

**GT – NONO
PERSPECTIVA DE USO DO
SUPERCOMPUTADOR SANTOS DUMONT**

17º WIRNP

Workshop RNP

**30 | 31 MAIO
SALVADOR | BA**

AUGUSTO C. GADELHA

LABORATÓRIO NACIONAL DE COMPUTAÇÃO CIENTÍFICA
LNCC / MCTI



USO DO SUPERCOMPUTADOR SANTOS DUMONT

17º **WRNP**
Workshop RNP

Laboratório Nacional de Computação Científica LNCC/MCTI Petrópolis, RJ



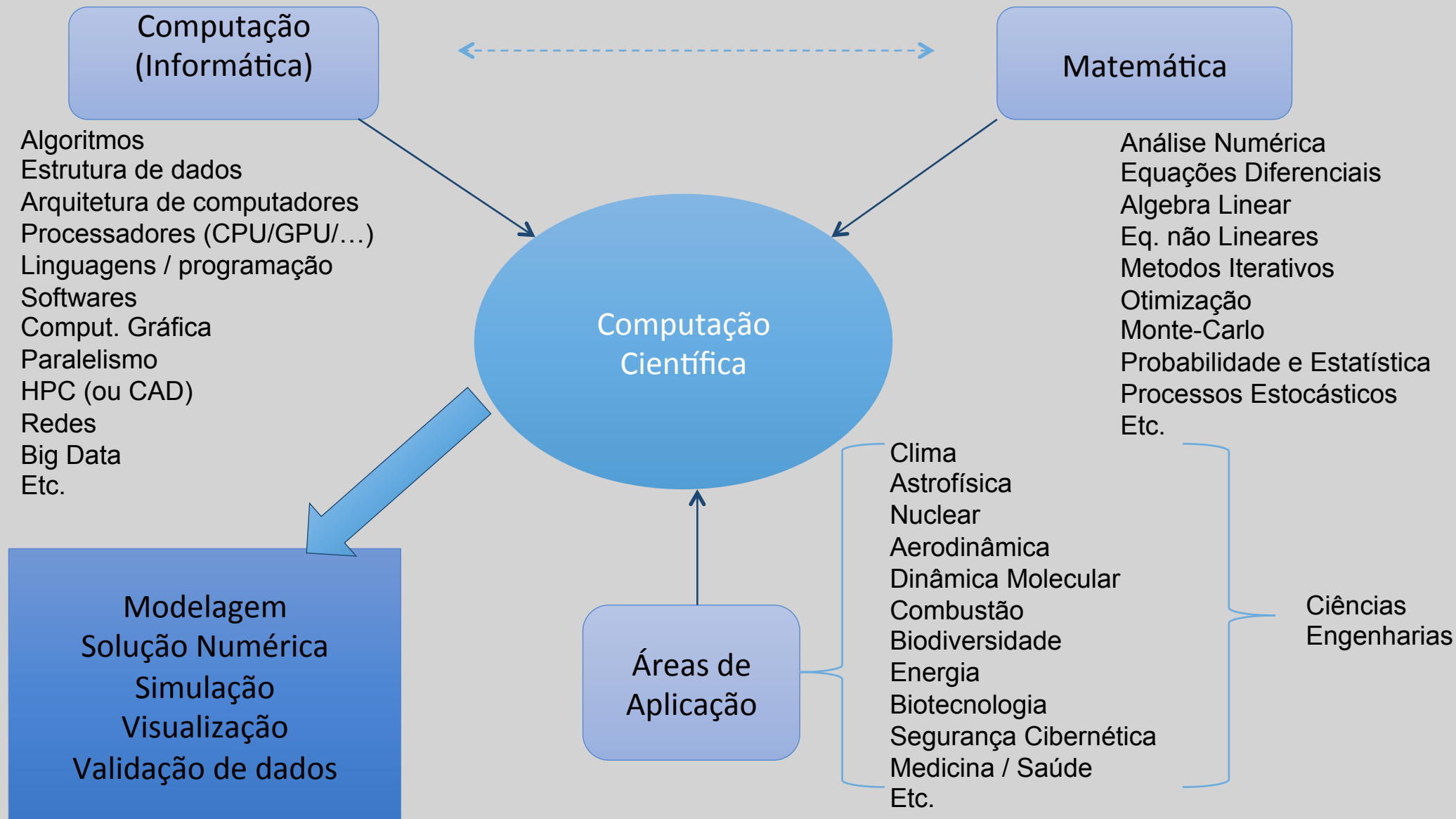
POR QUE O BRASIL NECESSITA DE COMPUTADORES DE
ALTO DESEMPENHO NA FAIXA DE CENTENAS DE
TERAFLOPS OU PETAFLÓPICOS?

