

CAP-387(2016) – Tópicos Especiais em Computação Aplicada: Construção de Aplicações Massivamente Paralelas

Aula 2: Sistemas Massivamente Paralelos Atuais

Celso L. Mendes, Stephan Stephany

LAC /INPE

Emails: celso.mendes@inpe.br, stephan.stephany@inpe.br



Lista Top500

- **Finalidade**

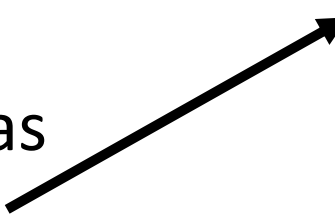
- Listar os 500 supercomputadores mais rápidos no mundo
- Foco na capacidade numérica – execução do Linpack

- **Periodicidade**

- 2 edições por ano: Junho (ISC-Europa) e Novembro (SC-EUA)
- Iniciada em 1993

- **Processo de Participação**

- Executar o Linpack sem mudanças
- Reportar valor de R_{max} em flops
 - Reportar também R_{peak} : máximo teórico



Sigla	Unidade	Flops/s
MF	Megaflops	10^6
GF	Gigaflops	10^9
TF	Teraflops	10^{12}
PF	Petaflops	10^{15}
EF	Exaflops	10^{18}

Lista Top500 Atual

- Edição mais recente: Junho/2016 (www.top500.org)
 - Sistema #1: Processador chinês, many-core, RISC
 - Sistema #2: Intel Xeon + Aceleradores Intel Xeon-Phi
 - Sistema #3: Cray XK7 (GPUs), 200 racks, Gemini interconnect
 - Sistema #4: BlueGene/Q (descontinuado)
 - Sistema #5: Processador Sparc

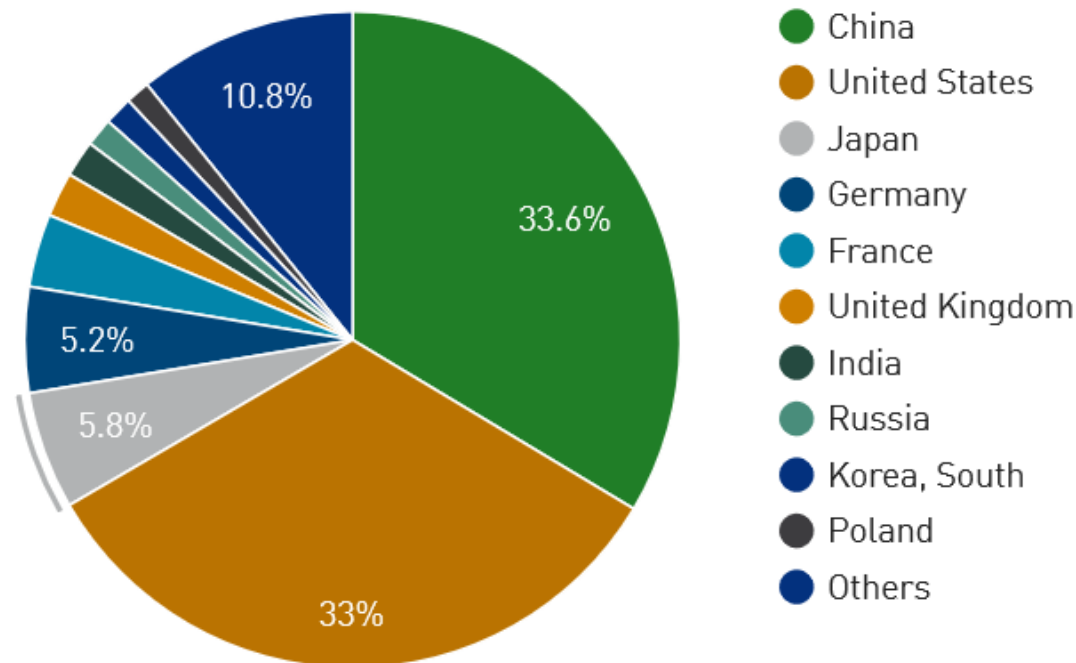
Posição	Sistema	País	R _{max} (PFlops)	R _{peak} (PFlops)
1	Sunway TaihuLight (NRCPC)	China	93,01	125,44
2	Tianhe-2 (NUDT)	China	33,86	54,90
3	Titan (Cray)	EUA	17,59	27,11
4	Sequoia (IBM)	EUA	17,17	20,13
5	K Computer (Fujitsu)	Japão	10,51	11,28



Lista Top500 - Geografia

- Participação por países: Junho/2016 (número de sistemas)
 - China > EUA pela primeira vez!

