



# RES (Red Española de Supercomputación) & PRACE (Partnership for Advanced Computing in Europe)



Montserrat González  
Técnica de gestión de la RES

Jornadas de Difusión de la RES  
26 y 27 de Septiembre, 2011 – Islas Canarias



## Índice

1. **El BSC-CNS (*Barcelona Supercomputing Center-Centro Nacional de Supercomputación*)**
2. **MareNostrum**
3. **MariCel**
4. **La RES (*Red Española de Supercomputación*)**
5. **El proyecto PRACE (*Partnership for Advanced Computing in Europe*)**



# Índice

1. **El BSC-CNS (*Barcelona Supercomputing Center-Centro Nacional de Supercomputación*)**
2. MareNostrum
3. MariCel
4. La RES (*Red Española de Supercomputación*)
5. El proyecto PRACE (*Partnership for Advanced Computing in Europe*)



# BSC-CNS Centro Nacional de Supercomputación

## Misión del BSC-CNS

- Investigar, desarrollar y gestionar la tecnología para facilitar el progreso científico.

## Objetivos del BSC-CNS

- I+D en Ciencias de la Computación, Ciencias de la Vida y Ciencias de la Tierra.
- Soporte de la supercomputación para la investigación externa al BSC-CNS.

## BSC-CNS es un consorcio que incluye:

- Gobierno de España (MICINN) – 51%
- Gobierno de Cataluña (DIUE) – 37%
- Universitat Politècnica de Catalunya (UPC) – 12%





## Arquitectura de Computadores:

- Superscalar y VLIW
- Hardware multithreading
- Diseño de exploración del espacio de memoria de chips multicore y aceleradores
- Memoria transaccional (hardware y hardware-assisted)
- SIMD y extensiones/unidades vector

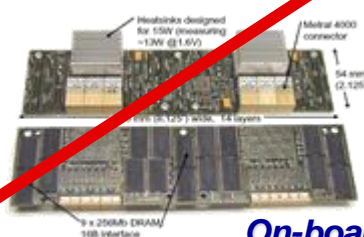
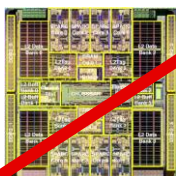
## Modelos de programación:

- Escalabilidad de MPI y UPC
- OpenMP para multicore, SMP y ccNUMA
- DSM para clusters
- CellSs, streaming
- Memoria transaccional
- Arquitecturas embebidas

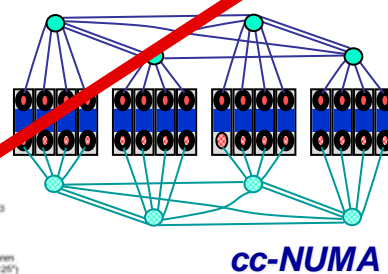
## Benchmarking, análisis y herramientas de predicción:

- Seguimiento de escalabilidad
- Identificación de patrones y estructuras
- Visualización i análisis
- Procesador, memoria, red y sistemas

Chip



On-board SMP



cc-NUMA



Grandes sistemas cluster



Futuros sistemas exaflop



Pequeños DMM



## Computación grid y cluster:

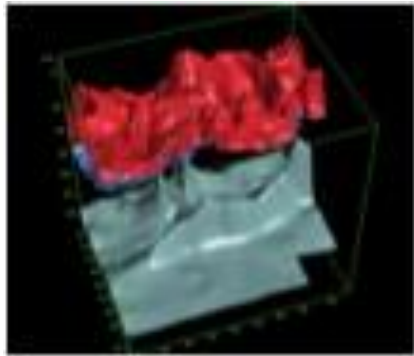
- Modelos de programación
- Gestión de recursos
- I/O para Grid

## Entornos operativos:

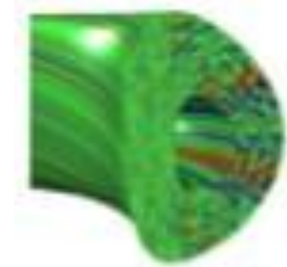
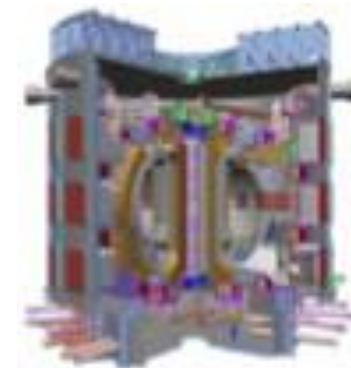
- Servidores de aplicaciones autónomos
- Gestión de recursos para cargas de trabajo heterogéneas
- Planificación coordinada y gestión de recursos
- Escalabilidad de sistemas de ficheros paralelos

# BSC-CNS: Aplicaciones en Ciencia e Ingeniería

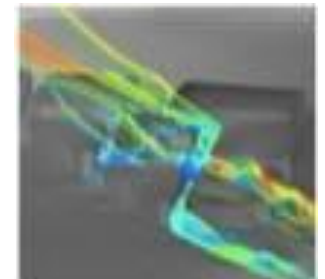
## Geofísica



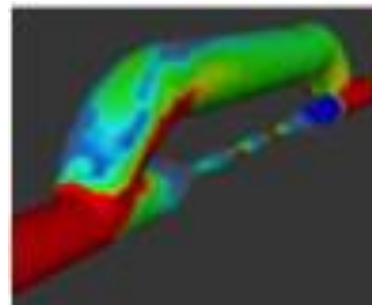
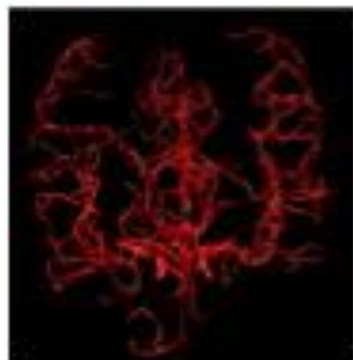
## ITER: Física del plasma



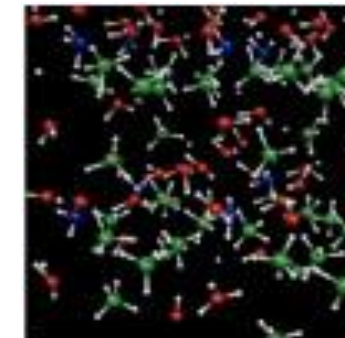
## Dinámica de fluidos computacional



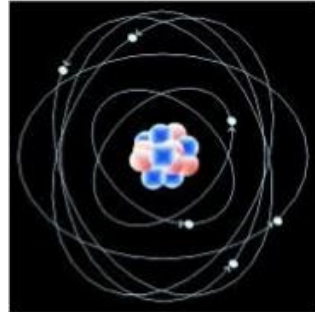
## Biomecánica



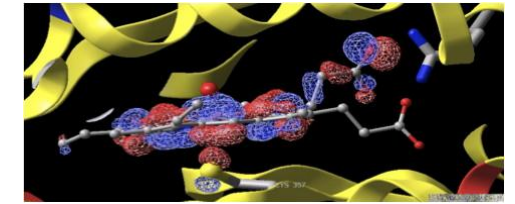
## Dinámica molecular ab-initio



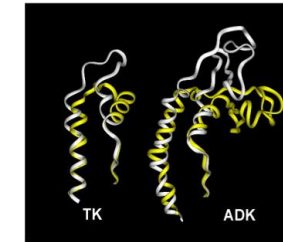




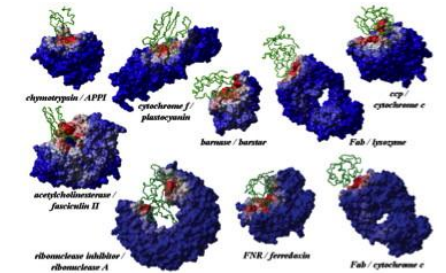
Modelado atómico (y electrónico) de bioquímica y biofísica de proteínas.



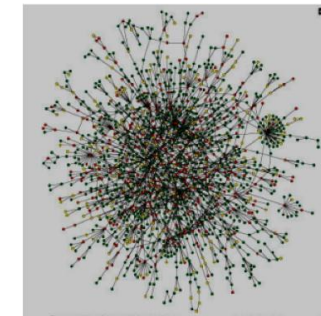
Modelado micro y mesoscópico de macromoléculas.  
Diseño de drogas.



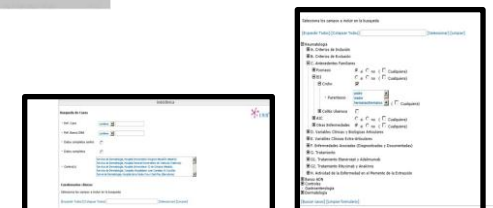
Identificación de bases estructurales de interacción proteína-proteína.