Paradigmas da Ciência

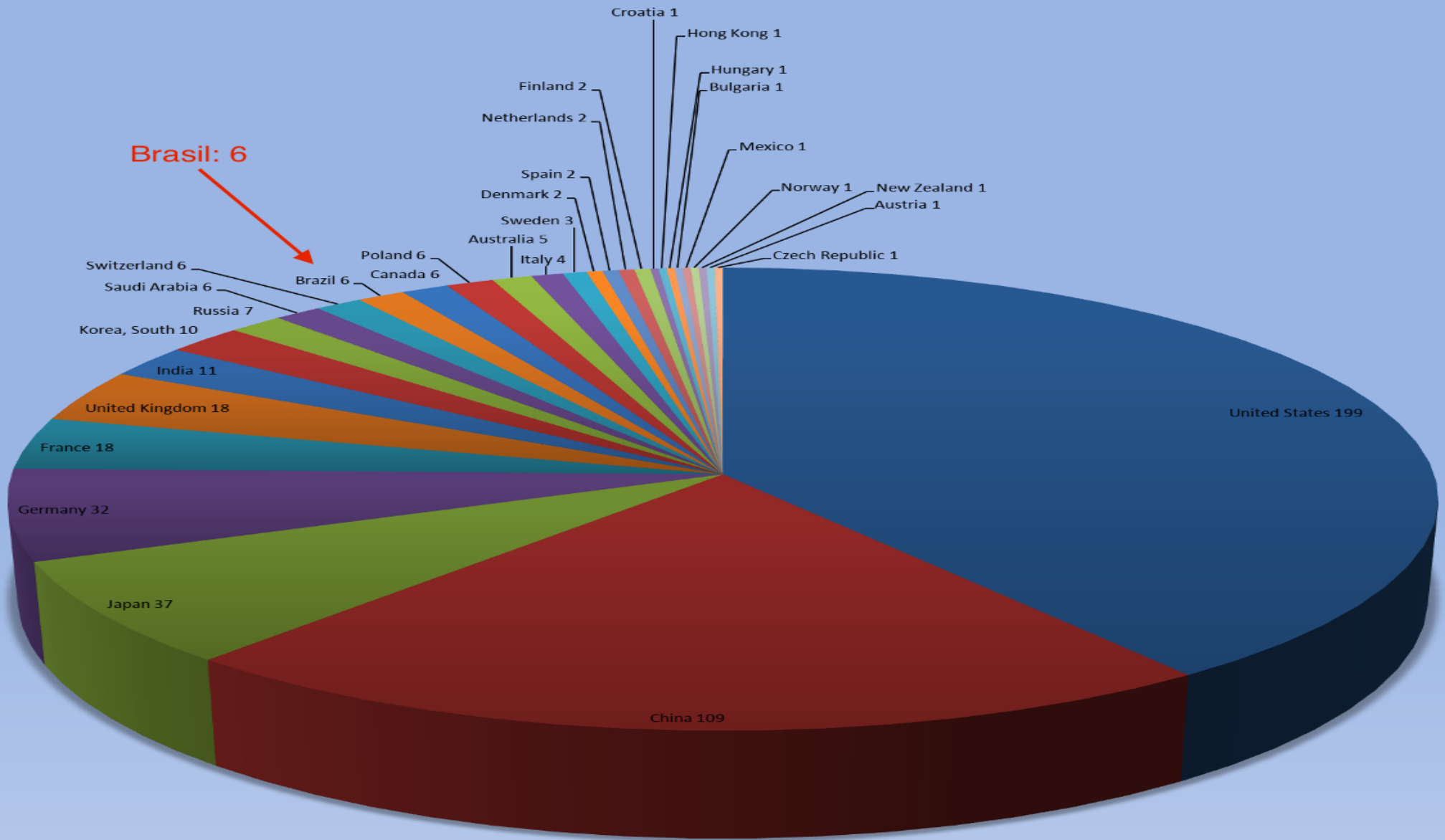
- ① Observação e experiência (Galileu, ...)
- ② Teoria (Newton, Leibnitz, Einstein ...)
- ③ Computação → Programação e simulação
- ④ Computação intensiva em dados → e-Ciência
→ Big Data

Meta dos países mais avançados é alcançar a velocidade EXAFLOPICA em 5 anos (Exascale HPC)

→ tratamento de problemas científicos de mais alta complexidade

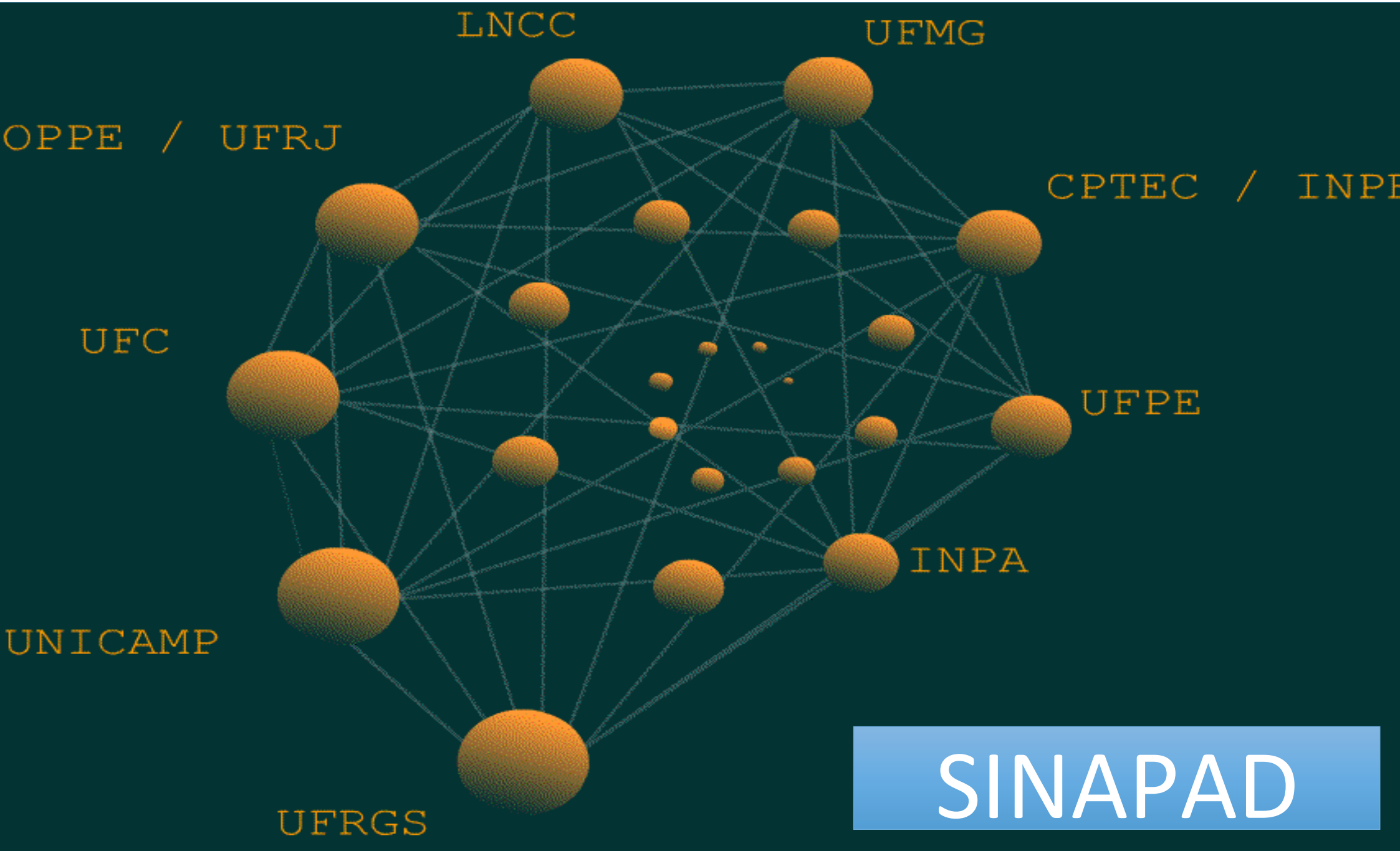


- BRASIL ESTÁ SE PREPARANDO PARA ENTRAR NA ERA DA “EXASCALE HPC”?
- ESTAMOS FORMANDO CIENTISTAS, ENGENHEIROS E ESPECIALISTAS EM HPC E COMPUTAÇÃO CIENTÍFICA PARA ACOMPANHAR AS PESQUISAS CIENTIFICAS MAIS AVANÇADAS?