Countries System Share

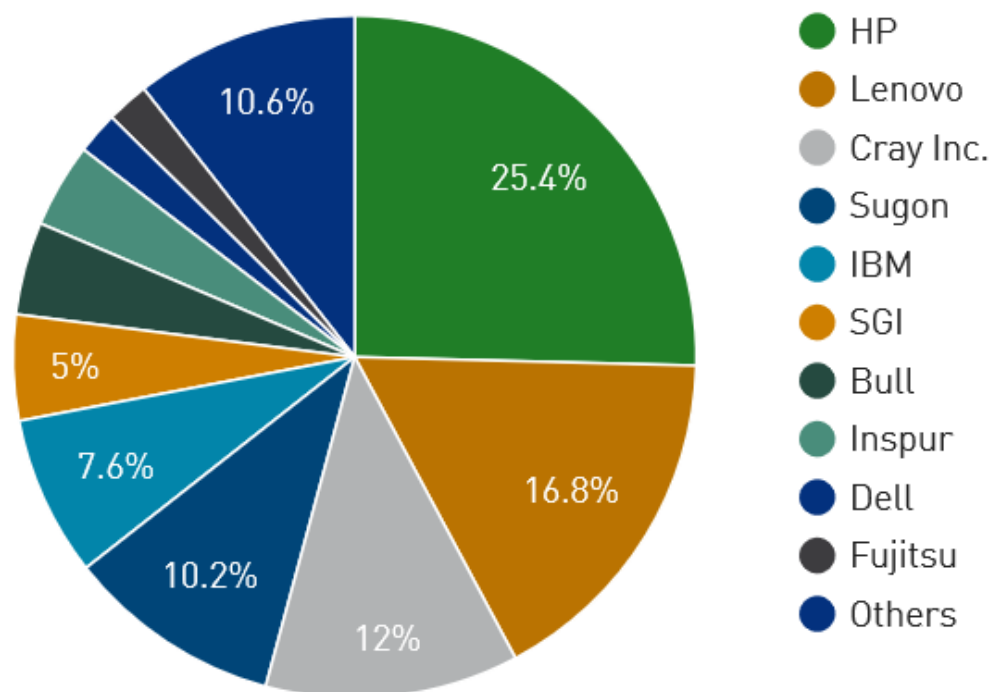


Lista Top500 - Fabricantes

- Participação por fabricantes: Junho/2016 (núm. de sistemas)