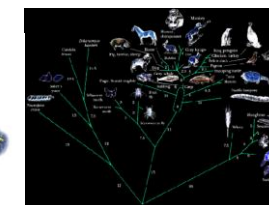
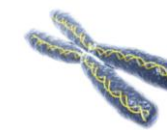
Redes de interacción proteína-proteína  
Sistemas biológicos.



Servicios web, aplicaciones y bases de datos.



Análisis de genomas y redes para modelar enfermedades, sistemas y evolución de organismos.





## Predicción de la calidad del aire

**Proyecto CALIOPE (financiado por el Ministerio de Medio Ambiente, Ref. 441/2006/3-12.1)**

MINISTERIO DE MEDIO AMBIENTE | BSC | Barcelona Supercomputing Center | Ciemat | MINISTERIO DE EDUCACIÓN Y CIENCIA | CSIC | CEAM

**Navegación**

- Inicio
- Descripción
- Justificación
- Implementación
- Pronóstico Calidad del Aire
- Pronóstico Meteorológico
- Pronóstico Emisiones
- Verificación: imágenes de satélite
- Participantes
- Publicaciones
- Enlaces
- Contacto
- Acceso privado

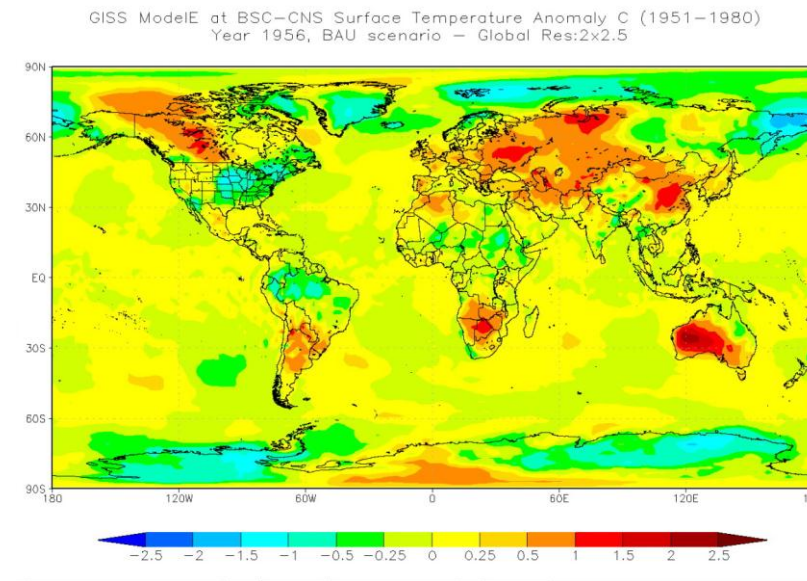
**Proyecto Caliope:**

El proyecto CALIOPE es un proyecto financiado por el Ministerio de Medio Ambiente (CALIOPE 441/2006/3-12.1) y cuyo principal objetivo es el desarrollo de un sistema de modelización de la calidad del aire operativo para España que proporcione un servicio de pronóstico de la calidad del aire con elevada resolución espacial para la Península Ibérica, Islas Baleares e Islas Canarias con andamientos en áreas urbanas con muy alta resolución.

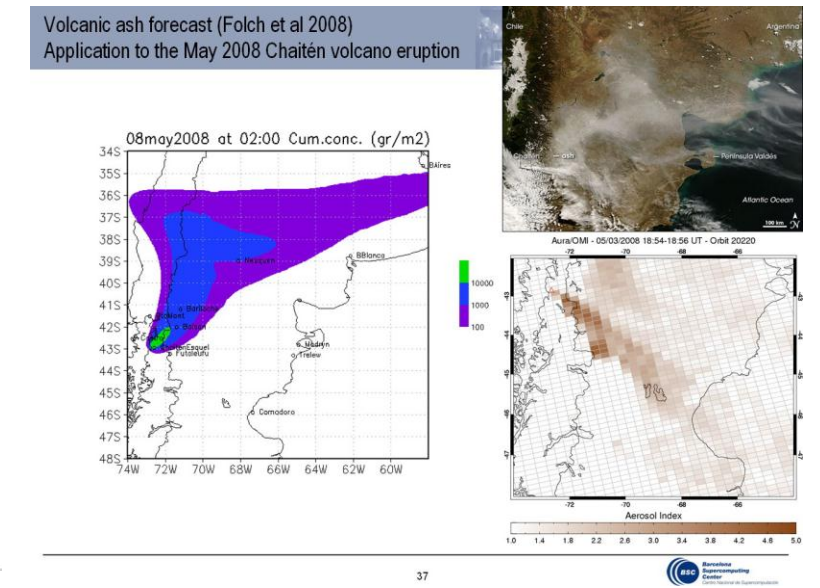
**NOTA:** El sistema se encuentra en fase de desarrollo, por lo que los resultados tienen un carácter preliminar y están en proceso de validación.

Europa (MareNostrum; 12km-ARW-CMAQ) | Península Ibérica e Islas Baleares (MareNostrum; 4km-ARW-CMAQ)

## Cambio climático



## Transporte de ceniza volcánica



## Transporte de polvo mineral

WMO Sand and Dust Storm Warning and Assessment System (SDS WAS)

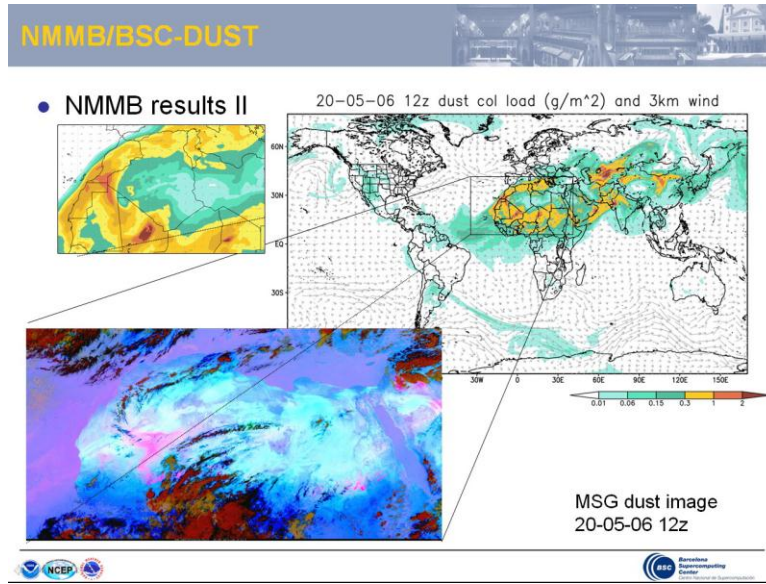
- To enhance the ability of participating countries to establish and improve systems for forecasting and warning to suppress the impact of Sand and Dust Storm by
- Establishing a coordinated global network of Sand and Dust Storm forecasting centers delivering products useful to a wide range of users in understanding and reducing the impacts of SDS

North Africa, Middle East and Europe | Asia

BSC-CNS AEMET, Spain | China Meteorological Administration (CMA) Xiao Ye Zhang xiaoye@cams.cma.gov.cn

WMO REGIONAL CENTRES

## Desarrollo de un modelo global de polvo mineral



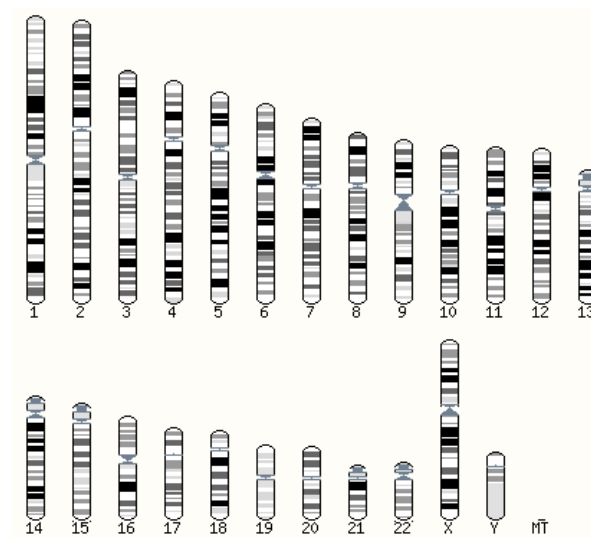
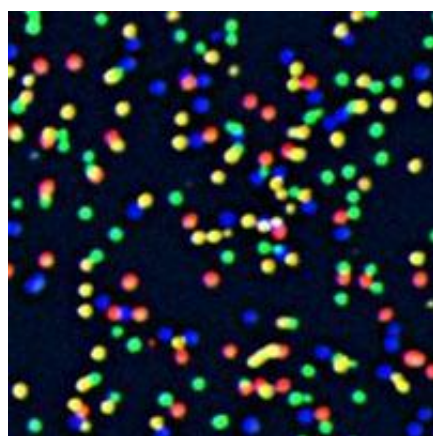
## Transferencia tecnológica: Estudios de impacto ambiental





