private локалната памет отколкото до главната обща памет
  - Тази иновационна организация на паметта на 3 нива (регистри, локална памет, и главна памет), с асинхронни DMA трансфери между локалните и главната памети, се различава радикално от конвенционалната архитектура и програмни модели
  - Тя дава възможност в явен вид да се осъществяват изчисленията паралелно с трансфера на данните и инструкциите, а също така и съхраняването на резултатите в главната памет

Иновационният модел на паметта в Roadrunner

- Главен мотив – през последните 25 години латентността на паметта (измервана в процесорни цикли) се е увеличила с 3 порядъка
- В резултат производителността на приложенията се ограничават по-скоро от латентността на паметта, отколкото от върховите скорости на изчисленията, определяни от бързодействието на процесорите (липси в кеша)
- Всеки от 8-те контролера за ДДП на SPE може да поддържа до 16 DMA трансфера едновременно и може да осигури навреме необходимите данни (just in time delivery)
- Осигурени са специални режими на ДДП за списъци (scatter gather списъци)

Суперкомпютър Ягуар JAGUAR 2009г.

- Cray Inc.
- Cray XT5-HE
- AMD x86_64 Opteron Six Core 2600 MHz (10.4 GFlops)
- Брой ядра 224 162

JAGUAR

- Jaguar е Cray XT4 система. Тя е предвидена като основна система LCF (Leadership Computing Facility);
- Jaguar има общо 7832 XT4 изчислителни възла. Всеки изчислителен възел съдържа Quad-Core 2,1GHz AMD Opteron процесор и 8GB памет. Общата производителност на системата е около 263 TF;
- Изчислителните възли работят на Compute Node Linux (CNL) OS. CNL е проектирана да намали системните разходи, като по този начин позволява мащабируемост с ниска латентност и глобална комуникация;
- Всеки възел е свързан с Cray "SeaStar" рутер чрез HyperTransport и топологията на свързване е 3-D Тороид. Резултатът е много висока честотна лента, ниска латентност и екстремна скалируемост.

JAGUAR SUPERCOMPUTER VIDEO

http://www.youtube.com/watch?v=zx677CceBAc

http://www.youtube.com/watch?v=3ClWeTIqzVc&feature=related

ТОП500 СУПЕРКОМПЮТРИ WWW.TOP500.ORG

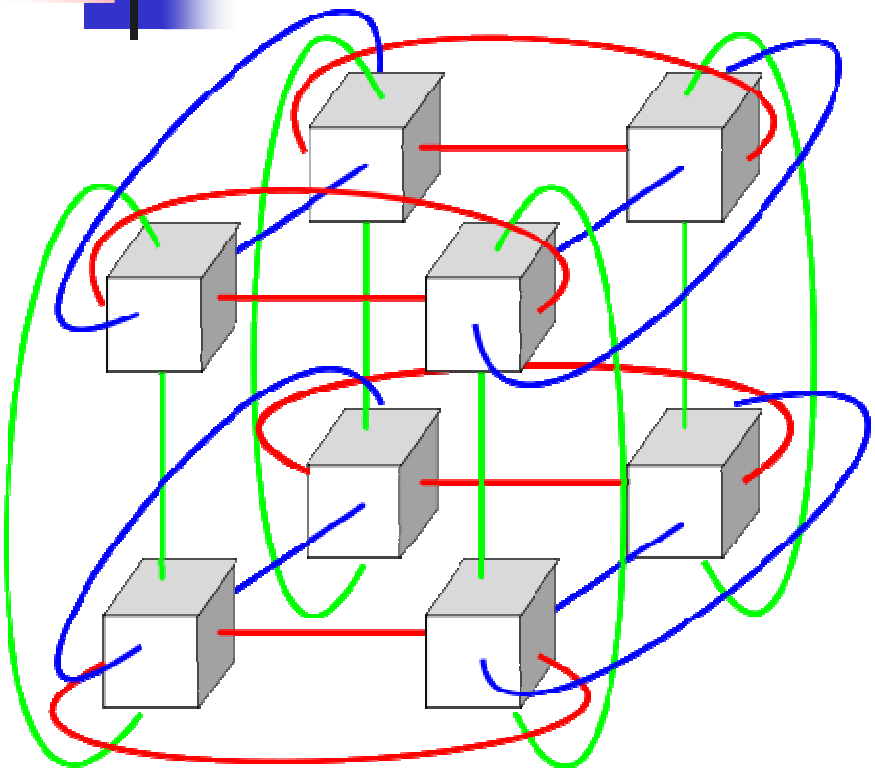
- Ноември 2011 г.
- “K Computer” – Япония, Fujitsu
- Инсталиран в RIKEN Advanced Institute for Computational Science (AICS) в Kobe, Япония
- 10.51 Petaflop/s за Linpack benchmark
- 705 024 ядра SPARC64
- Японската дума "kei" (京?), която означава 10 quadrillion
- През ноември 2011, TOP500 ранкира K като най-бързия суперкомпютър с производителност над 8 petaflops