Share of Supercomputers by Countries
Source:www.top500.org

NOV 2015



SINAPAD

O Supercomputador Santos Dumont



USO DO SUPERCOMPUTADOR SANTOS DUMONT

17º **WRNP**
Workshop RNP



PROJETO SDUMONT – SISTEMA DE COMPUTAÇÃO PETAFLÓPICA

INVESTIMENTO:

Equipamento computacional: R\$ 43 M

Infraestrutura (refrigeração, energia,...) R\$ 12 M

CUSTO 1º ANO DE OPERAÇÃO DA BULL: R\$ 1,9 M

CUSTO ANUAL DE APOIO AOS USUÁRIOS: R\$ 2,5 M

CUSTO ANUAL DE ENERGIA: R\$ 5,0 M

CONTRAPARTIDA BULL: € 7,5 M em 3 anos

Centro de Pesquisa da BULL em Petrópolis – LNCC

Centro de Aplicações da BULL no Rio de Janeiro – UFRJ

SUPERCOMPUTADOR SDUMONT

TECNOLOGIA DE PONTA EM PROCESSADORES

- Capacidade de processamento na ordem de 1,1 Petaflop/s (1,1 x 10¹⁵ float-point operations per second).
- Configuração híbrida de nós computacionais (CPU, GPU, MIC).

Configuração de Processamento SDUMONT

18.144 núcleos de CPU, distribuídos em 756 nós computacionais

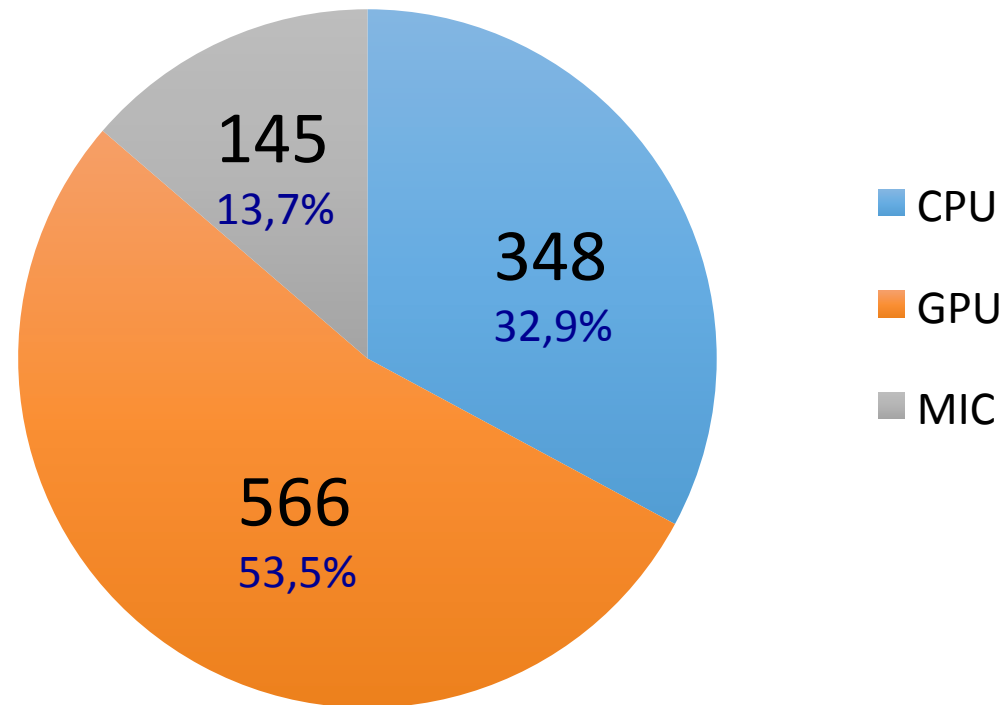
756 nós de computação de modelos B710/B715:

- 504 nós de computação modelo B710 ← CPU (66,7%)
- 198 nós de computação modelo B715 (com 2 GPU cada) ← CPU + GPU (26,2%)
- 54 nós de computação modelo B715 (com 2 Intel XeonPHI cada) ← CPU + MIC (7,1%)

OBS: Cada nó de computação dos modelos B710 e B715 possui: 2 CPU INTEL XEON E5-2695v2 Ivy Bridge (12cores, 2.4Ghz), 64 GIGABYTES RAM

1 nó de computação de modelo FAT NODE : com 16 CPU INTEL IVY BRIDGE (15cores, 2.3GHZ), 6 TERABYTES RAM

CAPACIDADE COMPUTACIONAL INSTALADA (Tflops)



SDUMONT: TECNOLOGIAS DE PONTA EM HPC

◆ SISTEMAS DE REFRIGERAÇÃO A LIQUIDO

Direct Liquid Cooling → PUE < 1,1

◆ REDE DE INTERCONEXÃO DE ALTO DESEMPENHO ENTRE NÓS

Comunicação de até 113 Gb/seg

FDR InfiniBand (56 Gb/s)

Fully non-blocking

Global topology is “Fat tree” (capability of optimized fat tree routing)

◆ SISTEMAS DE ARQUIVOS PARALELOS

Largura de banda agregada de I/O > 30GB/s

Lustre v.2.1 file system

2,5 Petabytes de armazenamento

◆ AMBIENTE DE SOFTWARE PARA PERFILAR E DEPURAR ATÉ 20.000 NÚCLEOS EM PARALELO

Software stack:

