



DESARROLLO DE SOFTWARE EN ARQUITECTURAS PARALELAS

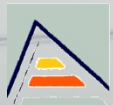


Dpnt. de Ciència de la Computació i Intel·ligència *a*rtificial
Dpto. de Ciencia de la Computación e Inteligencia *a*rtificial

Jose Penadés



Grupo de Computación de
Altas Prestaciones y Paralelismo



Universitat d'Alacant
Universidad de Alicante

DESARROLLO DE SOFTWARE EN ARQUITECTURAS PARALELAS

1. Motivación y aspectos de la programación paralela.
2. Tipos de sistemas paralelos. Paradigmas de programación paralela.
3. Conceptos básicos y medidas de paralelismo.
4. Diseño de programas paralelos.
5. La interface de paso de mensaje: el estándar MPI.
6. Paralelización de algoritmos: ejemplos y aplicaciones.

DESARROLLO DE SOFTWARE EN ARQUITECTURAS PARALELAS

1. Motivación y aspectos de la programación paralela.
 - Necesidades de la computación en paralelo.
 - Aplicaciones.
 - ¿Dónde se realiza cómputo en paralelo?
2. Tipos de sistemas paralelos. Paradigmas de programación paralela.
3. Conceptos básicos y medidas de paralelismo.
4. Diseño de programas paralelos.
5. La interface de paso de mensaje: el estándar MPI.
6. Paralelización de algoritmos: ejemplos y aplicaciones.



Necesidades de la computación en paralelo

❑ DEMANDA:

- Computadores con gran potencia de calculo.
- Velocidad en la computación.
- Problemas con complejidad elevada:
 - o Problemas de gran dimensión: coste polinomial de grado alto o aplicabilidad a grandes problemas.
 - o Problemas de tiempo real.
 - o Problemas de gran desafío: gran importancia social. Estudio del genoma humano, predicción meteorológica mundial, modelado fenómenos sísmicos, ...



Necesidades de la computación en paralelo

❑ LIMITACIONES de la computación secuencial:

➤ Restricciones lógicas:

- o **Complejidad intrínseca de los algoritmos.**

- o **Problemas de gran dimensión.**

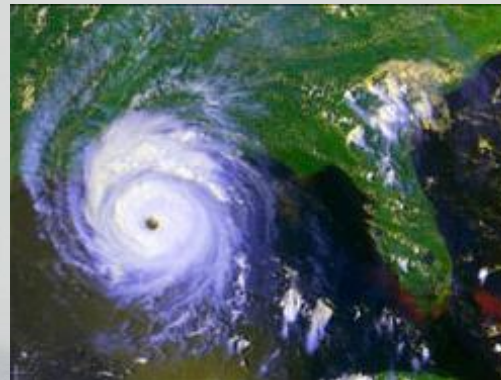
- Problemas que requieren PetaFLOPS (10^{15} FLOPS) y PetaBytes en recursos computacionales (ver por ejemplo “Grand Challenge” en.wikipedia.org/wiki/Grand_Challenge).

- Web search engines: se procesan millones de datos por segundo.

- o **Problemas en tiempo real.**



Formación de galaxias



Cambio global



Predicción climatológica



Necesidades de la computación en paralelo

❑ LIMITACIONES de la computación secuencial:

➤ Restricciones tecnológicas:

- o Velocidad de transmisión: La velocidad de un computador secuencial depende directamente de cómo los datos se mueven a través del hardware. Los límites absolutos son la velocidad de la luz (30 cm/nanoseg) y el límite de transmisión del cable de cobre (9cm/nanoseg). Un incremento de velocidad necesita de un incremento en la proximidad de los elementos de proceso.
- o Límite de integración: Cerca del límite máximo, aunque la ley de Moore (el número de transistores integrados por unidad de silicio se dobla cada 18 meses, 1965) se ha mantenido precisa durante un largo tiempo.
- o Estancamiento en los incrementos de la frecuencia de reloj (del orden de unos pocos nanosegundos). Más frecuencia → Más consumo + Temperaturas muy elevadas + interferencia electromagnética.

➤ Limitaciones económicas:

- o Es considerablemente muy caro hacer que un único procesador sea más rápido. Es mucho menos caro utilizar un número mayor de procesadores para conseguir el mismo propósito (o incluso mejorarlo).



Necesidades de la computación en paralelo

❑ SOLUCIONES:

- De tipo software: mejoras en los algoritmos.
- De tipo hardware: mejoras en la tecnología de computadores (mayor rapidez de instrucciones, dispositivos electrónicos más rápidos, ...)
- Paralelismo: replicar unidades de tratamiento de información con el objetivo de repartir tareas entre las mismas.