- Note: HP + SGI > 30%

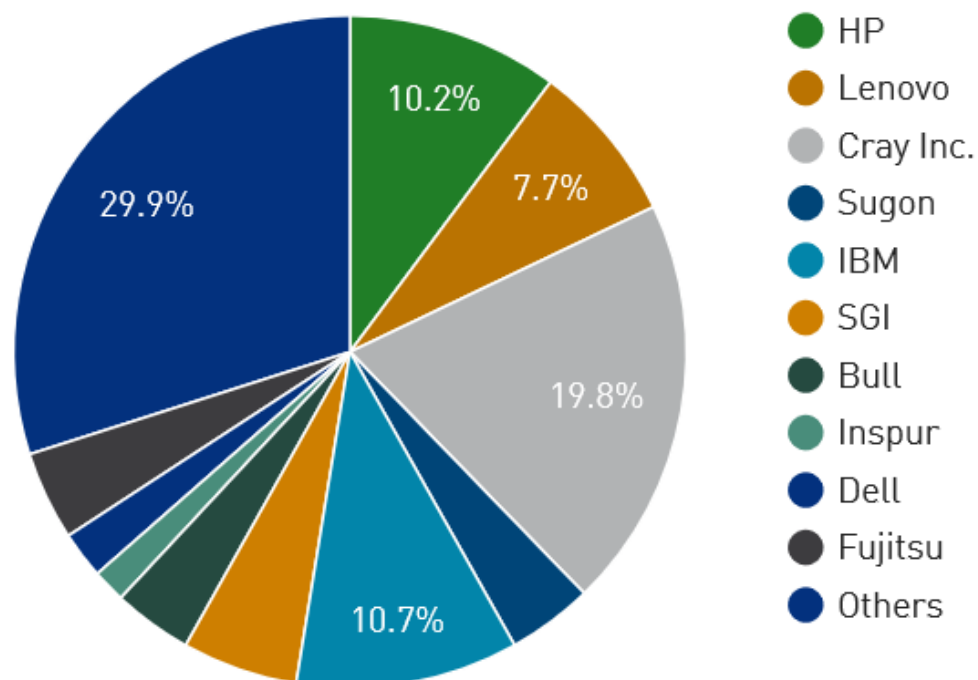
Vendors System Share



Lista Top500 - Fabricantes

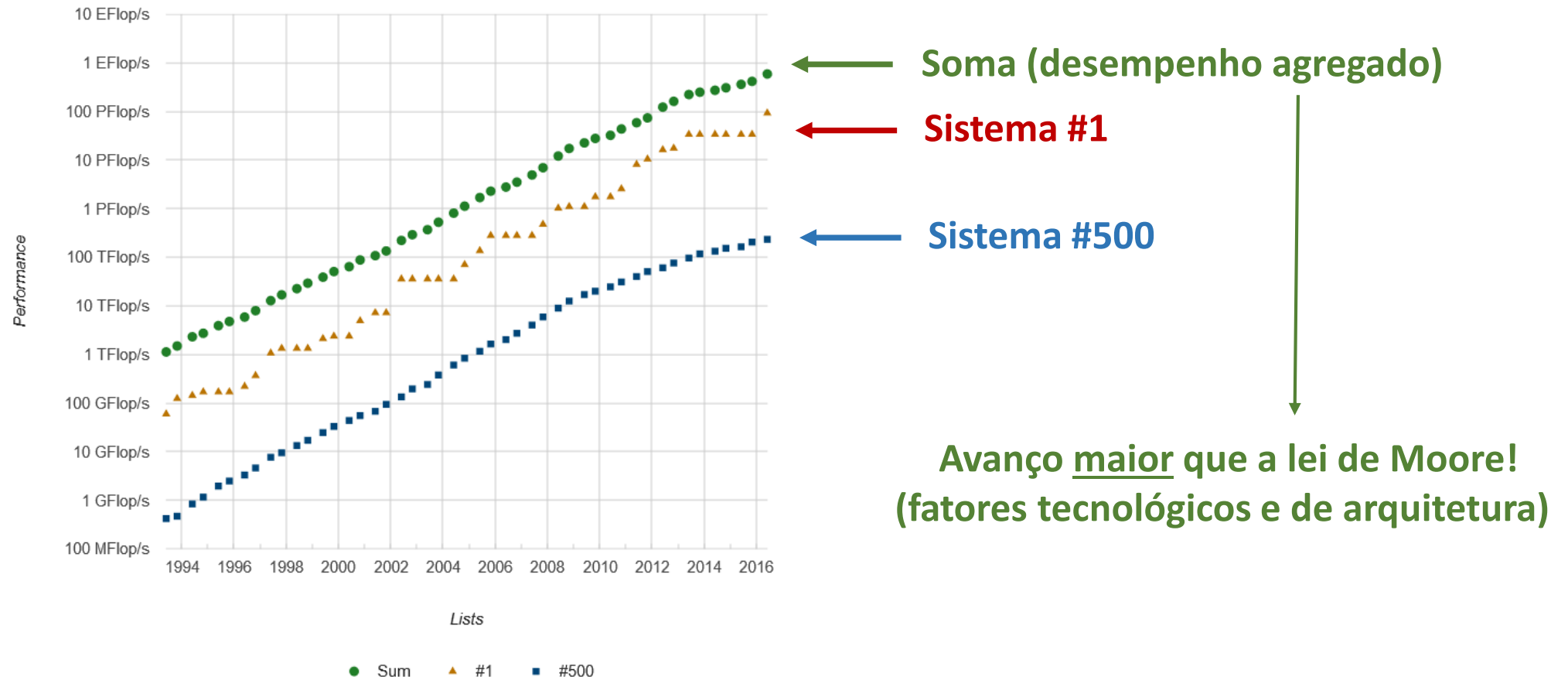
- Participação por fabricantes: Junho/2016 (desempenho total)
- OBS: fabricantes na mesma ordem do slide anterior