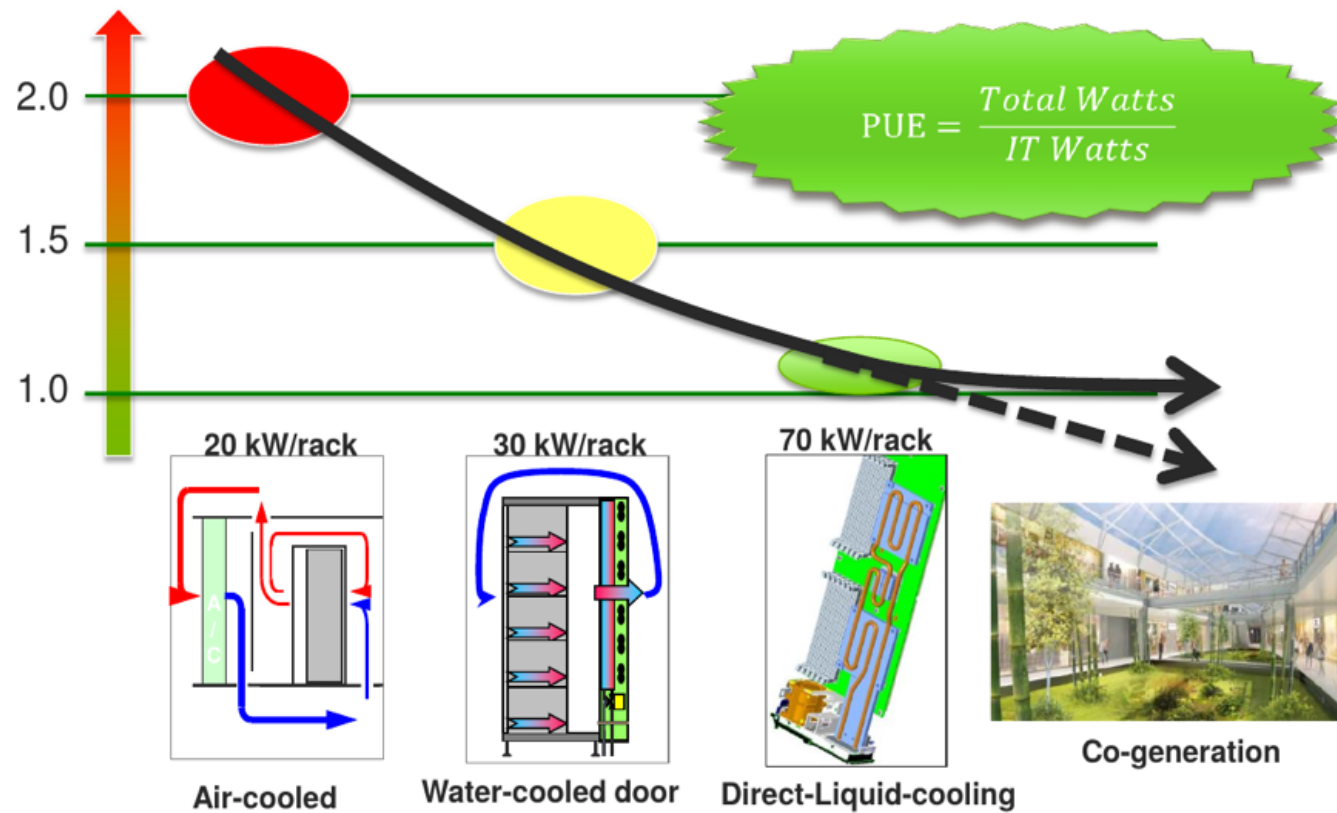
Red Hat 6.4;

Supercomputing Cluster suite AE4 u4;

INTEL compilers

Direct Liquid Cooling → Alta eficiência energética

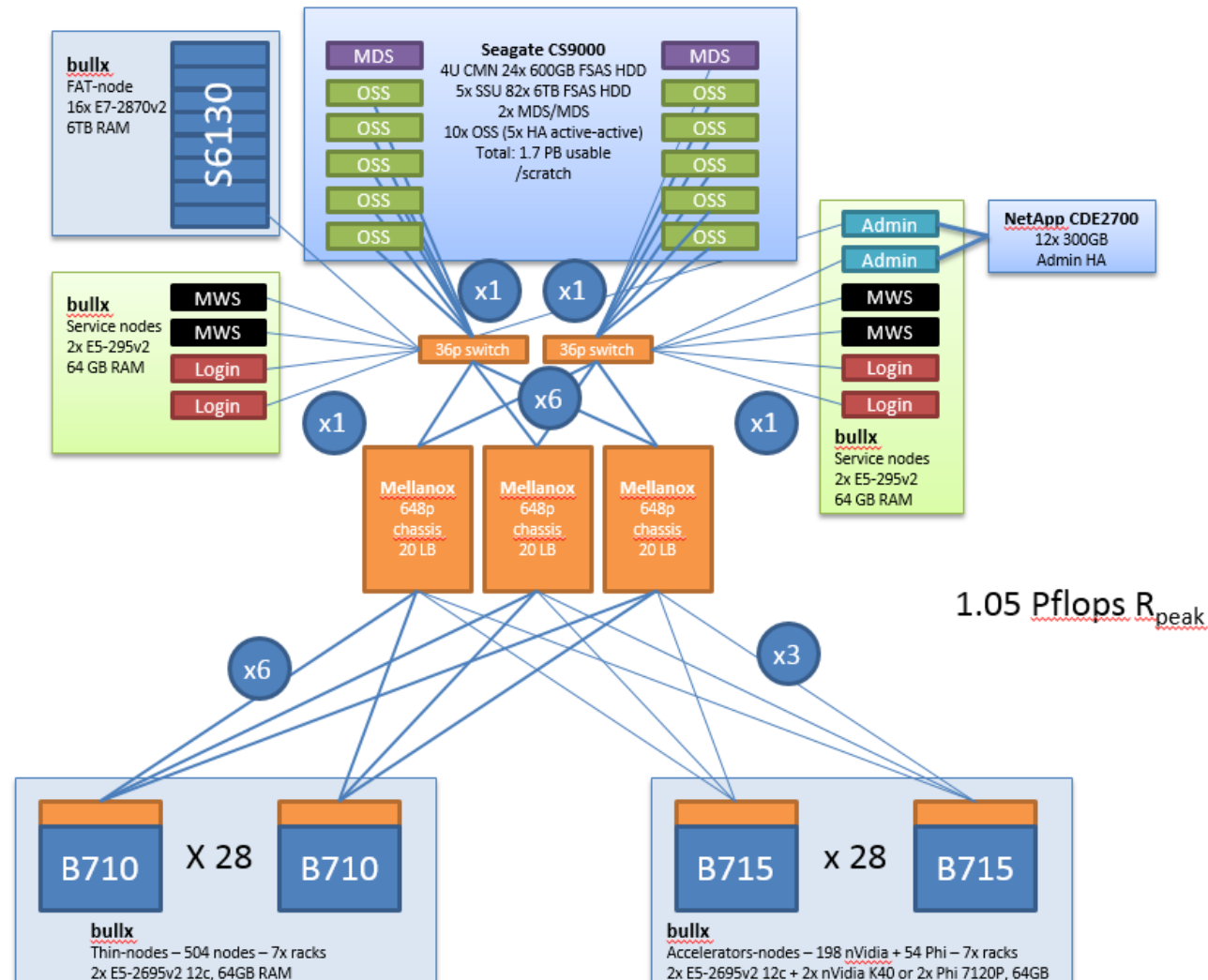
Cooling & Power Usage Effectiveness (PUE)