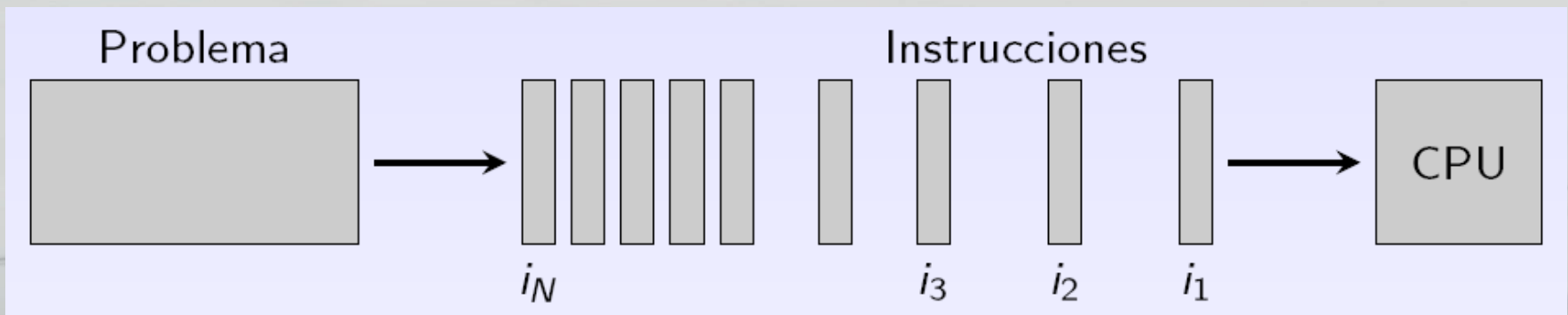




Necesidades de la computación en paralelo

Tradicionalmente los programas se han desarrollado para el cálculo en serie:

- ❑ Funcionan en un ordenador con una única CPU.
- ❑ Un problema se divide en un conjunto de instrucciones.
- ❑ Las instrucciones se ejecutan secuencialmente.
- ❑ Únicamente una instrucción se ejecuta cada vez.

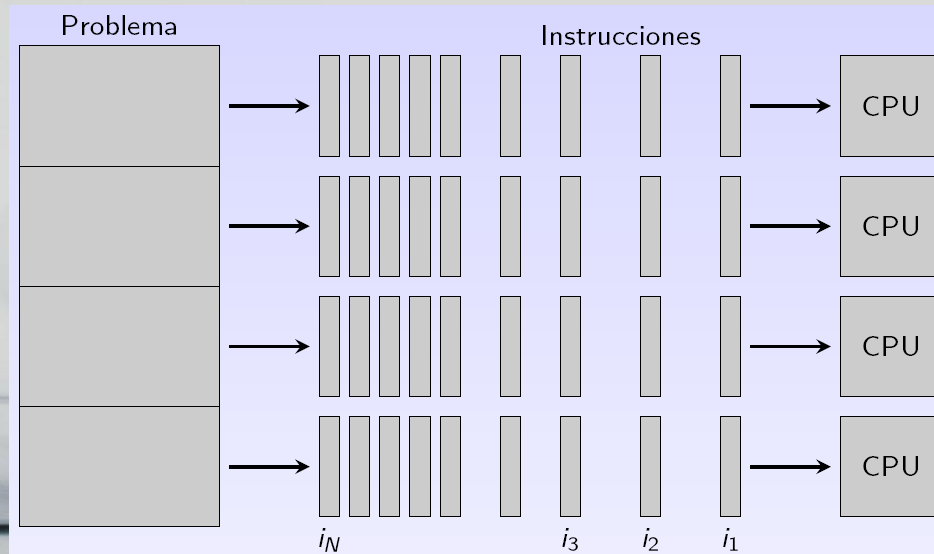




Necesidades de la computación en paralelo

El cálculo en paralelo consiste en usar múltiples recursos simultáneamente para resolver un problema dado:

- ❑ Hace uso de un ordenador con varias CPUs.
- ❑ El problema se divide en partes independientes.
- ❑ Cada parte se divide en un conjunto de instrucciones.
- ❑ Las instrucciones se ejecutan secuencialmente.
- ❑ Las partes se resuelven simultáneamente.