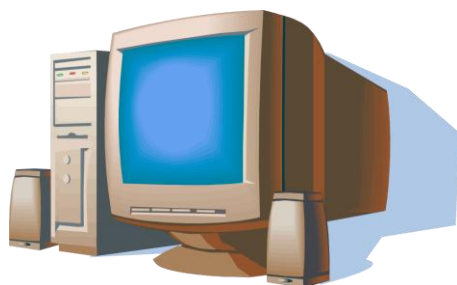
# **BSC-CNS: CNAG, Centro Nacional para el Análisis Geonómico**

- BSC-CNS proporciona servicio al CNAG a nivel de HPC, IT y datos.
- Secuenciación de nueva generación.



**Secuenciación rápida de individuos**

**Estudios detallados de procesos celulares**



**Raw Data:  
1-2TB/ejecución  
2 ejecuciones/semana  
10 máquinas**

**Procesado de imagen para generar secuencias de datos**

**Análisis de secuencia, alineación y clustering**

**Alineación de resultados  
250-500 GB/ejecución**



# Índice

1. **El BSC-CNS (*Barcelona Supercomputing Center-Centro Nacional de Supercomputación*)**
- 2. MareNostrum**
3. MariCel
4. La RES (*Red Española de Supercomputación*)
5. El proyecto PRACE (*Partnership for Advanced Computing in Europe*)



# MareNostrum







# MN: Evolución de MareNostrum

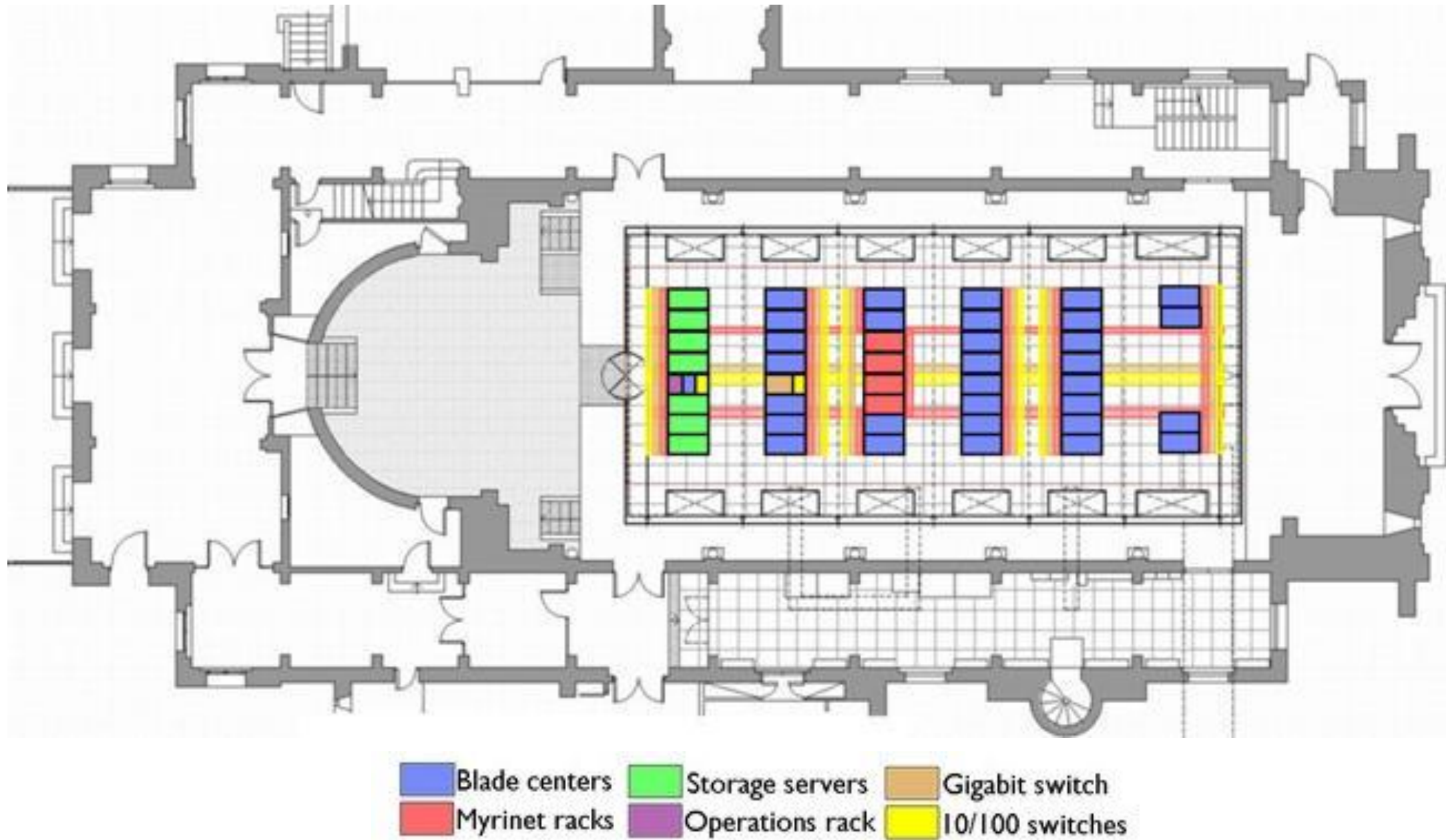
## MareNostrum 2004

- 42,35 Tflops capacidad máxima
- 4812 PowerPC 970 cores
- 2406 JS20 2.2 GHz
- 10 TB
- 4 GB por nodo
- 270 TB capacidad total de almacenamiento
- 3 redes
  - Myrinet
  - Gigabit
  - 10/100 Ethernet
- Sistema operativo
  - Linux 2.6 (SuSE)

## MareNostrum 2006

- 94,21 Tflops capacidad máxima
- 10240 PowerPC 970 cores
- 2560 JS21 2.3 GHz
- 20 TB de memoria
- 8 GB por nodo
- 480 TB capacidad total de almacenamiento
- 3 redes
  - Myrinet
  - Gigabit
  - 10/100 Ethernet
- Sistema operativo
  - Linux 2.6 (SuSE)

# Arquitectura de MareNostrum



*Plano de la capilla de Torre Girona donde se aloja MareNostrum*



## MN: Racks de Blade Centers (1/2)

- Hay 31 racks dedicados a cálculo con un total de 10240 procesadores PowerPC 970 a 2,3 GHz y con 20TBytes de memoria.
- Cada rack contiene 6 Blade Centers con 336 procesadores, 672 GBytes de memoria y proporciona una potencia máxima de cálculo de 3,1 Tflops.

### Blade Center

- Agrupa 14 nodos o blades JS21 con 2 procesadores duals, que suponen un total de 56 procesadores.
- Dispone de 2 entradas de alimentación para redundancia.
- Incluye un switch gigabit que conecta todos sus nodo.



*Blade Center IBM JS21*





## MN: Racks de Blade Centers (2/2)

### Blade o nodo JS21

- Formado por 2 procesadores duales PowerPC 970MP a 2.3 GHz, 8 Gbytes de memoria compartida y un disco SAS local de 36 GBytes.
- Contiene una tarjeta M3S-PCIXD-2-I que permite la interconexión a Myrinet y 2 conexiones a la red Gigabit.
- Puede funcionar sin disco, ya que el sistema operativo no está en el disco, sino en los racks de almacenamiento. Cuando un nodo se inicializa, el S.O. se carga mediante la red Gigabit.



*Blade IBM JS21*

### Procesador PowerPC 970MP

- Superscalar con arquitectura de 64 bits de propósito general
- Máximo de 4 instrucciones por ciclo de reloj y 200 instrucciones in-flight.

## MN: Racks de Myrinet

- Hay 4 racks dedicados a la interconexión de los 2560 nodos JS21 a través de la red de alta velocidad Myrinet y cables de fibra óptica.