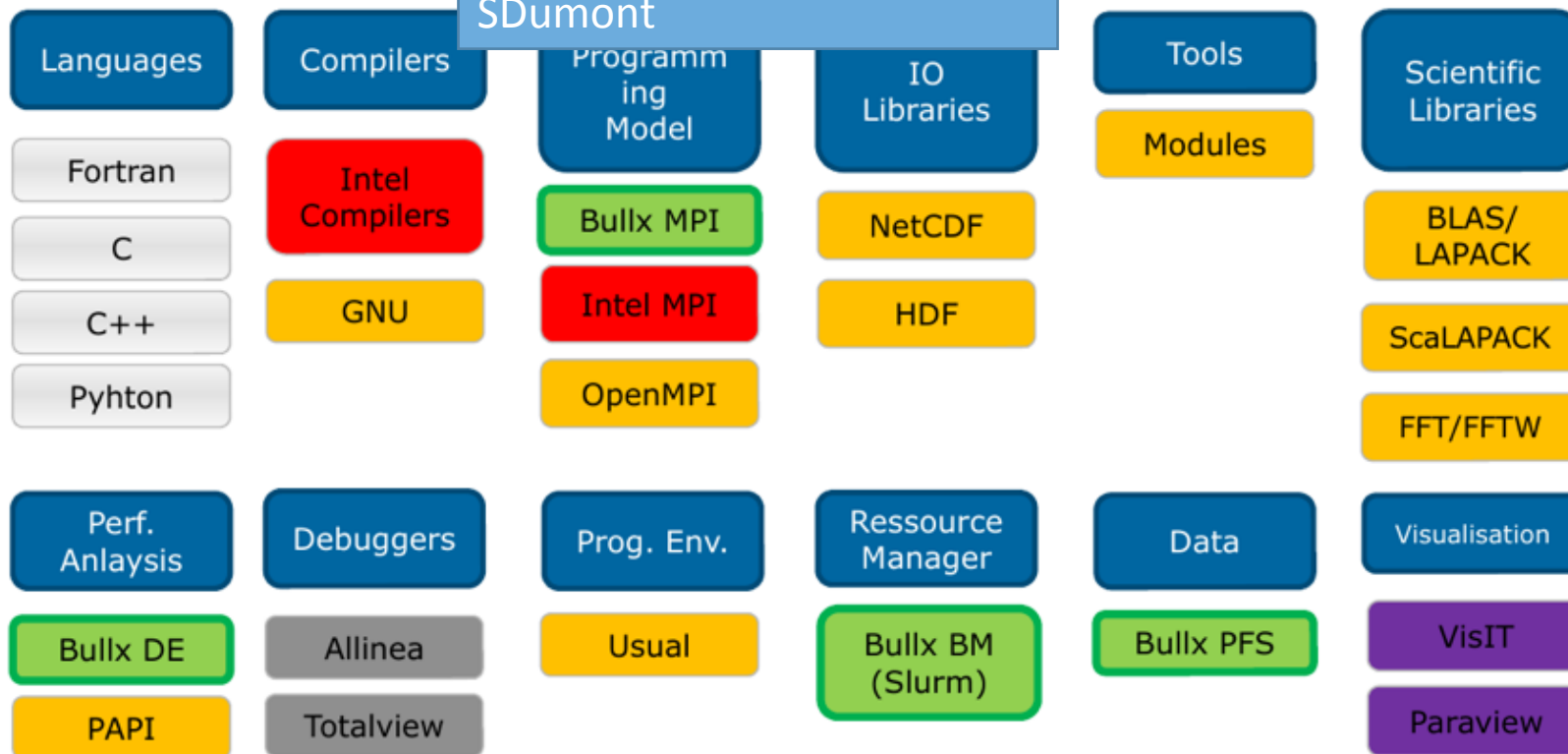
USO DO SUPERCOMPUTADOR SANTOS DUMONT



Rede de Interconexão de Alto Desempenho entre os nós de computação, comunicação de até 113Gb/s



Pilha de Software do SDumont



Bullx Linux

- Bull Products
- Standard GNU Package
- External OpenSource not provided
- External OpenSource provided
- External Commercial provided
- External Commercial not provided

Sistema Operacional:

Red Hat Enterprise Linux Server release 6.4

Gerenciamento de recursos:

Bullx Batch Manager (Bullx BM) baseado no SLURM 14.11

Linguagens:

Fortran, C, C++, Python, CUDA

Compiladores:

GNU, Intel, PGI

Implementação MPI:

Bullx MPI, Intel MPI e OpenMPI

Bibliotecas Científicas:

BLAS/LAPACK, ScaLAPACK, Intel MKL

Bibliotecas de I/O:

NetCDF, HDF5

Ferramentas de análise de desempenho:

PAPI
Bull performance monitor
HPC Toolkit
Open | SpeedShop
Intel Vtune, Intel Advisor

