

El Supercomputador MareNostrum

Microprocesadores para Comunicaciones

Omar El-Asmar Moreno

Indice

1. Introducción.....	Pag 2
2. Centro Nacional de Supercomputación de Barcelona.....	Pags 2-4
3. El supercomputador MareNostrum.....	Pags 4-12
3.1 MareNostrum I.....	Pags 4-6
3.2 MareNostrum II.....	Pags 6-8
3.3 MareNostrum III.....	Pags 8-12
4. Aplicaciones.....	Pags 14-15
5. Noticias de uso.....	Pag 15

1. Introducción

El supercomputador MareNostrum está situado en el Centro Nacional de Supercomputación de Barcelona (BSC-CNS) y actualmente es el supercomputador más potente de España y uno de los más potentes del mundo. Así lo ratifica el puesto 29 que ocupaba en junio de 2013 en la lista TOP500 que recoge los superordenadores más potentes del mundo. Su nombre ('Mar Nuestro' en latín) fue escogido tanto para representar el lugar donde se encuentra como para representar los grandes recursos informáticos de los que dispone.

Su construcción nace a partir del BSC-CNS, que es quien se encarga de su manejo y mantenimiento, de manera que a lo largo de los últimos años ha sufrido una serie de modificaciones que han mejorado sus prestaciones y capacidad de cálculo de manera impresionante. Actualmente forma parte del proyecto europeo PRACE cuya misión es permitir los descubrimientos científicos de alto impacto y la investigación en ingeniería, a la vez que mejora de la eficiencia energética de los sistemas de computación y la reducción de su impacto ambiental.

En este trabajo analizaremos cómo surgió la idea para construirlo, veremos las diferentes actualizaciones que ha sufrido y explicaremos las aplicaciones que tiene en la actualidad.

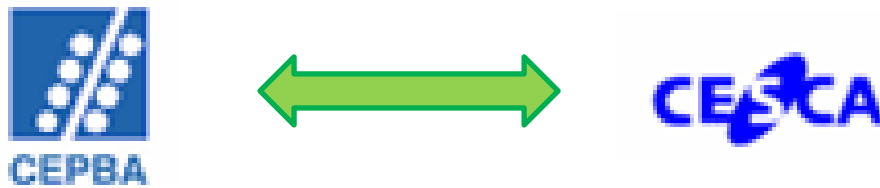
2. Centro Nacional de Supercomputación de Barcelona (BSC-CNS)

El BSC-CNS ha recorrido un largo camino durante estas dos últimas décadas para convertirse en lo que es actualmente. Básicamente, ha heredado la tradición del Centro Europeo Paralelismo de Barcelona (CEPBA), un centro de investigación, desarrollo e innovación en tecnologías de cómputo eficientes tanto para el mundo académico y la industria. Esta institución, perteneciente a la Universidad Politécnica de Cataluña (UPC), estaba patrocinada por la CICYT española (Comisión Interministerial de Ciencia y Tecnología) y el CIRIT catalán (Consell Interdepartamental de Recerca i Innovació Tecnològica).

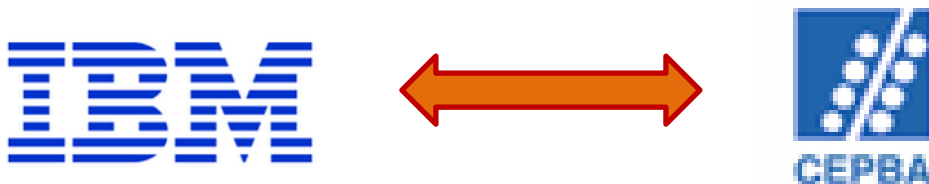
En 1991 CEPBA inició sus actividades, recopilando experiencia y las necesidades de los diferentes departamentos de la UPC. El Departamento de

Arquitectura de Computadores (DAC) se encargó de proporcionar la experiencia en el nivel más bajo del sistema informático (núcleos numéricos, sistemas operativos y arquitectura). Otros cinco departamentos de la UPC con alta demanda de cómputo se unieron para configurar el CEPBA .