Vendors Performance Share

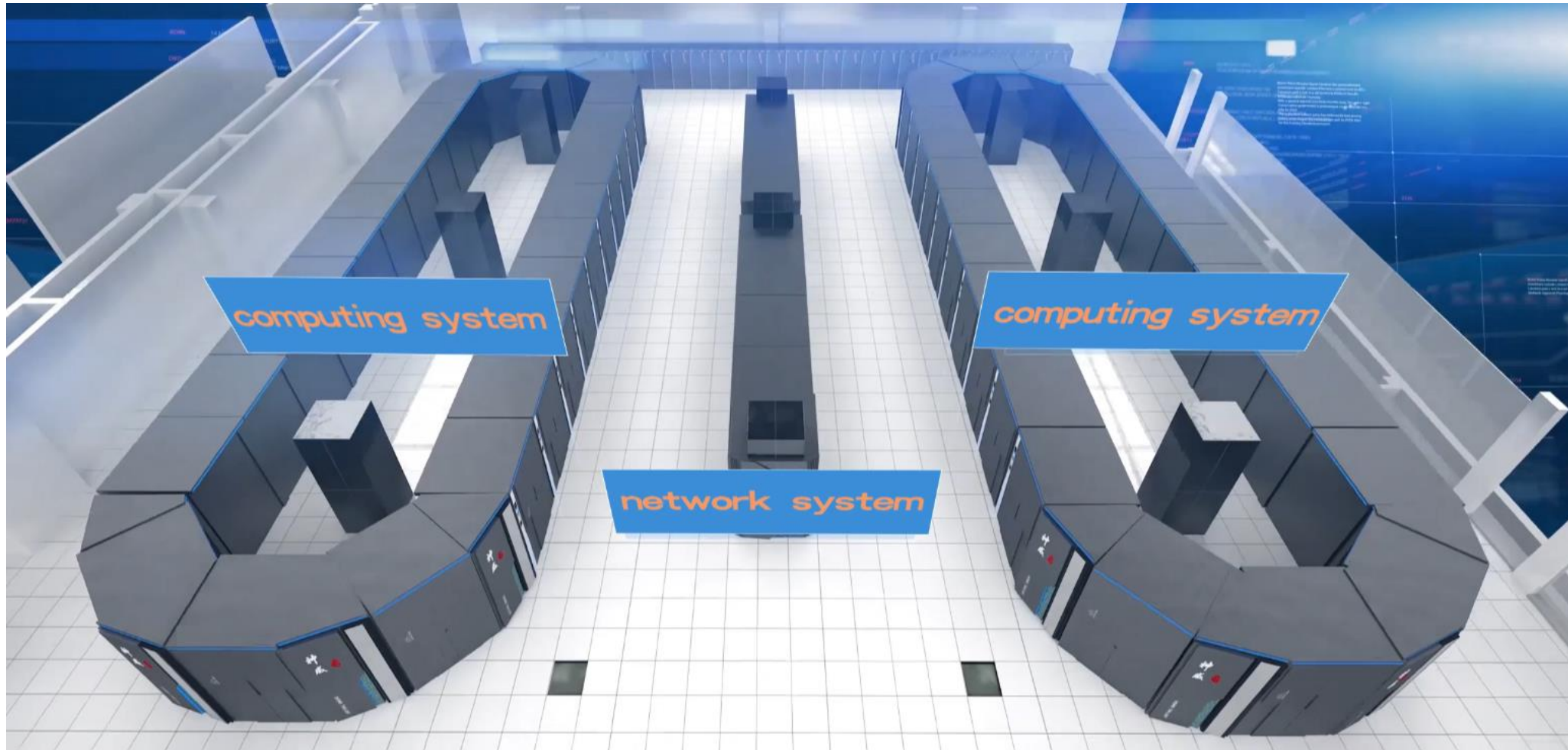


Lista Top500 - Evolução

- Evolução ao longo dos anos



Sistema #1: Sunway TaihuLight



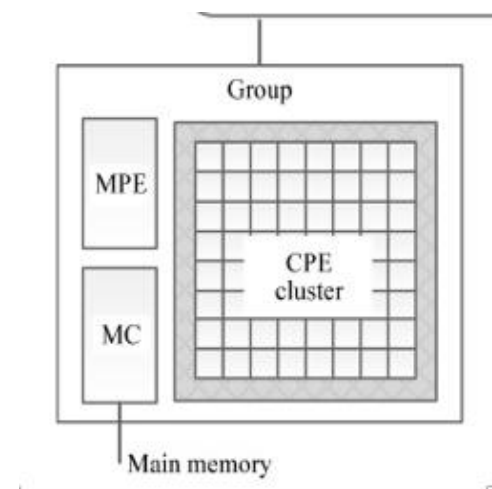
Sistema #1: Sunway TaihuLight

- Sunway TaihuLight – China, 2016
 - Instalado no centro de supercomputação em Wuxi
 - URL: <http://www.netlib.org/utk/people/JackDongarra/PAPERS/sunway-report-2016.pdf>
 - Total de 10.649.600 núcleos, em 40.960 nós (1 chip por nó)
 - Desempenho total de pico: 125 Pflops/s, em 40 racks
 - Desempenho no Linpack: 93 Pflops/s (~ 74.4% do pico)
 - Linpack: código numérico intensivo
 - Desempenho no HPCG: 0.3% do pico!
 - HPCG: código numérico + memória + comunicação
 - Conclusão: Sistema extremamente desbalanceado

Sistema #1: Sunway TaihuLight

- **Processador: SW26010 (Shanghai, China)**
 - Arquitetura many-core, RISC
 - Desempenho de pico de 11.6 Gflops/s em cada núcleo
 - 4 grupos de núcleos; cada grupo: 64 núcleos (CPE) mais 1 de controle (MPE)
 - 260 núcleos por chip: > 3 TFlops/s por chip
 - 32 GB de mem. externa por nó, 1.2 PB total

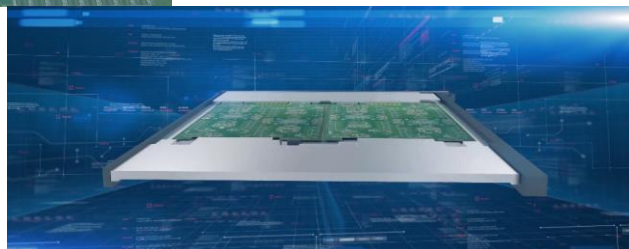
Grupo de núcleos:



Sunway TaihuLight - Hierarquia



2 chips por cartão (2 nós)



4 cartões por placa: 2+2 (8 nós)