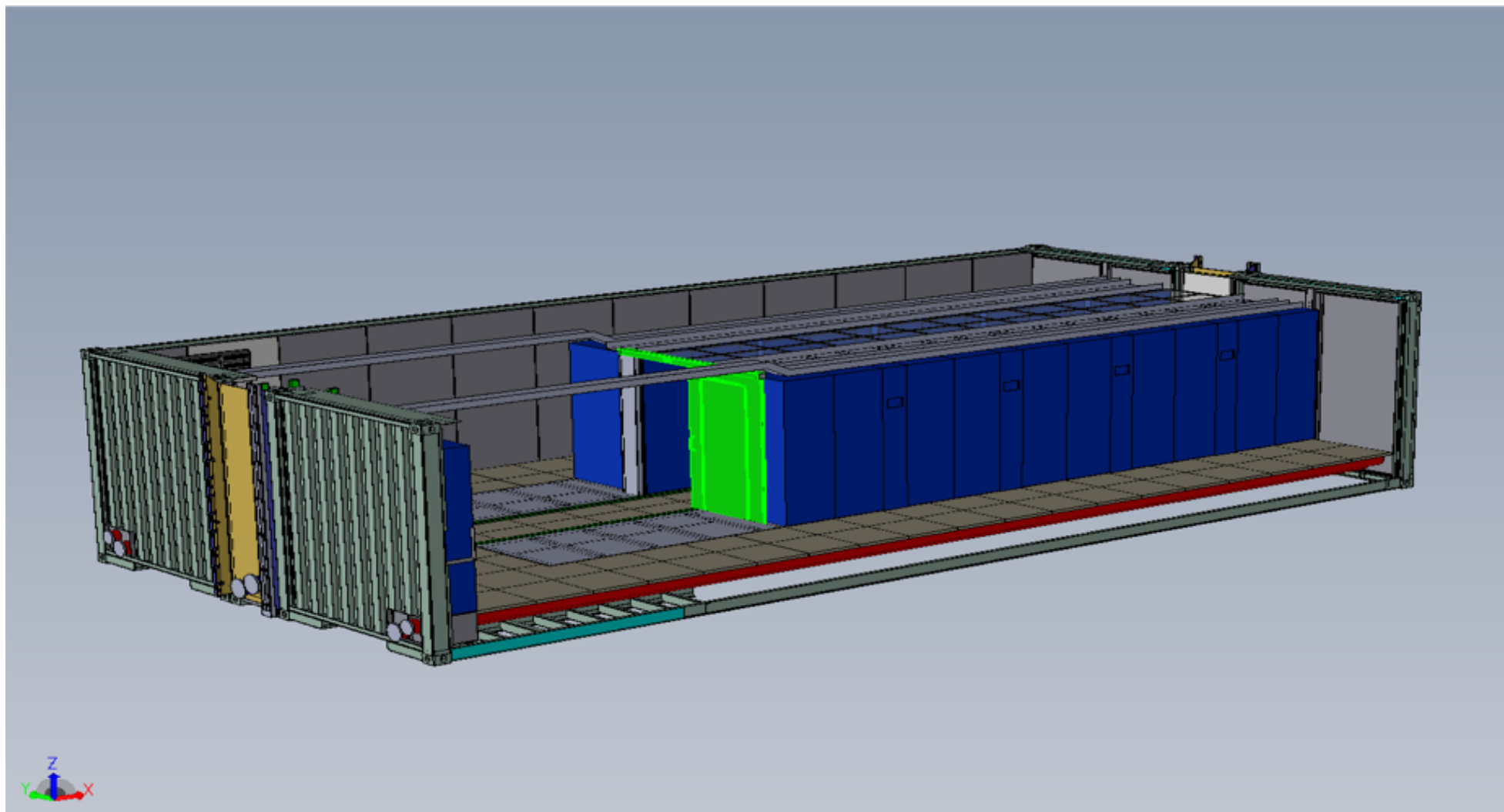
Ferramentas de análise de desempenho MPI:

Bull MPI Analyser
Scalasca
XMPI
Intel Itac / MPS

Ferramentas de análise de desempenho I/O:

IOTop]
Darshan

TECNOLOGIA DE CONTAINERS



USO DO SUPERCOMPUTADOR SANTOS DUMONT



Fase de TESTES do Santos Dumont com Aplicações Científicas e Tecnológicas

GAPF - Genetic Algorithm for Protein Folding

Predição de estrutura de proteínas. Modelagem Molecular com aplicações na construção de fármaco.

Parceria: LNCC, FIOCRUZ, UFPE, UFRJ

BRAMS - Brazilian developments on the Regional Atmospheric Modelling System

Modelo atmosférico de alta resolução, visando o teste de novas versões de um modelo do CPTEC/INPE.

Parceria: LNCC, CPTEC/INPE

Parallel Octree

Geração paralela de malhas para métodos de elementos finitos.