*3 racks con elementos de interconexión de Myrinet*



## MN: Racks de servidores de almacenamiento



*Racks de los servidores de almacenamiento*

- Hay 7 racks con 20 servidores de almacenamiento.
- Los 20 servidores disponen de 560 discos de 512GBytes que suponen un total de 280 TBytes de almacenamiento externo.
- Los 2560 nodo acceden a los discos a través de la red Gigabit.
- Los discos usan el sistema de ficheros GPFS (Global Parallel File System) que les ofrece una visión global del sistema de archivos y les permite el acceso en paralelo.





## MN: Rack de operaciones

- Hay 1 racks de operaciones, desde el que puede gestionarse el sistema a través del acceso a la consola de la máquina.
- Su contenido es el siguiente: :
  - 1 Monitor 7316-TF3
  - 2 nodos p615
  - 2 HMC (consolas) 7315-CR2
  - 3 nodos remotos asíncronos
  - 1 chasis BCIO BladeCenter IO
  - 4 Switches Cisco 3550



## MN: Rack de switch gigabit

- 1 rack de MareNostrum está dedicado a la interconexión de la red Gigabit y una parte de los elementos de la red Ethernet 10/100.
- Los elementos de este rack son:
  - 1 switch Force10 E600 Gigabit Ethernet. Contiene 7 slots en total, 6 de los cuales tienen tarjetas 10/100/1000 Base-T de 48 puertos que ofrecen un total de 288 puertos GigabitEthernet/IEEE 802.3
  - 4 Switches Cisco 3550 48-puertos Fast Ethernet

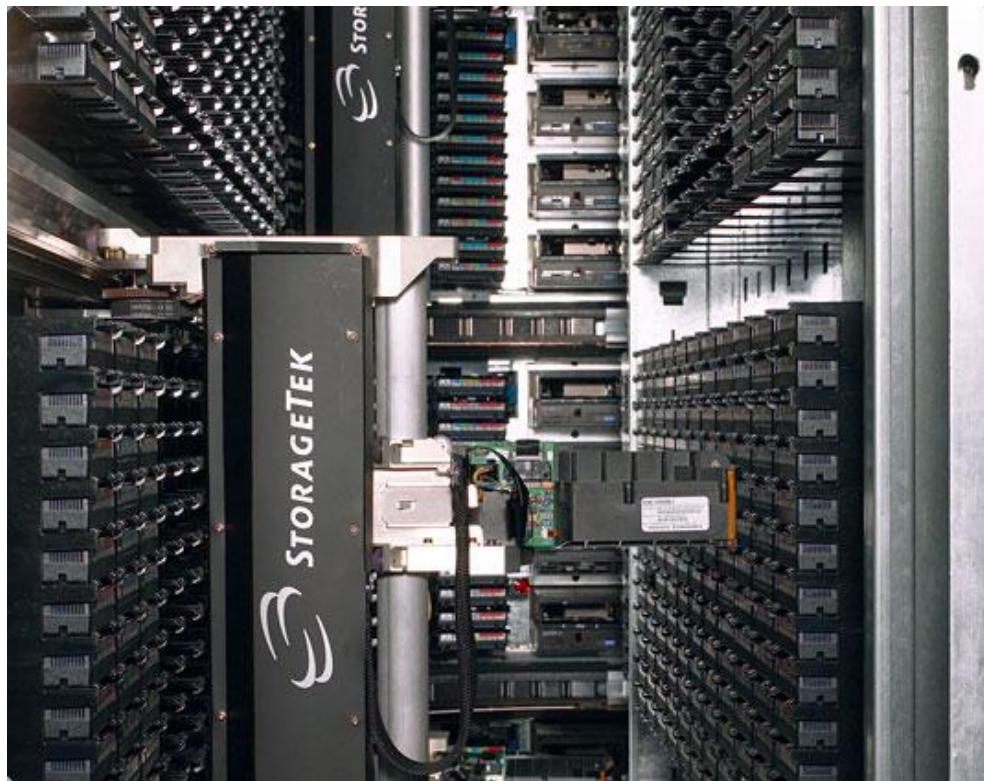


*Switch Force10 E600*



## Robot SL8500: HSM / Backup

- 7500 LTO4 cintas de 800 GB cada una  $\approx$  6 PB nativos
- 8 brazos trabajan a 4 alturas diferentes
- 23 LTO 4 unidades de lectura
- 5 Servidores (Sun Fire v445, 4 CPU UltraSPARC III, 8 GB RAM)
- Armarios de discos de 35 TB