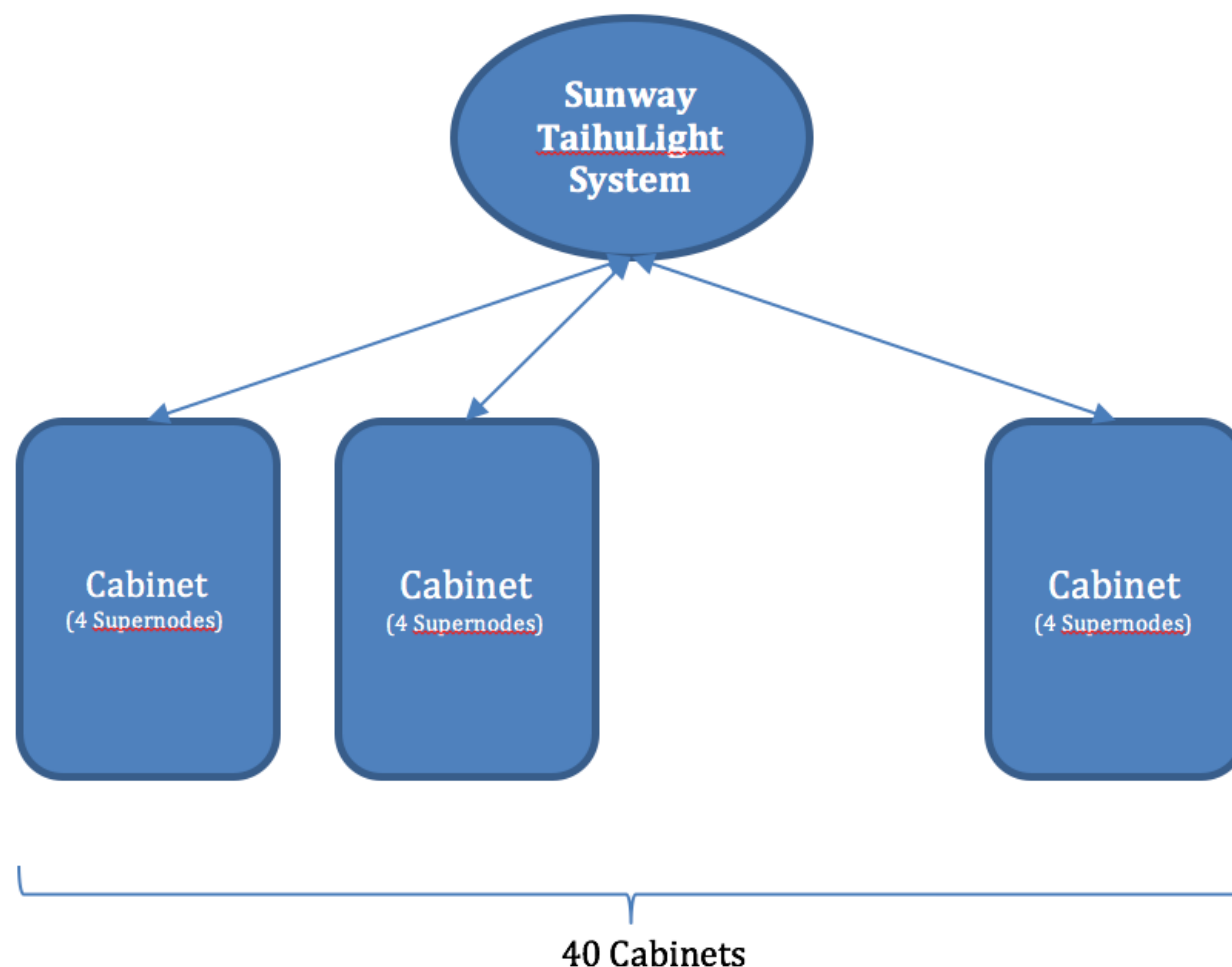


32 placas por Supernó (256 nós)

4 Supernós por rack (1024 nós)



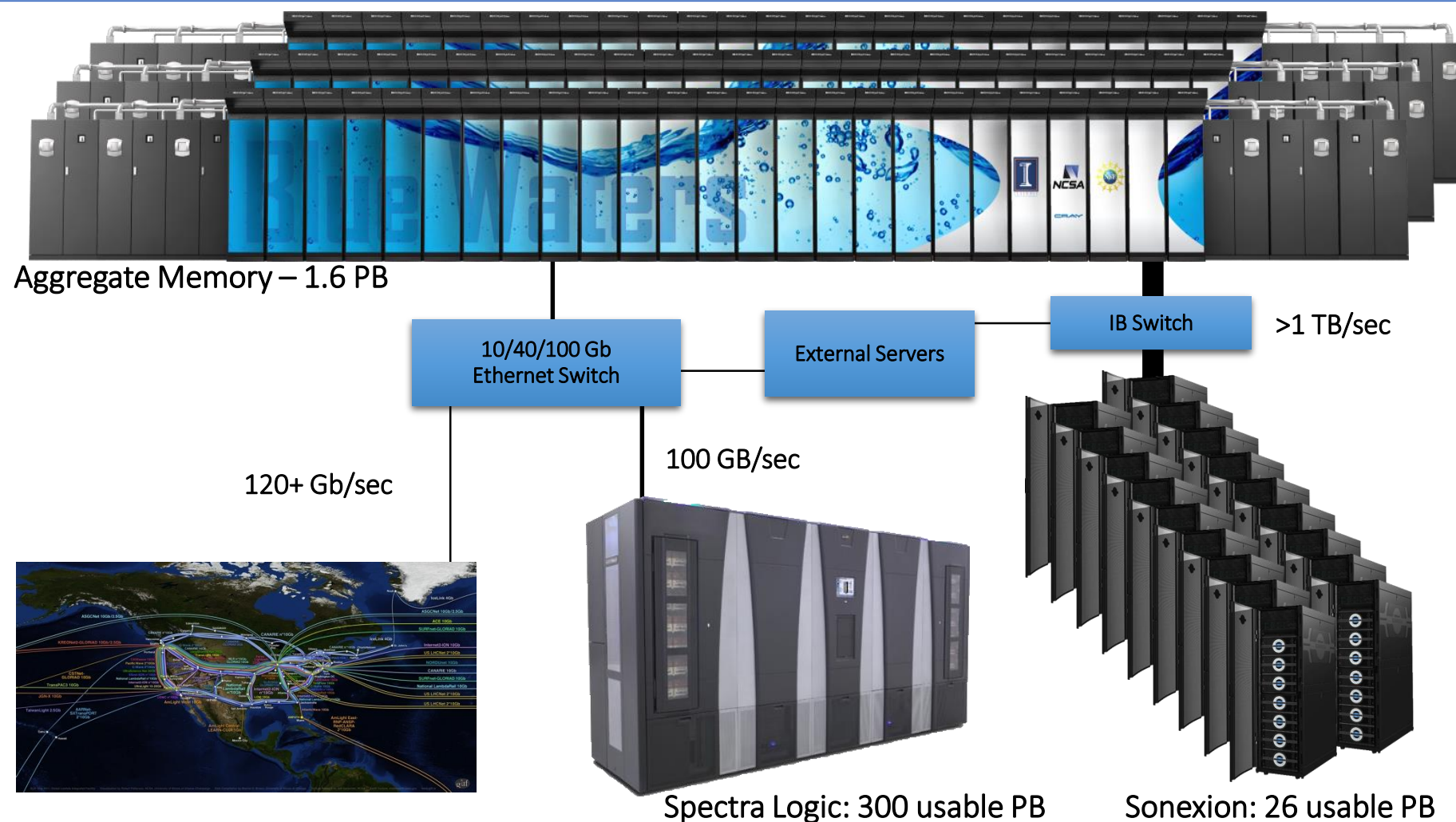
Sunway TaihuLight - Sistema



Além do Top500

- **Sistema Blue Waters – Univ. Illinois, EUA**
 - Financiado pela NSF (equivalente ao CNPq)
 - Cray XE6/XK7, 13.3 PF de pico
 - Arquitetura similar ao Titan/XK7 (#3 na lista Top500)
 - Desempenho de pico maior que o #5 da lista atual
 - Várias razões para não entrar no Top500:
 - Sistema híbrido: CPU (XE6), CPU+GPU (XK7)
 - Sistema balanceado:
 - Capacidades: computacional, armazenamento, I/O
 - Capaz de atender a vários tipos de aplicações
 - Avaliação: métrica SPP (*Sustained Petascale Performance*)
 - Mede o desempenho sustentado, em códigos reais
 - Mais que 1 PFlops/s medido em várias aplicações