Del año 1995 al 2000 el CEPBA coordinó sus actividades con el CESCA (Centre de Serveis Científics i Acadèmics De Catalunya) a través del C4 (Centro de Computación y Comunicaciones de Cataluña), fundado por el CIRIT, la Fundación Catalana de Investigación y la UPC.



En el año 2000 CEPBA firmó un acuerdo con IBM para crear el Instituto de Investigación CEPBA-IBM. Los objetivos de este acuerdo iban encaminados a la investigación sobre temas relacionados con Deep Computing, y el apoyo a la investigación local en otras áreas de la ciencia y la ingeniería. Esta Asociación de Investigación y Desarrollo tenía un compromiso inicial para 4 años.



En el año 2004 el Ministerio de Educación, la Generalitat de Catalunya y la Universidad Politécnica de Cataluña (UPC) tomaron la iniciativa de crear un Centro de Supercomputación en Barcelona, llamado Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC-CNS). En marzo de ese mismo año el Gobierno de España y la empresa IBM firmaron un acuerdo para construir unos de los ordenadores más potentes de Europa.

La principal misión del BSC-CNS consiste en investigar y desarrollar todo tipo de información tecnológica para facilitar el progreso científico en general. En este ámbito, se tiene una dedicación especial en las áreas de Ciencias de la Tierra, Ciencias de la vida y aplicaciones computacionales para la ciencia y la ingeniería.

En el año 2005 se constituyó oficialmente el BSC-CNS comenzando sus actividades de investigación y desarrollo. En ese mismo año, el supercomputador MareNostrum fue construido como la máquina más potente de Europa.

3. El supercomputador MareNostrum

Desde su inauguración en el año 2005, el supercomputador MareNostrum ha sufrido dos modificaciones que han mejorado sus prestaciones de memoria y capacidad de cálculo enormemente. A continuación, analizaremos las prestaciones de las tres versiones y veremos las modificaciones que se han ido produciendo.

3.1 MareNostrum I

El MareNostrum I fue construido en dos meses en el Centro Técnico de IBM en San Fernando de Henares (Madrid). Tras realizar las pruebas pertinentes y ejecutar el test LINPACK para su inclusión en la lista TOP500, fue trasladado al BSC-CNS donde se encuentra en la actualidad.



FIGURA 1. Interior de la capilla donde se encuentra el MareNostrum

El ordenador está físicamente instalado en el interior de la Capilla de Torre Girona, una edificación románica de principios del siglo XX que se encuentra en el Campus Nord de la Universidad Politécnica de Catalunya. Se encuentra en el interior de un cubo de cristal de 9 x 18 x 5 metros construido con más de 19 toneladas de cristal y 26 de hierro. El supercomputador ocupa una instalación de 170 metros cuadrados y pesa 40 000 kg. En la Figura 1 podemos ver de manera más clara su ubicación.

Para hacernos una idea del gran potencial del que disponía, vamos a ver un resumen de las características del sistema:

- a) Estaba formado por 2282 nodos de computación JS20 de IBM. Dentro de cada nodo había dos procesadores mononúcleo PowerPC 970FX, de arquitectura RISC de 64 bits a 1.6 GHz o 2.2 GHz, compartiendo 4 Gb de memoria.

De esta forma, se tenían 4812 procesadores y un total de 9128 Gb de memoria principal. A continuación, podemos ver el interior de un nodo JS20:



- b) Cada nodo individual JS20 disponía también 40 GB de almacenamiento además de un almacenamiento en red de 128 TB. Estos nodos estaban repartidos en 163 armarios, llamados BladeCenters, teniendo cada armario 14 nodos. En la imagen siguiente podemos ver el aspecto de un BladeCenter de IBM:



- c) Cuando se puso en marcha de forma definitiva se midió su rendimiento obteniendo un total de 27'91 GigaFlops, desde los anteriores 20'53 GigaFlops obtenidos en pruebas durante su construcción.

Como se puede comprobar, las características que presentaba la primera versión de MareNostrum eran asombrosas, no obstante, el BCS-CNS no quería que el proyecto se quedara ahí y apenas un año después se presentó una segunda versión del supercomputador, que mejoraría con creces las prestaciones que ya tenía.