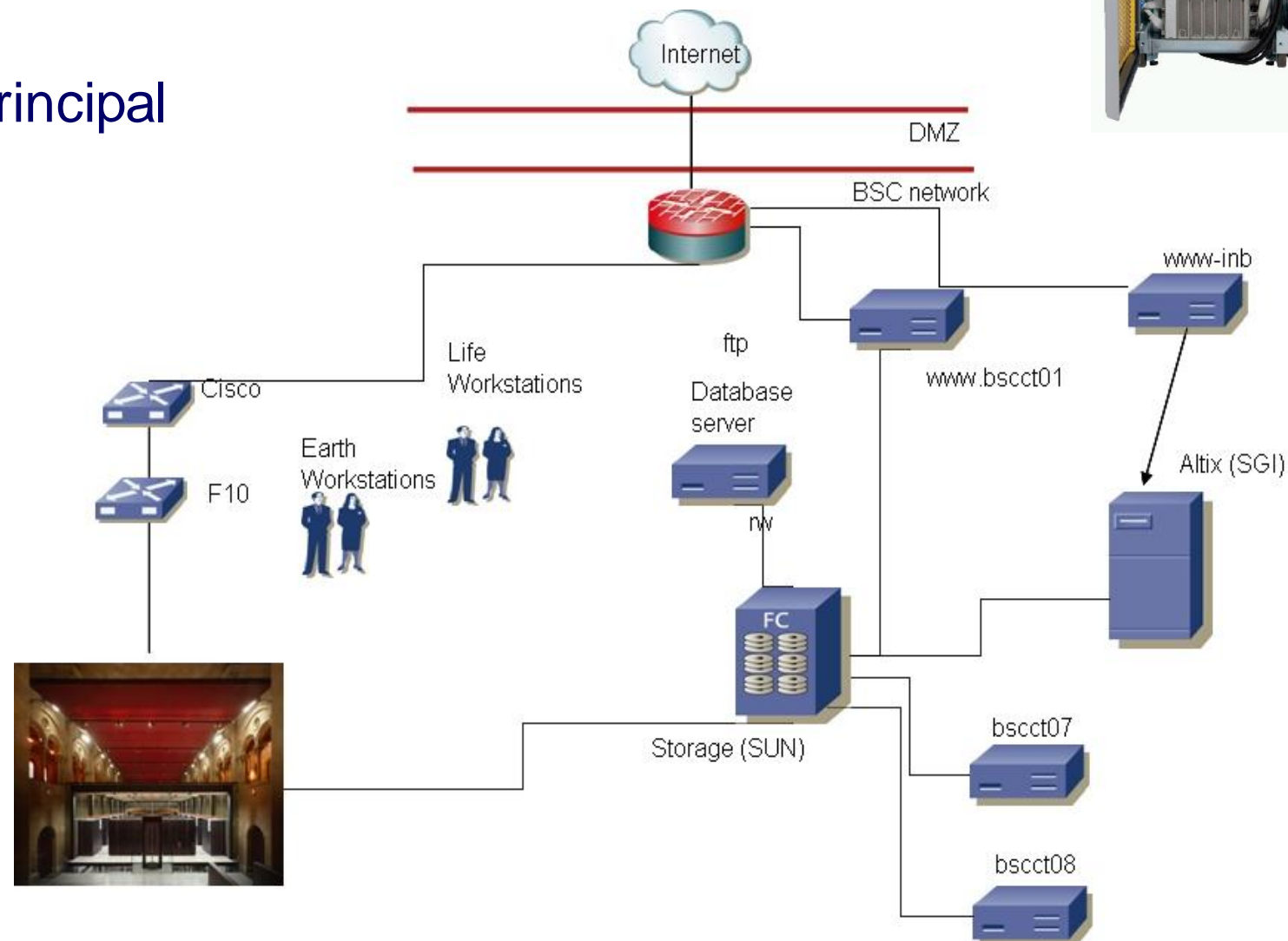
Robot SL8500



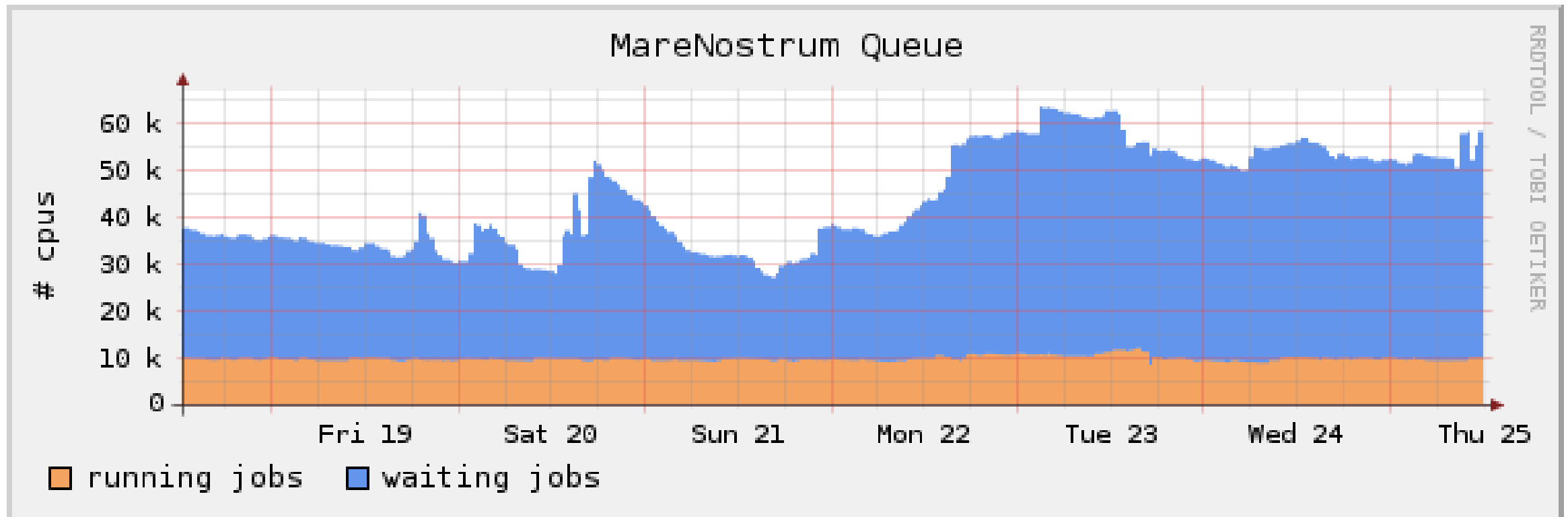
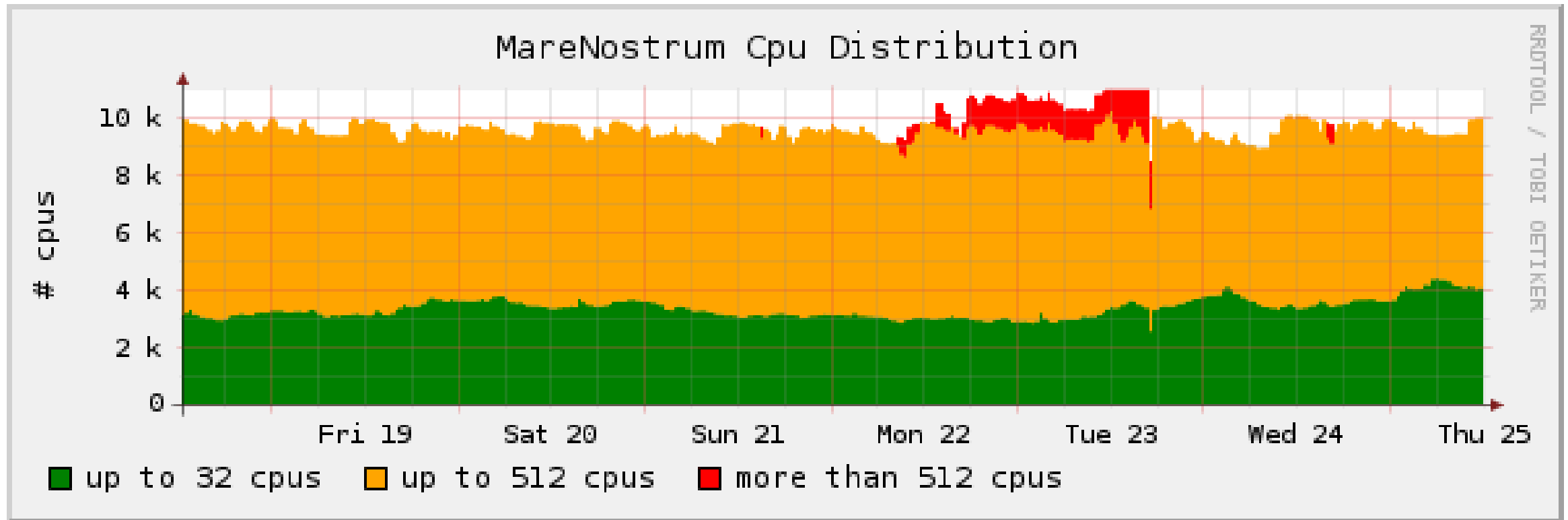


# Otro sistemas

- Sistema de memoria compartida
  - 64 nodes
    - 2 Montecito, dual core 1.6, total 256 cores
    - Cache 8 MB
  - 2.5 TB de memoria principal
- Databases
- PowerPC Cluster



# Rendimiento de MareNostrum





# Índice

1. El BSC-CNS (*Barcelona Supercomputing Center-Centro Nacional de Supercomputación*)
2. MareNostrum
- 3. MariCel**
4. La RES (*Red Española de Supercomputación*)
5. El proyecto PRACE (*Partnership for Advanced Computing in Europe*)





# MariCel: prototipo español en PRACE

## MariCel

- 72QS22 IBM servidores de Blades PowerXCell 8i 3.2Ghz CPUs
- 864 GB total de RAM
- 1296 cores
- 12 JS22 IBM servidores de Blades Power6 64 bits 4.0 GHz dual core
- 96 GB RAM total
- 48 cores
- 4xDDR InfiniBand (16GBytes/s de datos)
- MPI, GPFS
- 4 x 24 ports de switch
- Máximo rendimiento 14.4 Tflops en 2 unidades de rack
- 20 kW de consumo



MariCel está en 2 racks de la sala de MareNostrum



IBM QS22 Blade Center





## Índice

1. El BSC-CNS (*Barcelona Supercomputing Center-Centro Nacional de Supercomputación*)
2. MareNostrum
3. MariCel
4. **La RES (Red Española de Supercomputación)**
5. El proyecto PRACE (*Partnership for Advanced Computing in Europe*)



# La RES: Miembros de la RES



## BSC-CNS (MareNostrum)

Procesador: 10240 PowerPC 970 2.3 GHz  
 Memoria: 20 TB  
 Disco: 280 + 90 TB  
 Redes: Myrinet, Gigabit, 10/100  
 Sistema: Linux

## UPM (Magerit II - renovado en 2011)

Procesador: 3.920 IBM Power7 3.3 GHz  
 Memoria: 8700 GB  
 Disco: 190 TB  
 Redes: Infiniband, Gigabit  
 Sistema: Linux

## IAC, UMA, UC, UZ, UV (LaPalma, Picasso, Altamira, Caesaraugusta, Tirant)

Procesador: 512 PowerPC 970 2.2 GHz  
 Memoria: 1 TB  
 Disco: 14 + 10 TB  
 Redes: Myrinet, Gigabit, 10/100  
 Sistema: Linux