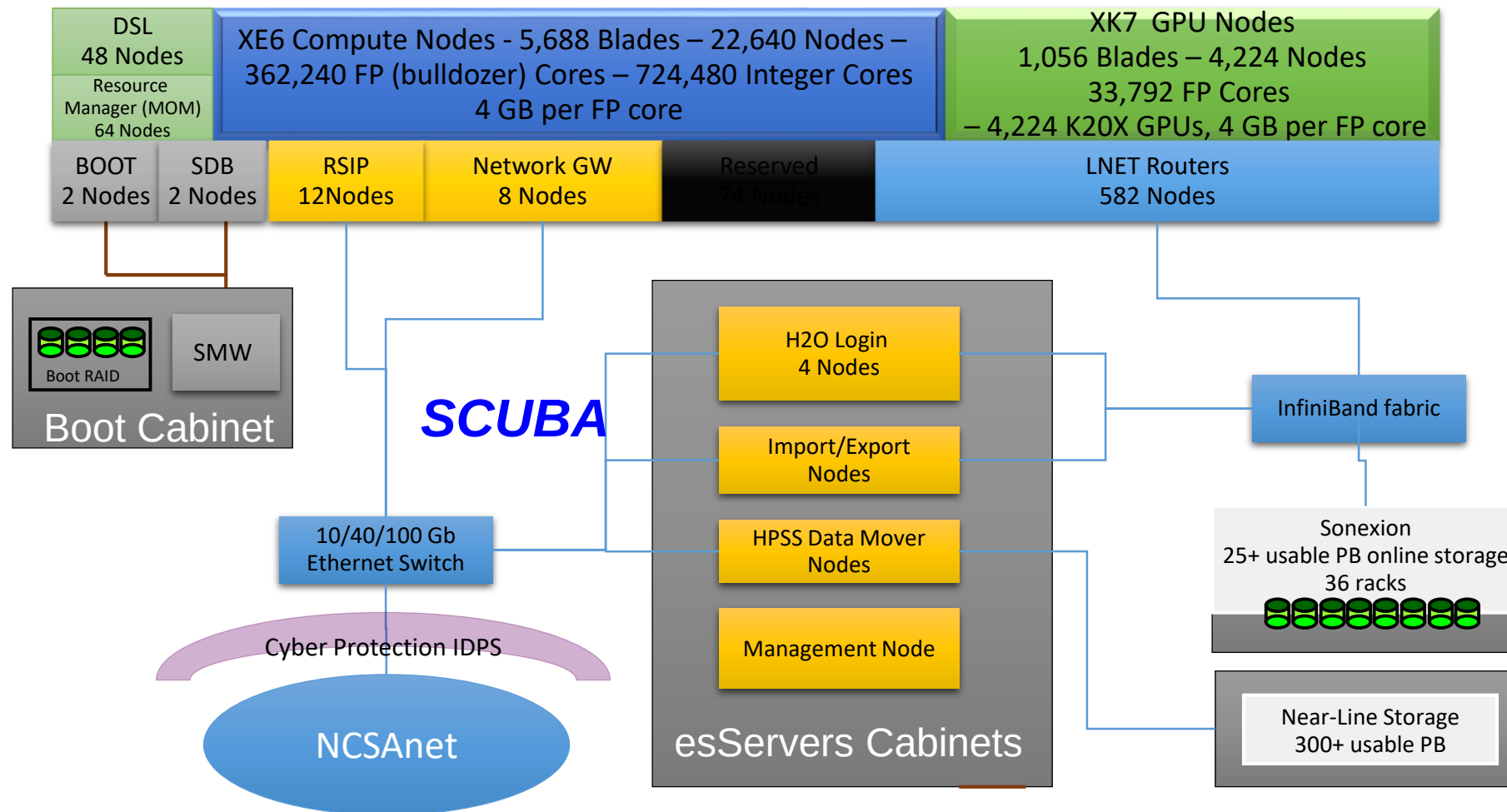
Sistema Blue Waters



Arquitetura do Blue Waters

Rede Gemini (HSN)