3.2 MareNostrum II

Debido a la amplia demanda de uso que tenía el MareNostrum por parte de numerosos proyectos científicos, en el año 2006 el BSC-CSN decidió sacar una segunda versión del supercomputador, doblando su capacidad de cálculo.

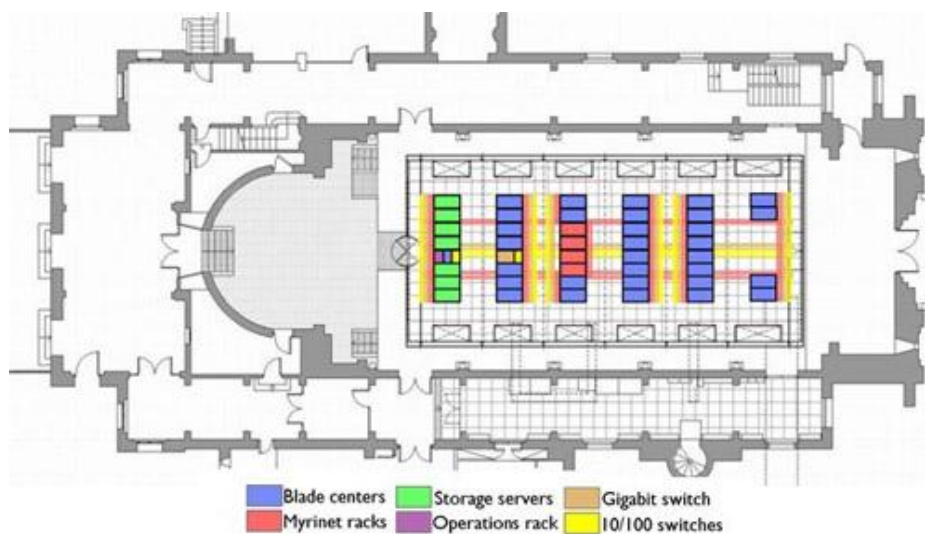


FIGURA 2. Distribución de los racks en el MareNostrum II

El nuevo sistema estaba formado por un total de 44 racks de los cuales 31 eran dedicados a capacidad de cálculo y los demás formaban la red de interconexión y control del sistema. En la figura 2 podemos ver la forma en que se distribuyen los racks.

Para poder analizar las modificaciones sufridas respecto a la versión anterior, vamos a ver un resumen de las características del sistema:

- a) En total estaba formado por 2560 nodos JS21 de IBM con procesadores de doble núcleo PowerPC 970MP de 64 bits funcionando a una velocidad de 2'3 GHz. Cada nodo dispone de 2 procesadores (4 núcleos), 8 GB de memoria principal y 36 GB de almacenamiento local.
- b) Los nodos JS21 están enfundados en armarios BladeCenter en grupos de 14, y a su vez estos BladeCenter están enfundados en los racks en grupos de 6. Cada rack tenía un consumo de 23 KW.

En total por rack suman hasta 336 núcleos, juntando así entre los 31 racks dedicados a operaciones de cálculo los 10.024 procesadores que conforman el sistema. Esto nos da en total una capacidad de cálculo de 63'83 TeraFlops pudiendo alcanzar picos de 94'21 TeraFlops con un consumo de 700 KW.

- c) En cuanto al sistema de memoria del ordenador, cuenta con 20 TB de memoria principal y 390 + 90 terabytes de almacenamiento en red, que se encuentran en 7 racks del sistema.



FIGURA 3. Distribución de las conexiones de la red Myrinet

- d) Los 2560 nodos se comunican entre sí a través de una red de interconexión de alta velocidad llamada Myrinet, que ocupa 4 racks dentro del sistema.

La red Myrinet consiste en dos cables de fibra óptica, conectados mediante un solo conector, de manera que la interconexión se realiza mediante conmutadores y encaminadores. En total está formada por más de 63 kilómetros de fibra óptica. En la figura 3 podemos ver como se realiza la interconexión.

- e) Para acceder al almacenamiento en red se utiliza una red Gigabit Ethernet, que se encuentra en uno de los racks del sistema y emplea 7,5 Km de fibra óptica. Está formada por un switch Force10 E600 Gigabit Ethernet, ofreciendo una cantidad total de 288 puertos de conexión.