Necesidades de la computación en paralelo

□ Paralelismo:

- **Computador paralelo:** Capaz de ejecutar varias instrucciones simultáneamente.
- **Computación paralela:** Uso de varios procesadores trabajando juntos para resolver una tarea común:
 - o Cada procesador trabaja en una porción del problema.
 - o Los procesos pueden intercambiar datos, a través de las direcciones de memoria compartidas o mediante una red de interconexión.
- A las arquitecturas de este tipo se las denomina **Arquitecturas Paralelas.**





Necesidades de la computación en paralelo

Ejemplo: Ordenar un conjunto de libros

- ❑ Libros organizados en **estantes**.
- ❑ Estantes agrupados en **estanterías**.
- ❑ **Una persona** → Velocidad limitada.
- ❑ **Varias personas** → Enfoques:
 - Repartir libros entre trabajadores + ordenación simultánea.
 - Trabajadores deben desplazarse.
 - Repartir libros y estanterías.
 - Si un trabajador encuentra un libro suyo lo almacena.
 - Si no, lo pasa al responsable.



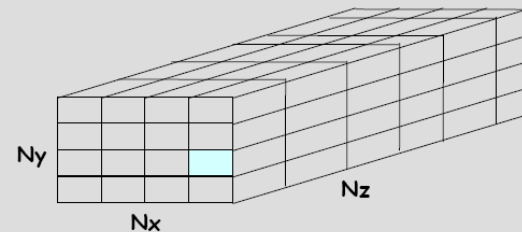
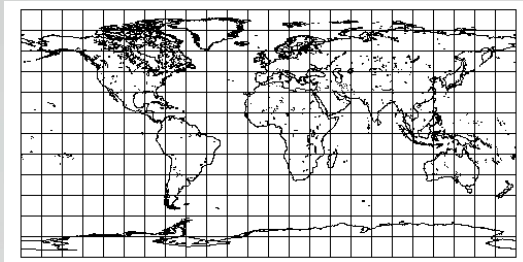
- ❑ **Aspectos de la solución paralela:**
 - **Descomposición:** La tarea se divide en subtareas.
 - **Asignación:** Las tareas se asignan a trabajadores.
 - **Comunicación:** Los trabajadores deben cooperar.
 - **Diferentes soluciones paralelas** de un problema.



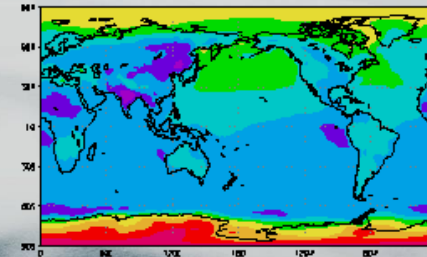
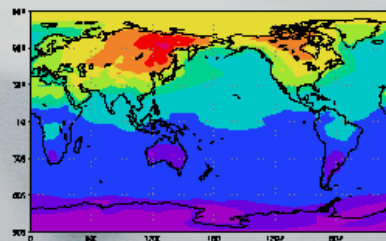
Aplicaciones

❑ Modelado predictivo y simulaciones. Ejemplo: modelado del clima:

- Mallado de $N_x = N_y = 3000$ Kms, $N_z = 11$ kms.
- Dominio descompuesto en segmentos cúbicos $0.1 \times 0.1 \times 0.1 \approx 10^{11}$ segmentos.



- Evolución temporal (2 días): recálculo segmento en paso de tiempo (30 min.) \rightarrow 100 operaciones.

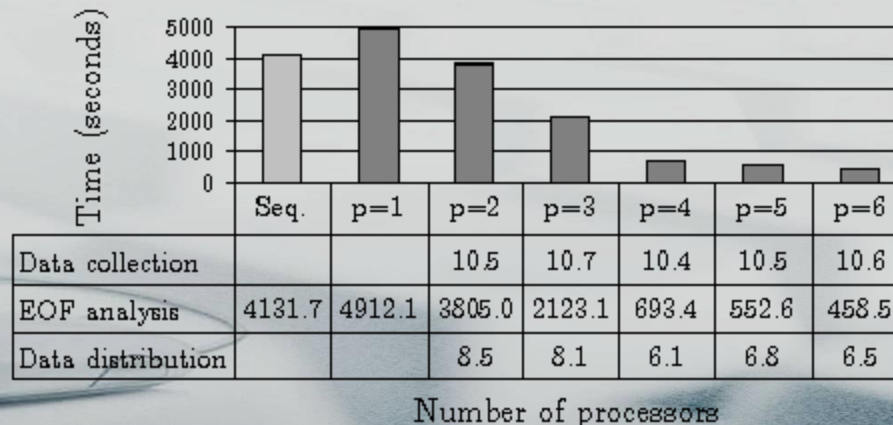




Aplicaciones

❑ Modelado predictivo y simulaciones. Ejemplo: modelado del clima:

- Ordenador secuencial 10^9 inst./seg $\rightarrow \approx 12$ días !
- Ordenador paralelo con 1000 procesadores: resolución en **menos de 3 horas**.
- Ventajas paralelismo:
 - o Resolución problemas antes “irresolubles”.
 - o Mallado más fino para incrementar la precisión.