Cray XE6/XK7 - 288 Gabinetes

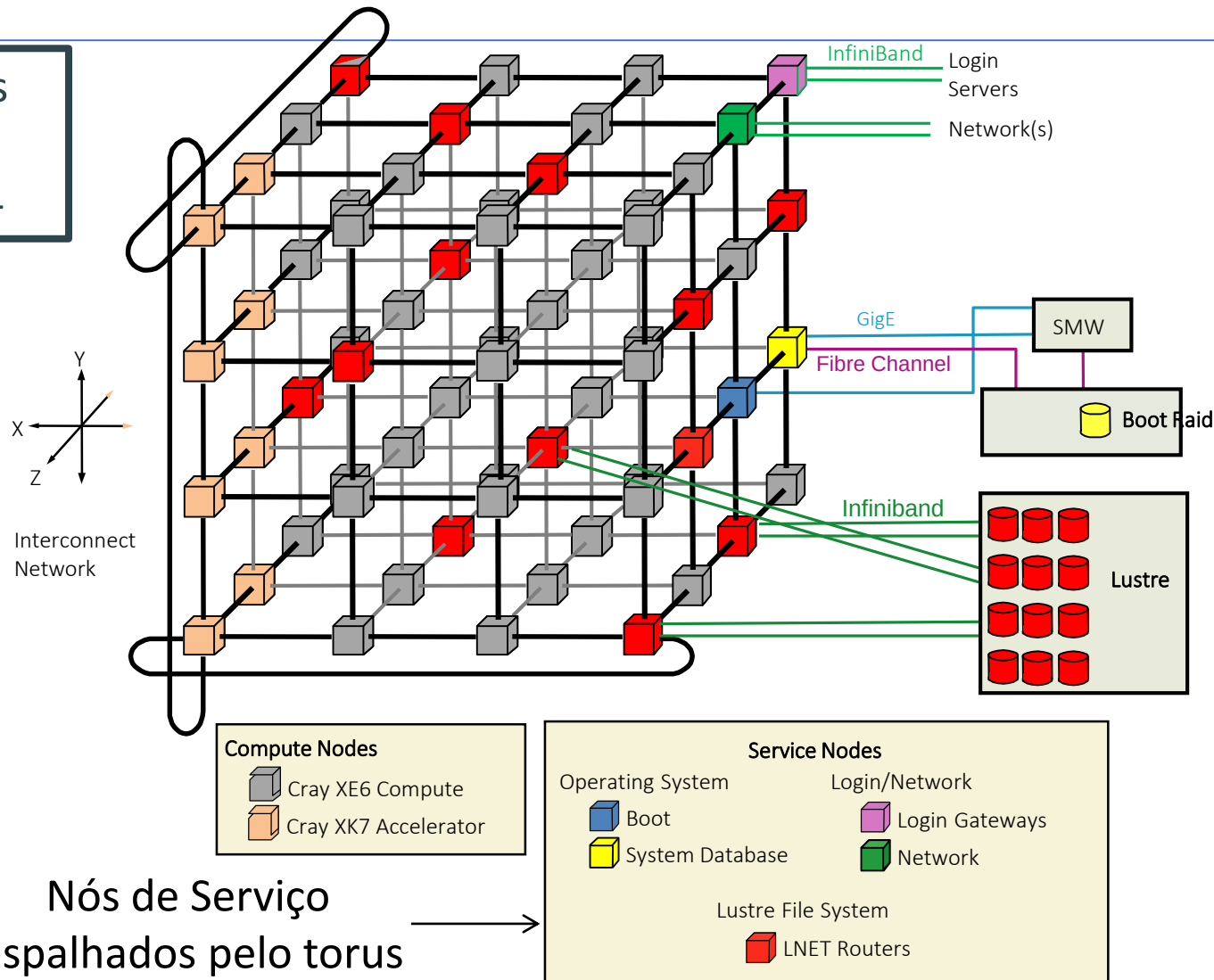


Números do Blue Waters

- Total de Gabinetes XE/XK: 288
 - Gabinetes XE: 243
 - Gabinetes XK: 45
- Desempenho de Pico: **13.3 PF**
 - CPUs x86: 7.1 PF
 - GPUs: 6.2 PF
- Memória: 1.6 PB
- Disco: 26 PB
- Arquivamento em Fita: ~ 300 PB