Esta red incluye GPFS (Global Parallel File System), un software de IBM que permite que cualquier usuario desde cualquiera de los nodos tenga acceso a los datos, pueda modificarlos y eso no cree conflictos con otros usuarios que estén accediendo a los mismos datos.

- f) El sistema utiliza el entorno SUSE Linux, una de las distribuciones de Linux existentes a nivel mundial más sencillas de instalar y administrar, ya que cuenta con varios asistentes gráficos que facilitan el desarrollo de las diferentes tareas.

Tras realizar la nueva actualización del sistema, en el año 2007 se renovó el acuerdo entre el BSC-CNS e IBM. En este acuerdo se tipificaba el desarrollo y la compra de un nuevo supercomputador, a cambio del cual IBM financiaría diversos proyectos de investigación del BSC para mejorar las arquitecturas del procesador Cell para poder ser usado en la computación de altas prestaciones.

La idea era actualizar el supercomputador a su tercera versión lo antes posible, pero numerosos problemas a los que se añadió el inicio de la crisis económica española, hicieron que se retrasara varios años.

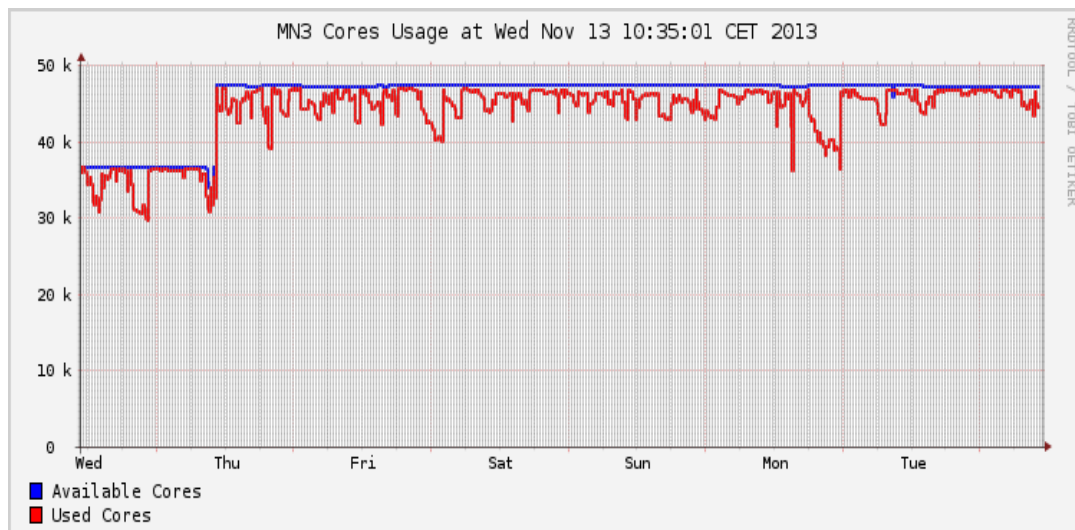
3.3 MareNostrum III

Tras varios años funcionando con la segunda versión, finalmente en el año 2012 llegó la tercera actualización del sistema, MareNostrum III, que se encuentra aún vigente. Esta nueva versión es 10 veces más potente que su predecesora y tiene una complejidad mucho mayor estando formada por 52 racks, de los cuales 36 están dedicados para la realización de cálculos.

El sistema se ha renovado por completo y ahora está basado en procesadores Intel SandyBridge, racks de computación iDataPlex y una red de interconexión Infiniband. A continuación podemos ver las características que presenta la nueva versión:

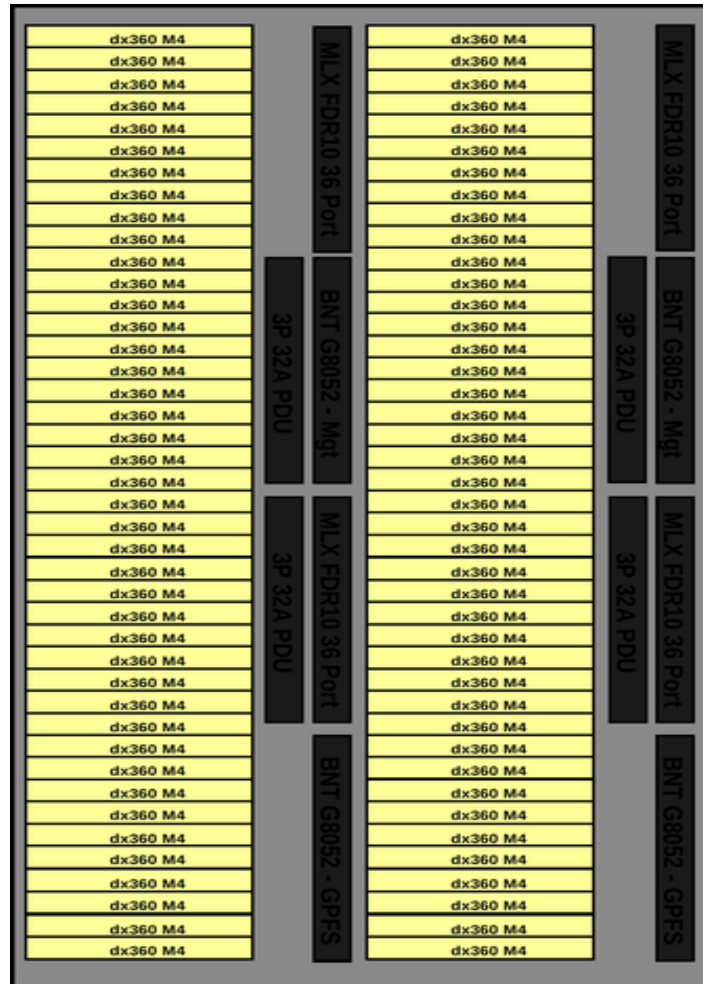
- a) El sistema está formado por 3056 nodos, cada uno de los cuales contiene 2 procesadores Intel SandyBridge (cada uno con 8 núcleos) a 2,6 GHz.

De esta manera tenemos en total 48896 procesadores trabajando en conjunto, lo que nos da una capacidad de cálculo de 1,1 PetaFlops por segundo con un consumo total de 1 MW. A continuación podemos ver una gráfica que nos muestra el porcentaje de uso de procesadores:

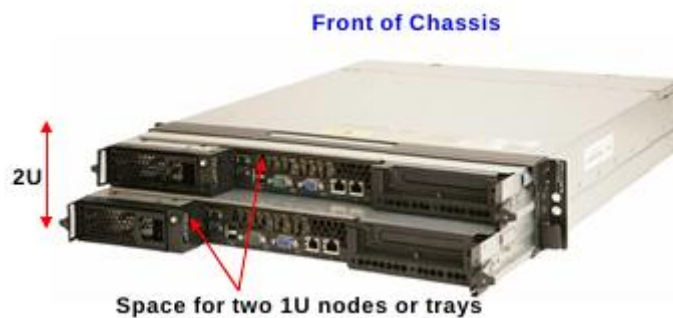


Todos estos procesadores están almacenados en 36 racks de cálculo IBM iDataPlex, de manera que tenemos una capacidad de cómputo de 27,95 TeraFlops por rack, con un consumo de 28 KW por rack. Si nos damos cuenta, este consumo es mayor que en el caso del MareNostrum 2, lo que implica un cambio en los sistemas de refrigeración del sistema.