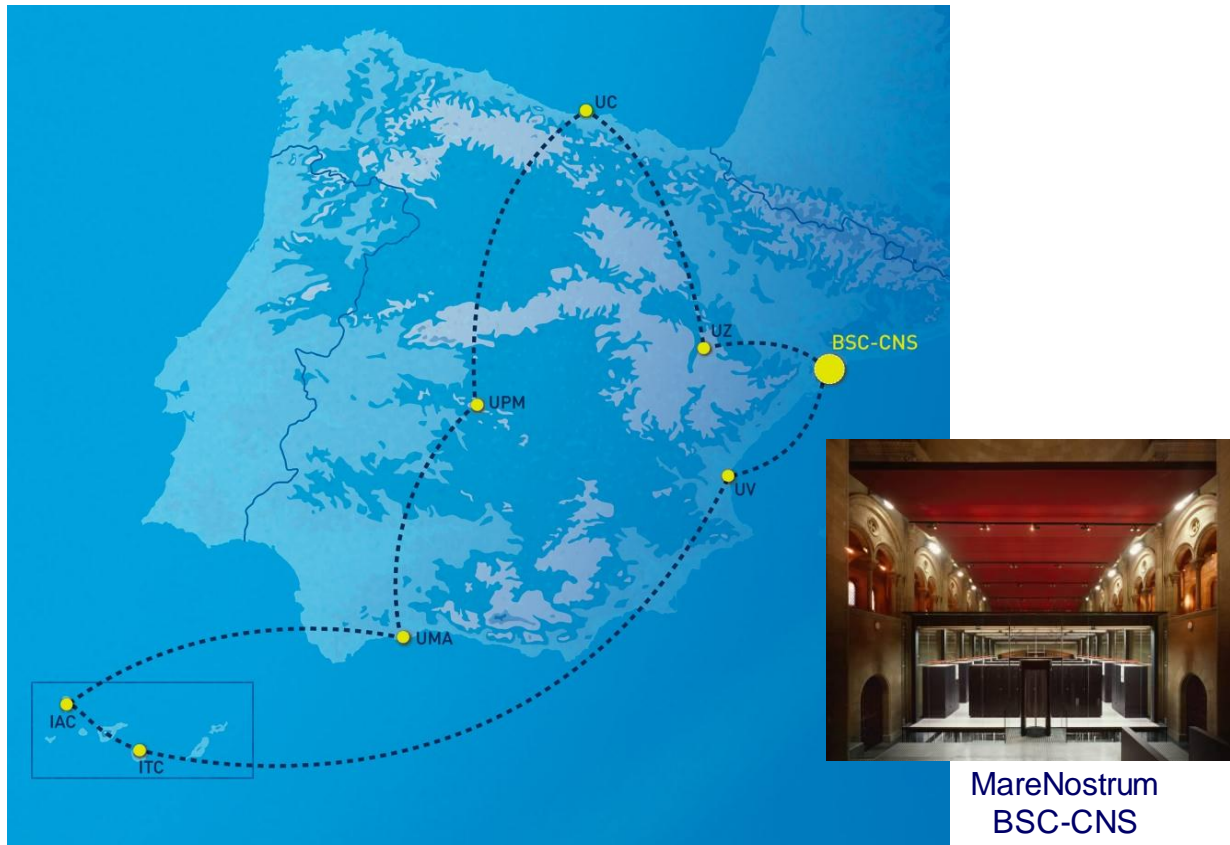
## Gobierno de las Islas Canarias - ITC (Atlante)

Procesador: 336 PowerPC 970 2.3 GHz  
 Memoria: 672 GB  
 Discp: 3 TB  
 Redes: Myrinet, Gigabit, 10/100  
 Sistema: Linux