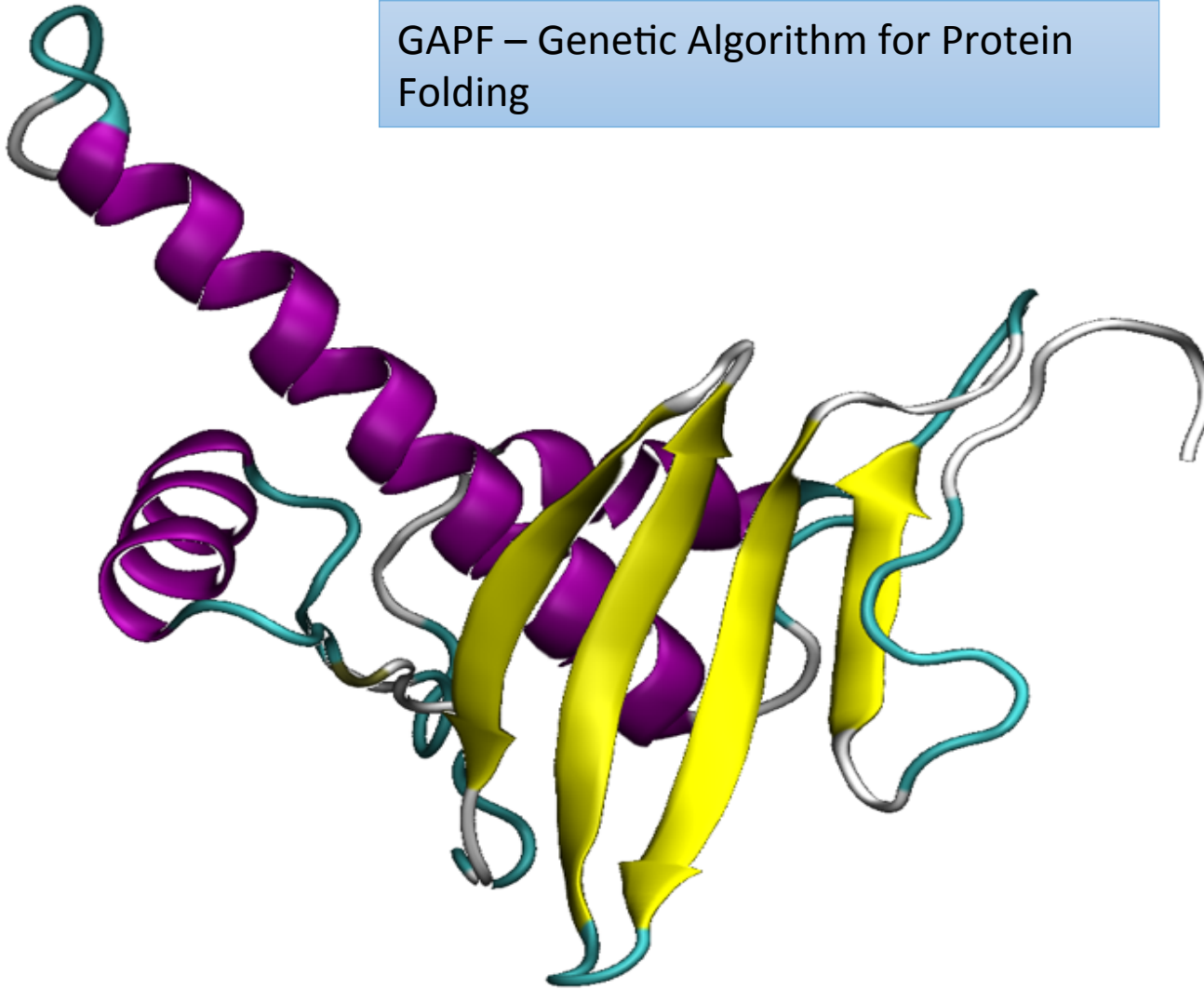
COPPE/UFRJ

MHM: Multiscale Hybrid-Mixed

Método de elementos finitos multi-escalas.

LNCC

GAPF – Genetic Algorithm for Protein Folding



GAPF: Genetic Algorithm for Protein Folding

Algoritmo estocástico no qual melhores resultados são obtidos quanto mais repetidas execuções são realizadas.

Assim, quanto mais capacidade de processamento tiver a máquina melhores são os resultados para um dado período de tempo de execução.

Esse foi um grande ganho que o SDUMONT proporcionou: permitir fazer muitas execuções ao mesmo tempo.

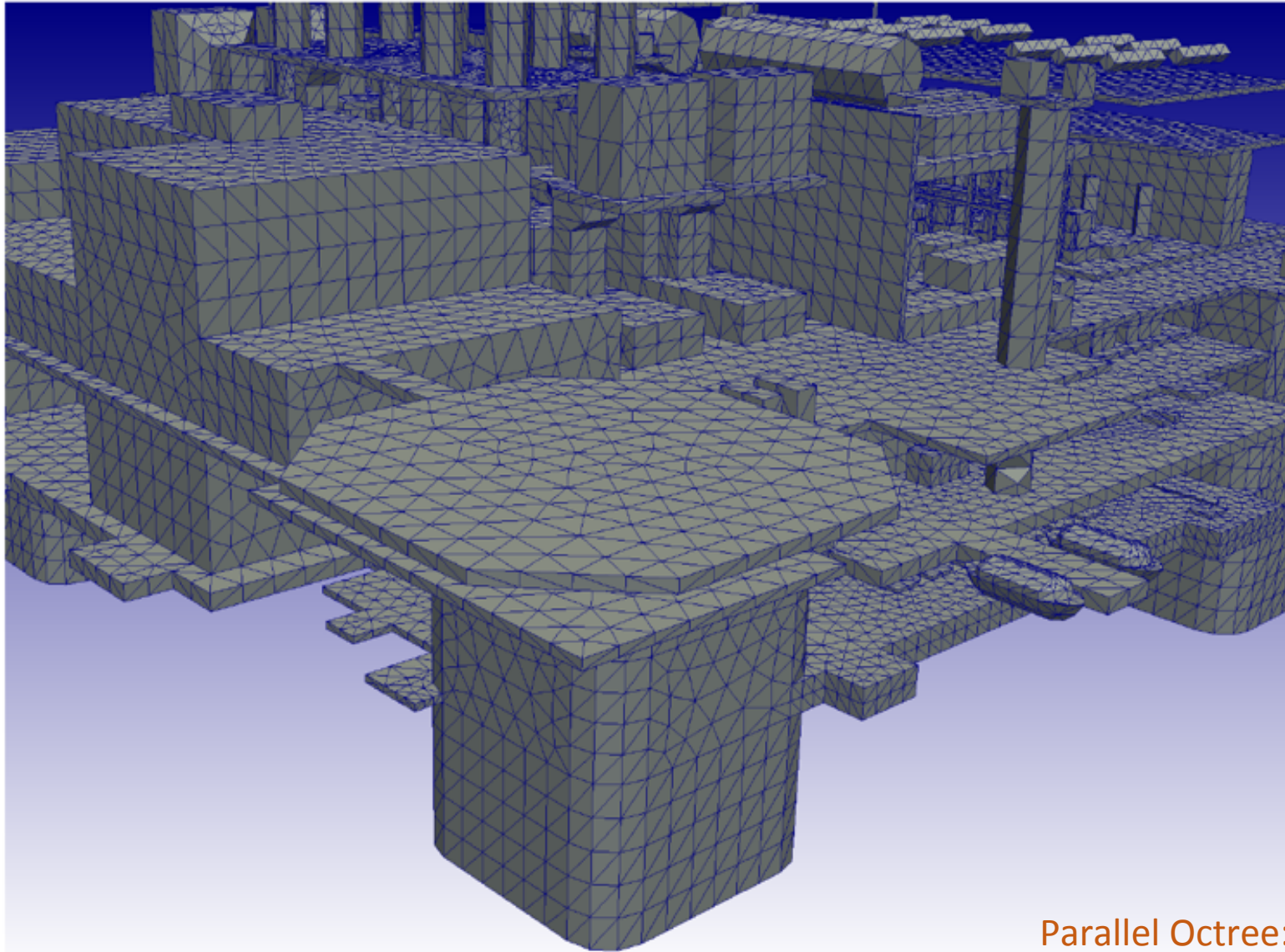
Atualmente estão sendo realizadas 100 execuções para cada alvo de aminoácidos, com resultados muito superiores aos obtidos com as 30 execuções realizadas anteriormente devido a restrições dos computadores então disponíveis.