No 1 - “K Computer”

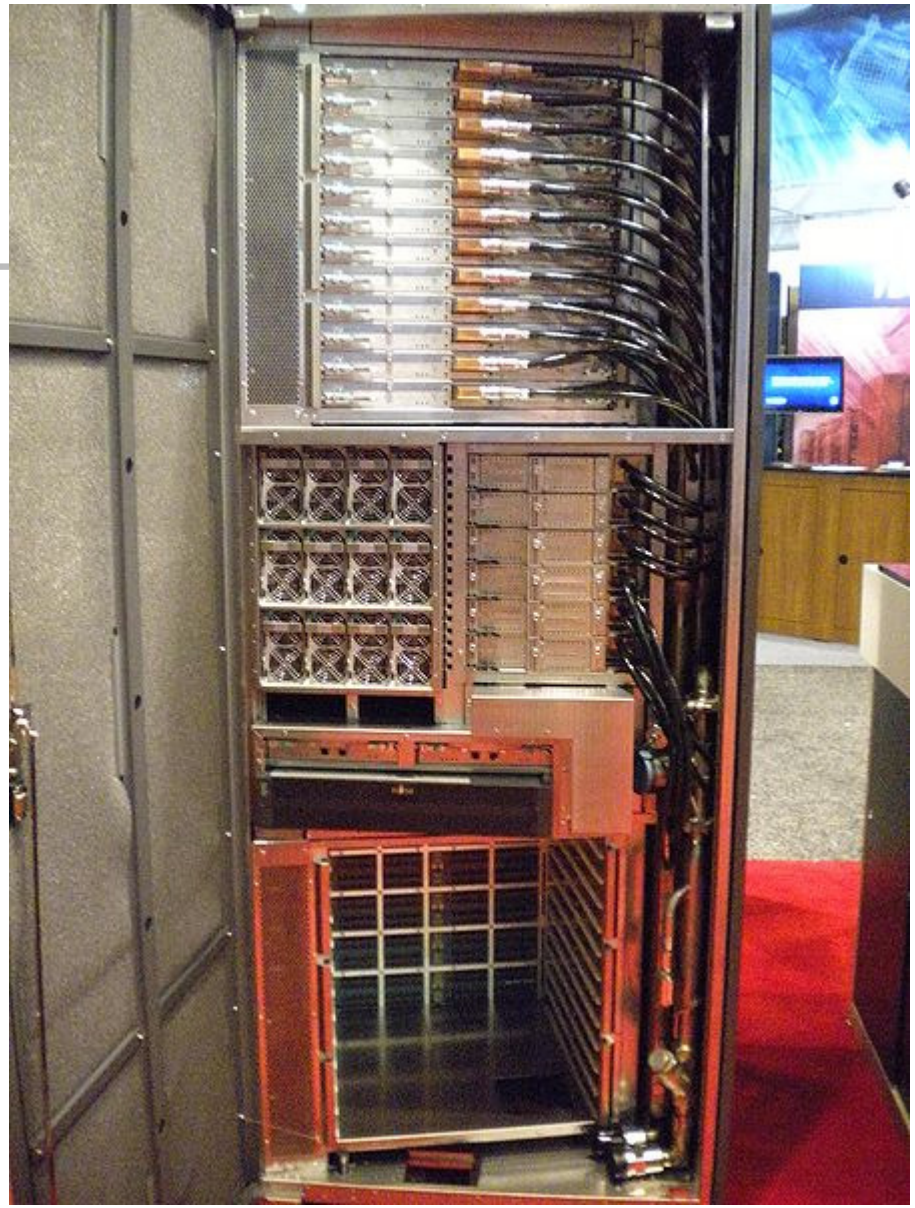
Япония, Fujitsu

Системна мрежа Tofu
Памет - 1410048 Gb
Консумирана мощност - 12659.89 Kw

“K Computer”



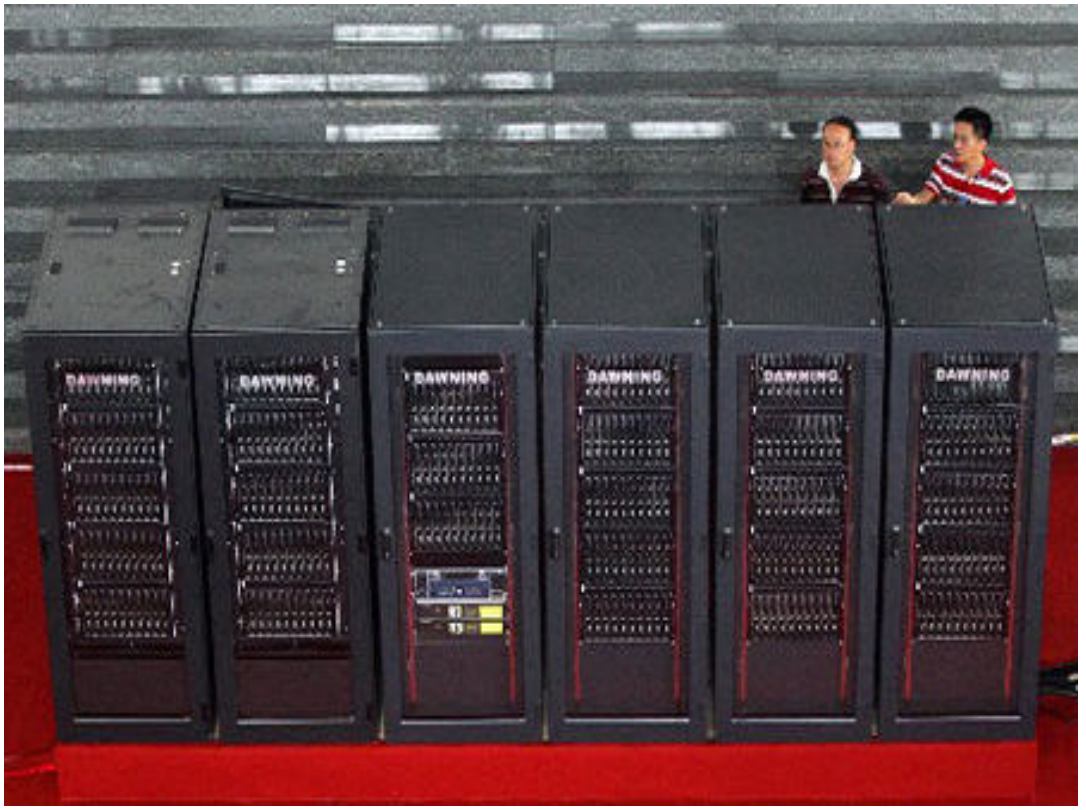
Системната комуникационна мрежа е 6-мерен тороид *Tofu*



No 2 Tianhe - Китай

- Национален суперкомпютърен център в Тянджин
- National University of Defense Technology - NUDT
- 186 368 ядра
- 2.566 petaFLOPS
- Изчислителен възел - 2 Xeon X5670 6-ядрени процесора и един процесор Nvidia M2050 GPU
- Системната мрежа е базирана на Infiniband

Суперкомпютър NEBULAE - "Мъглявина" Китай



120 640 ядра

КРАЙ