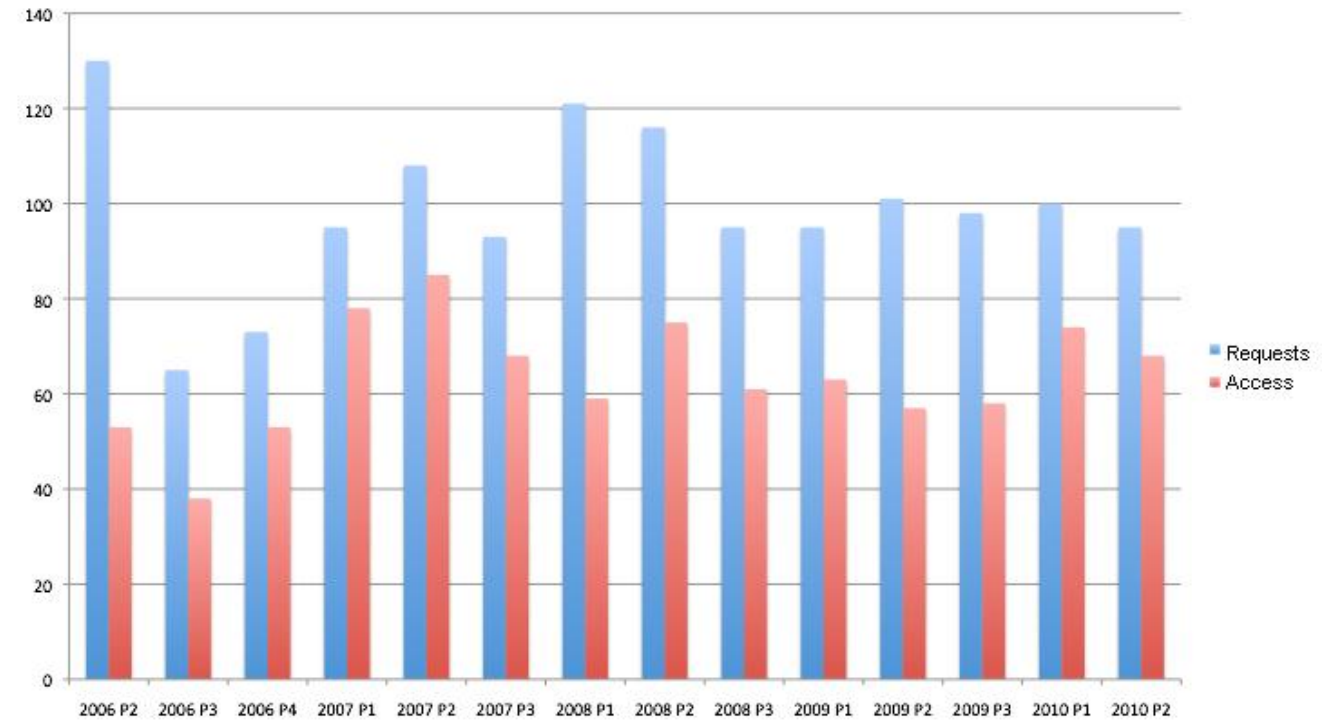




# BSC-CNS: Vertebrador de la supercomputación en España

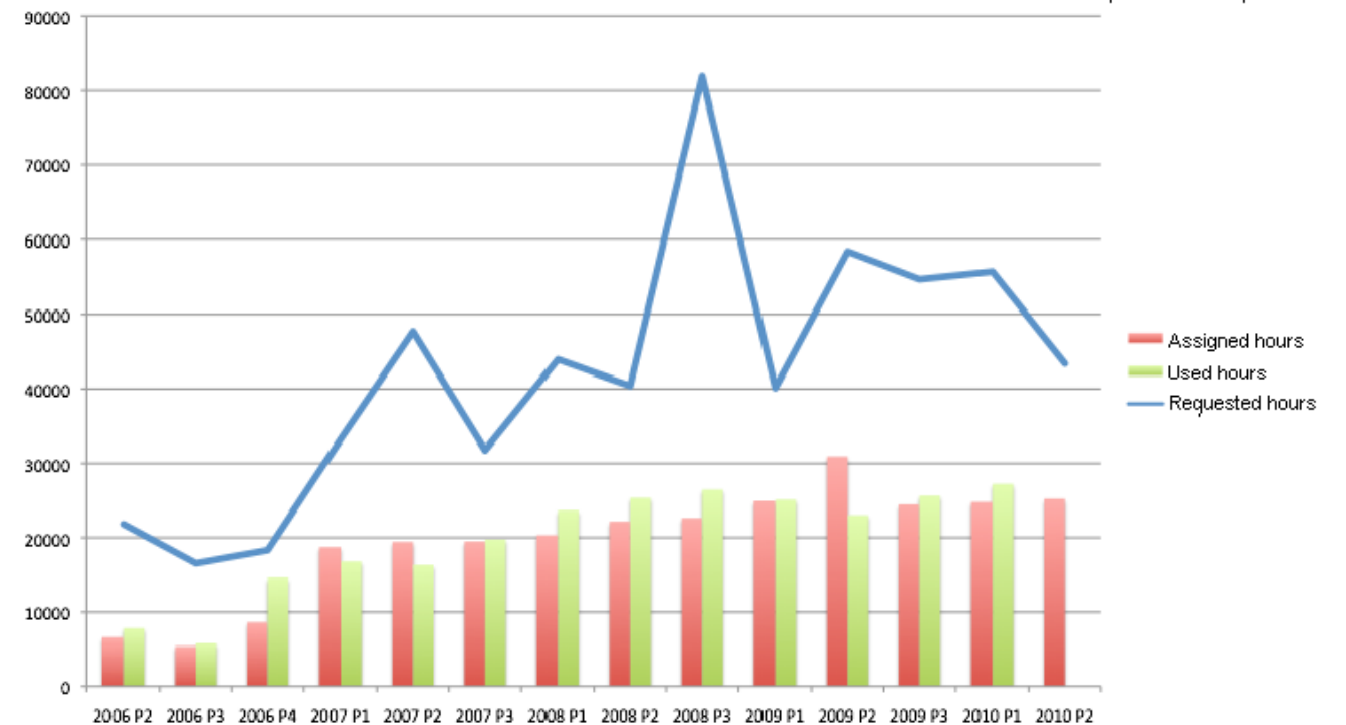


Actividades en la RES

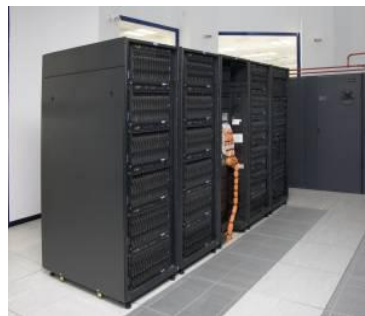


Miles de horas en la RES

Data updated June 30, 2010



Magerit  
Universidad Politécnica de Madrid



La Palma  
IAC



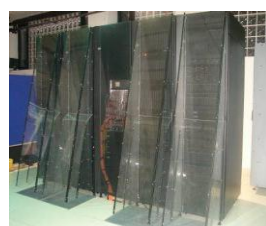
Atlante  
ITC



CaesarAugusta  
Universidad de Saragossa



Picasso  
Universidad de Màlaga



Altamira  
Universidad de Cantabria



Tirant  
Universitat de València



# Proyectos científicos en la RES

Astronomía, Espacio y Ciencias de la Tierra	12	17	12	16	21	14	14	11	14	12	14	11	14	18	16	216
Biomedicina y Ciencias de la Vida	20	26	18	18	26	24	20	28	24	22	23	26	29	24	19	347
Física e Ingeniería	19	24	17	20	32	29	22	21	32	26	24	21	23	21	25	356
Química y Ciencia y Tecnología de los Materiales	21	27	23	18	33	32	31	26	29	26	22	22	16	23	26	375
Investigación del BSC-CNS	14	19	14	14	22	29	17	17	20	17	17	16	16	17	17	258
	2005	2006			2007			2008			2009			2010		1552



## Índice

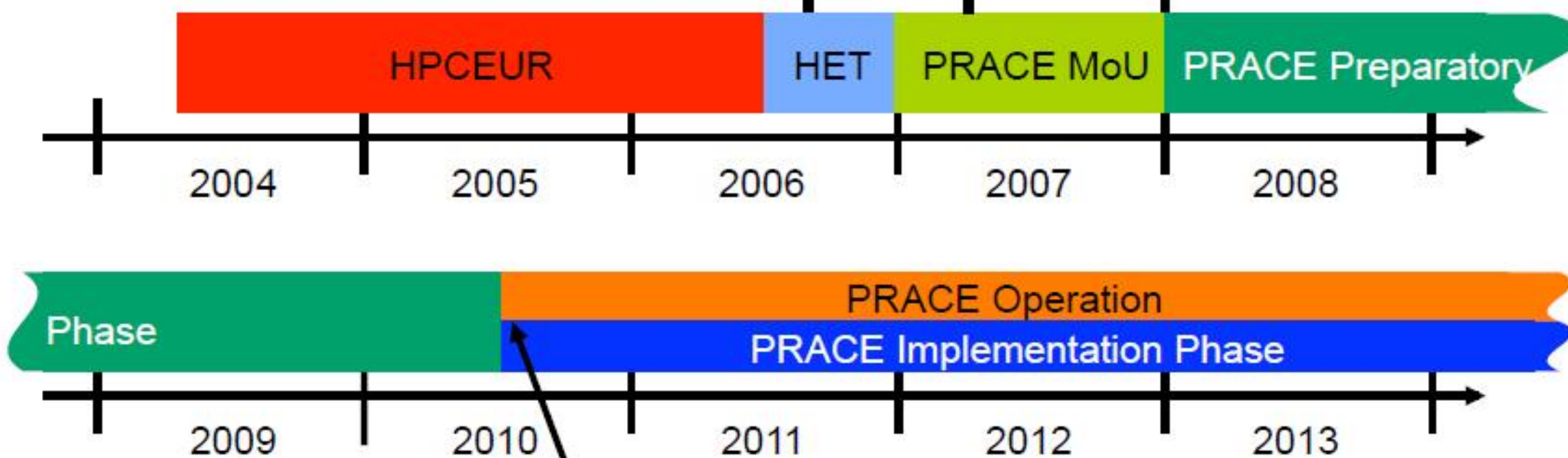
1. El BSC-CNS (*Barcelona Supercomputing Center-Centro Nacional de Supercomputación*)
2. MareNostrum
3. MariCel
4. La RES (Red Española de Supercomputación)
5. El proyecto PRACE (*Partnership for Advanced Computing in Europe*)



PARTNERSHIP  
FOR ADVANCED COMPUTING  
IN EUROPE



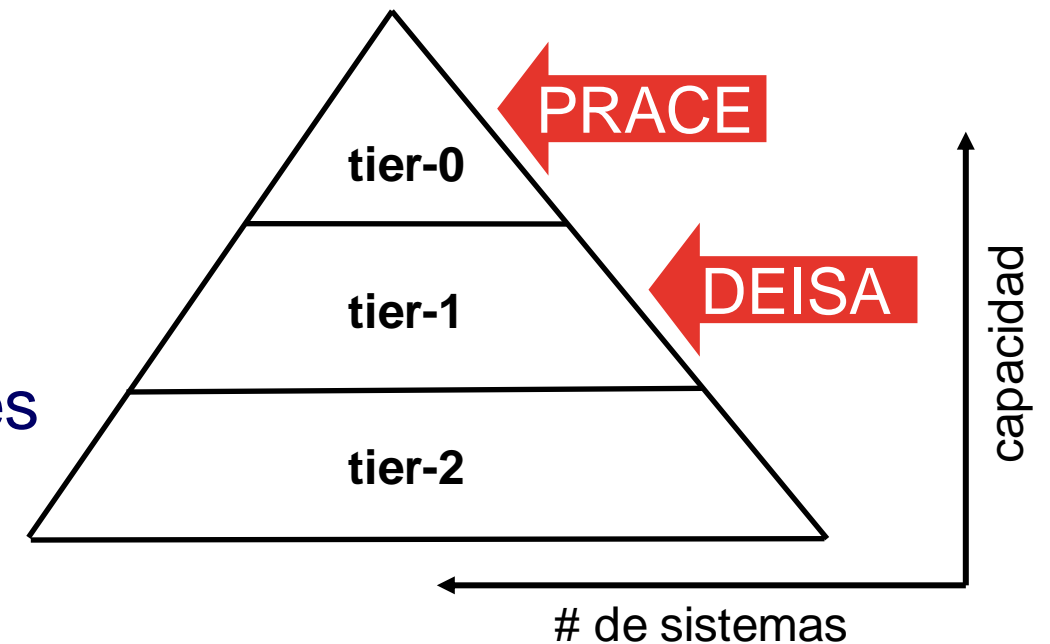
## Historia de PRACE y primeros pasos



Los centros Tier-0 proporcionan capacidad y servicio HPC a través de una entidad legal

## Visión ESFRI del servicio HPC Europeo

- Tier-0: 3-5 centro europeos
- Tier-1: centros nacionales
- Tier-2: centros regionales/universidades



- Los sistemas HPC europeos involucran todos los participantes:
  - Proveedores de servicios HPC en cada nivel
  - Infraestructuras Grid
  - Comunidades de usuarios industriales y científicos
  - Industria europea de HPC hardware y software



# PRACE: La unión de Europa

PARTNERSHIP  
FOR ADVANCED COMPUTING  
IN EUROPE



## Tier 1 – Miembros generales



## Tier 0 – Miembros







# JUGENE IBM BlueGene/P (FZJ Jülich, Alemania)

## Características generales:

- Potencia máxima de cálculo 1 Pflops
- 144 TB de memoria principal total
- 72 racks con 73.728 nodos
- 294.912 cores Power PC 450, 32-bits, 850 MHz
- 2 GB de memoria principal por core
- 5 PB capacidad total de almacenamiento