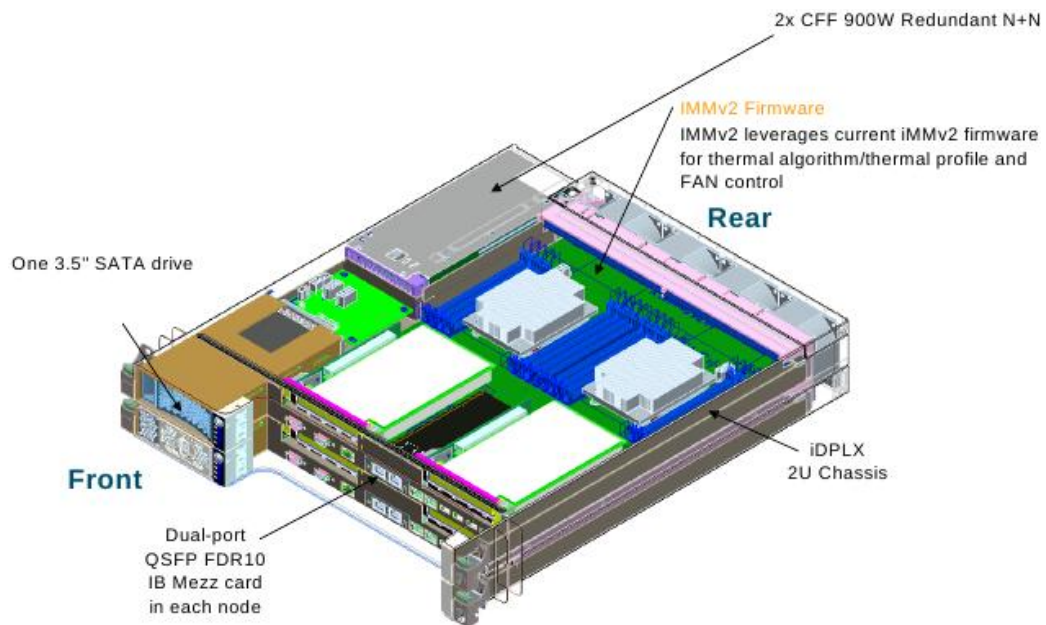
Por otro lado, cada uno de los racks iDataPlex está formado por 84 nodos de cómputo IBM dx360 M4, tal y como podemos ver en la siguiente imagen:



Cada nodo dx360 M4 está formado por una estructura 2U, de manera que cada uno de ellos incluye dos nodos de computación. En la siguiente imagen podemos ver como es la estructura externa:



La estructura interna de cada uno de los nodos la podemos ver en la imagen siguiente. En la parte izquierda de la imagen podemos ver los dos procesadores Intel que conforman cada uno de los nodos.



- b) En cuanto al almacenamiento tenemos 100,8 TeraBytes de memoria principal y aproximadamente 2 PetaBytes de almacenamiento en disco repartidos en 30 servers.
- c) Con las nuevas redes de comunicación se ha multiplicado por 40 la velocidad de comunicación. Cada nodo está conectado a 4 redes, dos de gestión y dos de datos:
- En cuanto a las redes de gestión una de ellas nos permite encender y apagar los nodos, mientras que la otra permite que los nodos tomen la imagen del sistema operativo.
 - En cuanto a las redes de datos la primera de ellas es una red Infiniband, que es la tecnología actual en las mejores máquinas del mundo, no bloqueante y de baja latencia que permite conectar todos los nodos a una velocidad de 40 Gbps empleando aproximadamente 71 km de fibra óptica.

La segunda de ellas, es la que implementa GPFS, que como ya vimos en la versión anterior nos permite mover datos hacia el sistema de almacenamiento de manera paralela sin molestar a otros usuarios.

- d) El sistema operativo que se ha montado sobre esta computadora es el SUSE Linux versión 10.

Como podemos observar, en la tercera versión del supercomputador se ha realizado una remodelación total respecto a las versiones anteriores. Esto ha permitido, no solo mejorar las prestaciones de cómputo del supercomputador, sino también emplear las tecnologías más innovadoras en lo que se refiere a unidades de procesamiento y redes de interconexión.

Todas estas mejoras han supuesto un salto en la calidad de las aplicaciones que se pueden realizar en el supercomputador. Algunas de las más importantes las veremos en el siguiente apartado.

4. Aplicaciones

Como ya hemos podido comprobar, las capacidades de computación que ofrece el MareNostrum son muy buenas, por ello, sus servicios se prestan para investigaciones científicas que requieren cálculos tremendamente complejos. Su uso está disponible para la comunidad científica nacional e internacional, controlada por un comité de asignación que asigna tiempo de cómputo en función de la valía de los proyectos a realizar.