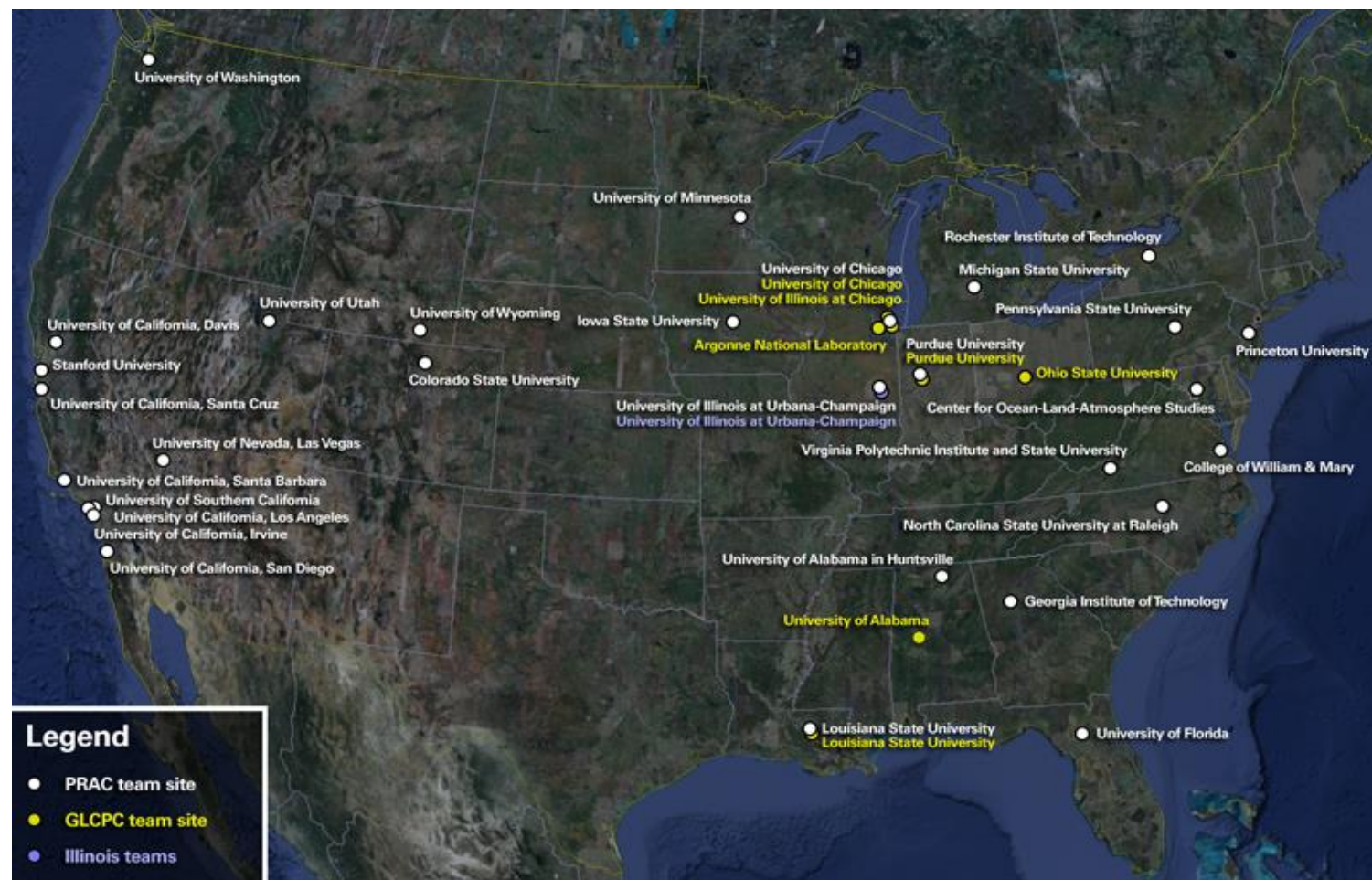
Interconexão no Blue Waters

Blue Waters
Torus 3D
24 x 24 x 24



Nós de Serviço
espalhados pelo torus

Usuários do Blue Waters



Aplicações no Blue Waters

Códigos PRAC

(80% do tempo de máquina)

Sumário

Total: 35 amostras

Fortran: 25 (71%)

C: 16 (46%)

C++: 20 (56%)

MPI: 31 (89%)

OpenMP: 17 (49%)

GPU: 10 (29%)

Disciplina	Código	Linguagem	Modelo de Progr.
Lattice QCD	MILC, Chroma, CPS	C/C++	MPI/pthreads, GPU
Biomolecular dynamics	NAMD	C++	Charm++, GPU
Astrophysics	ENZO/Cello	C++/Fortran	MPI+OpenMP
Astrophysics	Pgadget/ENZO	C++/Fortran	MPI
Chemistry	ACCESSIII	C, C++, Fortran	MPI
Social networks	Episimdemics	C++	Charm++
Astrophysics	Cactus, LazEv, Harm3D, Whisky	C, C++, Fortran	Cactus, MPI, OpenMP
Fluid dynamics	PSDNS	Fortran	MPI
Climate/weather	GCRM, CCSM-IE, SP-CCSM (POP, CICE-4, CAM3.5)	C and F77/F90	MPI/OpenMP, MCT
Geophysics	AWP-ODC, Hercules	F77, F90	MPI, OpenMP, GPU
Material science	QMCPACK/ AFQMC	C++/Fortran	MPI/OpenMP, GPU
Astrophysics	Cactus, CCATIE, McLachlan, Whisky, DiFranco	C, C++, Fortran	Cactus, MPI, OpenMP
Chemistry	GAMESS/ NWChem	C and Fortran	DDI/ GA/ARMC, OpenACC
Astrophysics	PPM	Fortran	MPI/OpenMP, GPU
Biology	ecology simulator	C++	MPI
Climate/weather	CM1	Fortran/F90	MPI/OpenMP
Geophysics	SPECFEM3D / LSQR	F90 / C++	MPI
Turbulence in Fluids	DISTUF	F95	MPI PETSc
Space Physics	MS-FLUKSS with Chombo	Fortran/C++ python	MPI+OpenMP
Astrophysics	ChaNGa	C++	Charm++
Space-based Earth Science	GRIPS SHARK Revisit	C/F90	MPI
Materials Science	OMEN NEMO5	C++	MPI PETSc
Software		C/C++/F90	MPI
Biophysics	Gromacs	C	MPI+OpenMP, GPU
Climate/weather	CCSM	C/Fortran	MPI+OpenMP
Space Physics	Magtail OSIRIS	F90	MPI
Biophysics	AMBER	F90/C	MPI, GPU
Astrophysics	Maestro Castro	F90 C++/F90	MPI+OpenMP
Biophysics	LAMMPS	C++/F90	MPI+OpenMP
Heliophysics	H3D/VPIC	F90/C	MPI+OpenMP
Heliophysics	BIFROST STAGGER RADMHD		
Materials Science	PHOTONPLAMSA	c/c++	MPI, OpenMP, GPU
Materials Science	RMG, LAMMPS	C/C++	MPI, GPU
Astrophysics	PGADGET	C/C++	MPI+threads
Weather	WRF	F90/C	MPI+OpenMP
Earth Sciences	SPECFEM3D	F90	MPI