## 4 Redes:

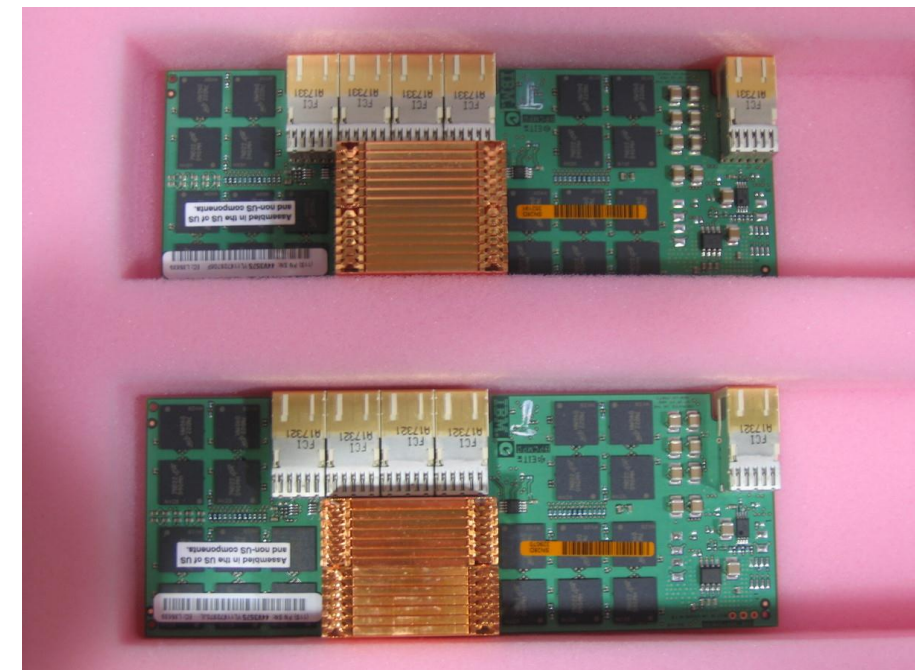
- Red propietaria
- 10 GigaEthernet

## Sistema operativo:

- SuSE Linux Enterprise (SLES 10)



*Sala del computador Jugene*



*Nodos de Jugene*

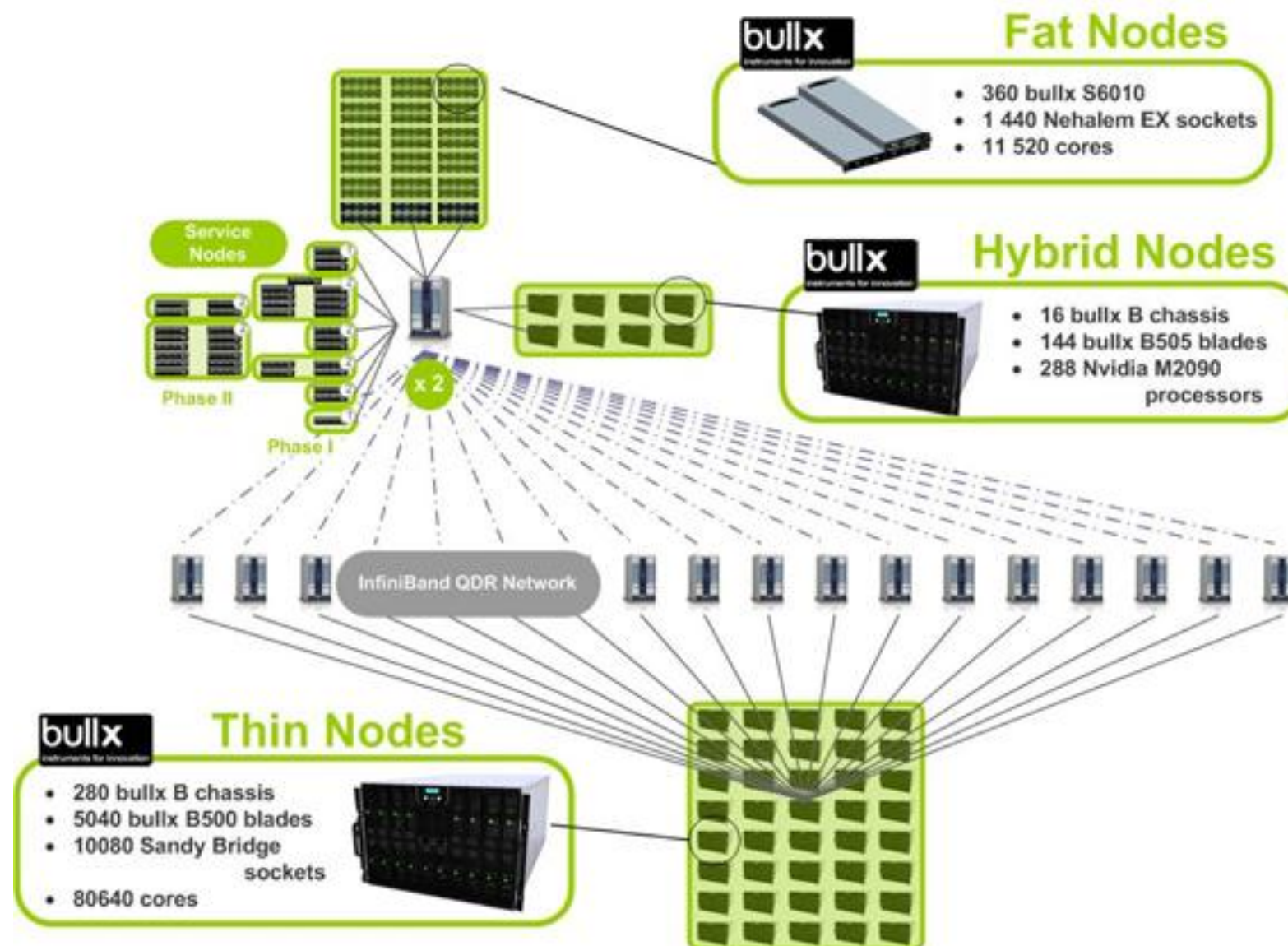
Más información en: <http://www2.fz-juelich.de/jsc/jugene>



# CURIE Bull XCluster (CEA Bruyères-Le-Châtel, Francia) (I)

Sistema a construir en 2 fases:

- Enero de 2011: 360 “Fat nodes” que proporcionan 105 Tflops
- En 2012: 5040 “Thin nodes” que proporcionarán 1,5 Pflops



Más información en: <http://www-hpc.cea.fr/en/complexes/tgcc-curie.htm>



## CURIE Bull XCluster (CEA Bruyères-Le-Châtel, Francia) (II)

### Características actuales:

- Potencia máxima 105 teraflops pico
- 360 S6010 bullx nodes (fat nodes)
- 4 eight-core x86-64 CPUs, 128 GB memoria principal y 1 disco local de 2TB.
- 45 TB de memoria principal total
- 720 TB de capacidad total de almacenamiento

### Red:

- InfiniBand QDR Full Fat Tree network

### Sistema operativo:

- Linux



*Sala del computador CURIE*



*Fat Nodes de CURIE*



## HERMIT CRAY XE6 (HLRS Stuttgart, Alemania)

Sistema a construir en 3 fases.

Características generales de la fase de test:

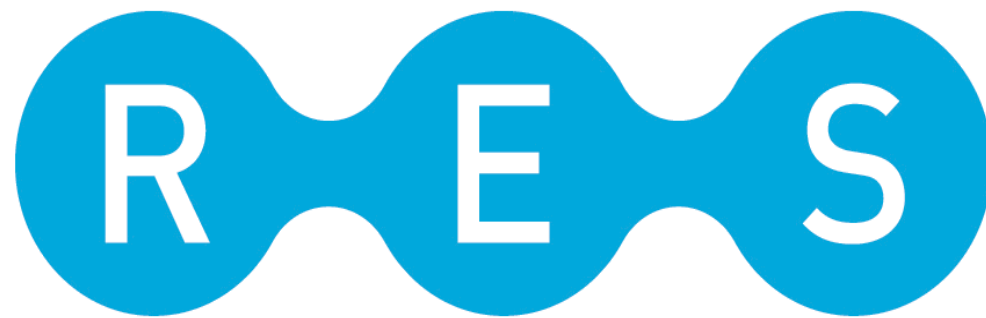
- Potencia máxima de cálculo 10 Tflops
- 1 rack de 84 nodos AMD Opteron 8-Core
- 13.440 AMPD Opteron, 2 GHz
- 32 GB de memoria/nodo
- 5 PB capacidad total de almacenamiento

Red:

- Cray Gemini



*Prototipo de Hermit*



RED ESPAÑOLA DE  
SUPERCOMPUTACIÓN



**Muchas gracias por vuestra atención**

<http://www.bsc.es>

Puedes visitar MareNostrum a través de [http://www.bsc.es/plantillaC.php?cat\\_id=37](http://www.bsc.es/plantillaC.php?cat_id=37)