Así las diferentes versiones del MareNostrum han sido esenciales para desarrollar importantes proyectos de investigación tanto públicos como privados. Algunos de los más emblemáticos son:

- ❖ **METEOROLOGÍA Y CLIMA.** El departamento de ciencias de la Tierra del BSC-CNS utiliza la supercomputación para mejorar la predicción meteorológica, los estudios de calidad del aire y las investigaciones sobre el clima.

- ❖ **CUERPO HUMANO.** El departamento de ciencias de la vida puede estudiar, gracias al MareNostrum, cómo actúan las proteínas en el interior de las células y desarrollar proyectos de genómica.
- ❖ **ENERGÍA EÓLICA.** El BSC y la compañía Iberdrola tienen una colaboración científica para mejorar el diseño de los molinos de viento a partir de simulaciones informáticas de la interacción entre los flujos de aire y las turbinas.
- ❖ **PETRÓLEO.** Repsol y el BSC han creado un centro de investigación conjunto para utilizar la supercomputación con el objetivo de mejorar la detección de depósitos de hidrocarburos y su explotación.
- ❖ **SUPERCOMPUTACIÓN.** Algunas de las compañías líderes en el sector de los superordenadores, como IBM, Intel y Nvidia, tienen acuerdos de colaboración estable con el BSC para mejorar sus productos.
- ❖ **INFORMÁTICA PERSONAL.** Microsoft y el BSC han creado un centro de investigación conjunto para mejorar la computación en ordenadores personales y dispositivos móviles.

A continuación podemos ver un resumen del uso del MareNostrum en un año por parte de diferentes fuentes, así como las diferentes aplicaciones que se encuentran en el:

- ❖ Aplicaciones instaladas permanentemente en el MareNostrum → 259
- ❖ Usuarios del MareNostrum en 1 año → 650

DISTRIBUCIÓN DEL TIEMPO DE CÁLCULO	
Proyectos Europeos PRACE	70 %
Proyectos de la Red Española de Supercomputación	24 %
Proyectos propios del BSC-CNS	6 %

5. Noticias de uso

Un corazón en 3D : el supercomputador MareNostrum clave para simular el órgano

El Barcelona Supercomputing Center-Centro Nacional de Supercomputación trabaja actualmente en una simulación 3D del corazón. Este proyecto se enmarca en el programa de excelencia Severo Ochoa y el supercomputador MareNostrum es clave. En el proyecto intervienen médicos del hospital de Sant Pau y de la Universidad de Lleida y científicos del Centro de Visión por Computador de la Universidad Autónoma de Barcelona, que proporciona las imágenes que ayudan a reproducir el funcionamiento del corazón.

El superordenador MareNostrum simulará las neuronas en el Human Brain Project

El Human Brain Project (proyecto del cerebro humano, HBP en sus siglas en inglés) está considerado el proyecto neurocientífico más ambicioso del mundo con una inversión de 1.200 millones de euros.

Según han informado los responsables del BSC, el reto de este macroproyecto cofinanciado por la Unión Europea es el de comprender y simular el funcionamiento del cerebro humano en sus diferentes capas, desde el genoma y niveles celulares a neuronas, circuitos, regiones del cerebro y finalmente el cerebro entero.

Según los científicos, tener una comprensión profunda de cómo opera el cerebro humano será clave para el desarrollo de nuevas herramientas para tratar las enfermedades neurológicas y crear nuevas tecnologías de la información.

6. Bibliografía

❖ Información general sobre el MareNostrum

- <http://es.wikipedia.org/wiki/MareNostrum>
- <http://www.bsc.es/marenostrum-support-services/documentation>

❖ Noticias

- <http://www.lavanguardia.com/tecnologia/20121027/54353933541/mare-nostrum-3-superordenador-barcelona.html>
- <http://www.rtve.es/noticias/20131007/superordenador-marenostrum-simulara-neuronas-proyecto-human-brain/759744.shtml>
- <http://www.noticias24.com/tecnologia/noticia/568/marenostrum-la-supercomputadora/>

❖ Vídeos

- <http://shelf3d.com/lwFccNUwWpQ#Corporate Video in English>
- <http://www.youtube.com/watch?v=1qOYIZwVUgc>
- <http://www.youtube.com/watch?v=wTVuhQkUb6Y>
- <http://www.youtube.com/watch?v=D8vGWoKtYck>