BRAMS: BRAZILIAN DEVELOPMENTS ON THE REGIONAL ATMOSPHERIC MODELLING SYSTEM



Desempenho paralelo do BRAMS-5.0 no cluster SDUMONT

nodes	#cores	time(s)	Ideal speedup	Reached speedup	eficieny
64	1024	679	1.00	1.00	100%
128	2048	348	2.00	1.95	98%
256	4096	180	4.00	3.77	94%
512	8192	97	8.00	7.03	88%
645	10320	77	10.08	8.83	88%
670	13400	67	13.09	10.14	78%



modelo da
plataforma
Offshore
P-51 com
87.618
facetas

Parallel Octree: COPPE/UFRJ

Parallel Octree: COPPE/UFRJ

modelo da plataforma Offshore P-51 com 87.618 facetas

Cores	levels	elements	nodes	MPI	time (s)
64	11	296,394,446	3	BULL	29.80
256	12	1,182,028,956	11	BULL	38.18
1024	13	4,725,678,648	43	BULL	50.13
4096	14	18,895,327,518	171	BULL	74.94
16384	15	75,519,891,076	683	BULL	133.74

A cada aumento de nível de refinamento o tamanho do problema aumenta aproximadamente em quatro vezes (4x).

A fim de tentar manter o mesmo tempo de processamento a quantidade de cores também é aumentada em quatro vezes (4x).

MHM: Multiscale Hybrid-Mixed

(Método de elementos finitos multi-escalas)

Os métodos MHM são adaptados à resolução de modelos tridimensionais com múltiplas escalas ou grandes contrastes, caracterizando-se por alta ordem de precisão (baixas taxas de erro) e por incorporar a granularidade e heterogeneidade das novas gerações de arquiteturas massivamente paralelas.

Capacidade do algoritmo MHM em resolver em paralelo, modelos com bilhões de graus de liberdade em poucos minutos.

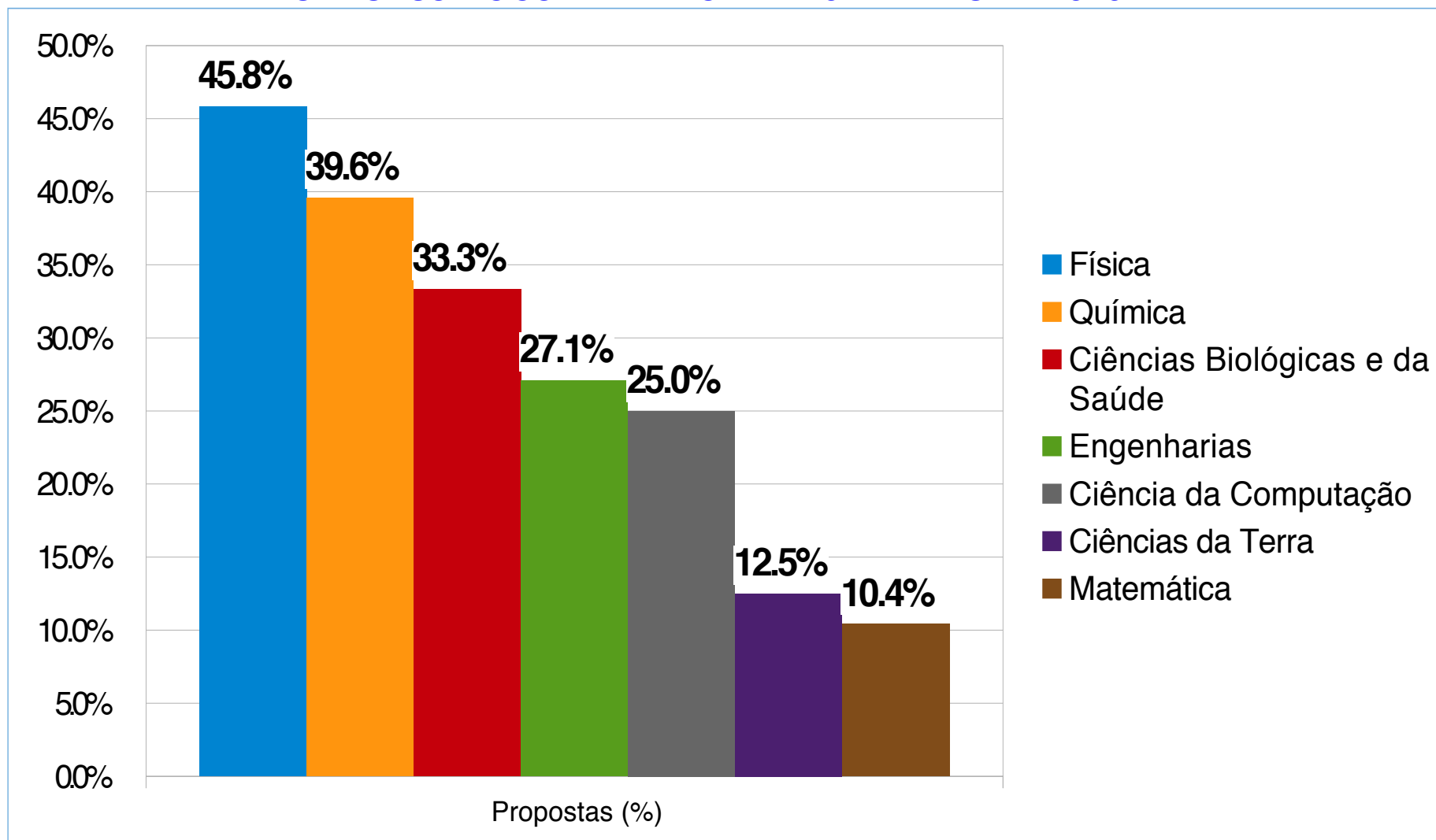
COMO USAR O SDUMONT

- Pesquisador de qualquer instituição de ensino ou pesquisa no Brasil submete proposta ao LNCC;
- Proposta deve demonstrar relevância técnica ou científica e propriedade no uso de computação de alto desempenho (HPC);
- Proposta é analisada por comissão de especialistas de várias instituições de pesquisa e universidades para recomendação de uso do SDumont;
- Informações detalhadas: <http://sdumont.lncc.br/>

48 PROPOSTAS SUBMETIDAS ATÉ 16 DE MAIO DE 2016

<u>SUDESTE</u>	-	<u>38</u>
<u>SUL</u>	-	<u>6</u>
<u>NORDESTE</u>	-	<u>2</u>
<u>NORTE</u>	-	<u>1</u>
<u>CENTRO-OESTE</u>	-	<u>1</u>

48 PROPOSTAS SUBMETIDAS ATÉ 16 DE MAIO DE 2016



17º **WRNP**

Workshop RNP

OBRIGADO !

Augusto Gadelha
diretoria@Incc.br



Ministério da
Cultura

Ministério da
Saúde

Ministério da
Educação

Ministério da
Ciência, Tecnologia
e Inovação

GOVERNO FEDERAL
BRASIL
PÁTRIA EDUCADORA