Aplicaciones

❑ Técnicas de validación cruzada:

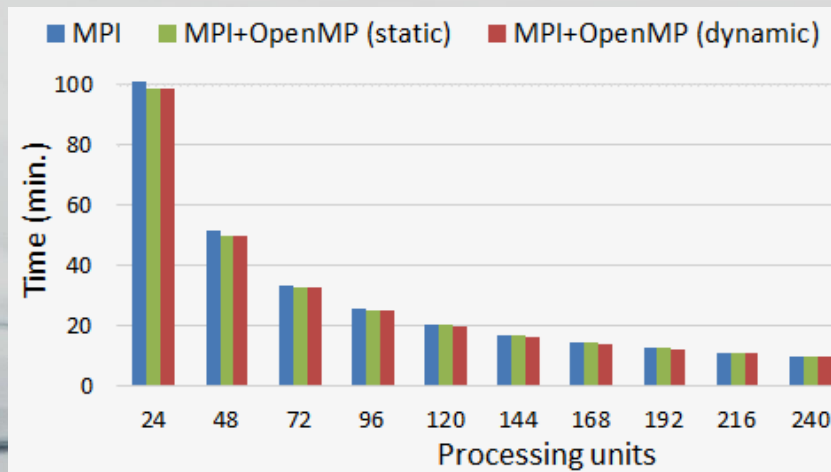
- Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar la precisión de un modelo que se llevará a la práctica.
- Divide los datos de muestra en dos conjuntos.
- Realiza el análisis de un subconjunto (datos de entrenamiento o training set), y
- valida el análisis en el otro subconjunto (datos de prueba o test set)



Aplicaciones

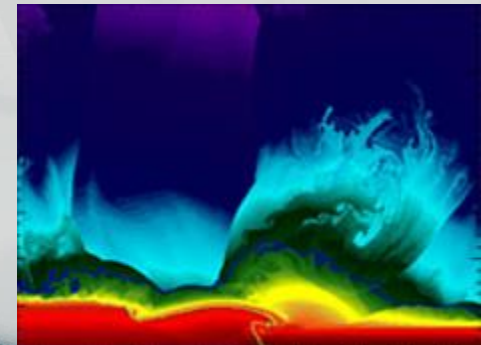
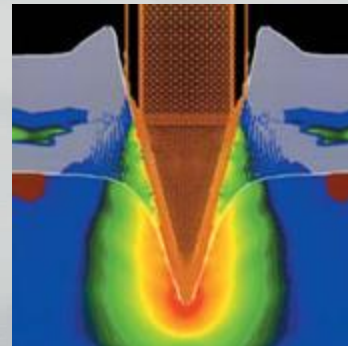
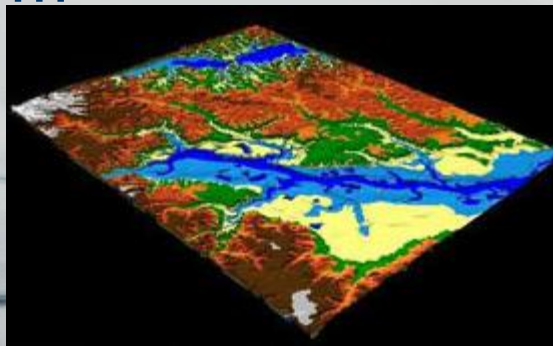
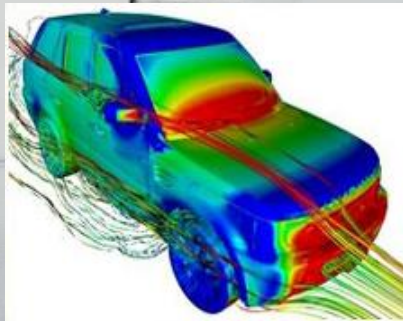
❑ Técnicas de validación cruzada. Ejemplo:

- Datos de la muestra: 27988 datos con 7 variables cada dato.
- Training set: el 80% de la muestra, 22391.
- Test set: el resto de los datos.
- Ordenador secuencial $\rightarrow \approx 37,7$ horas !
- Ordenador paralelo con 240 procesadores (cluster IUII): resolución en **menos de 10 minutos**.



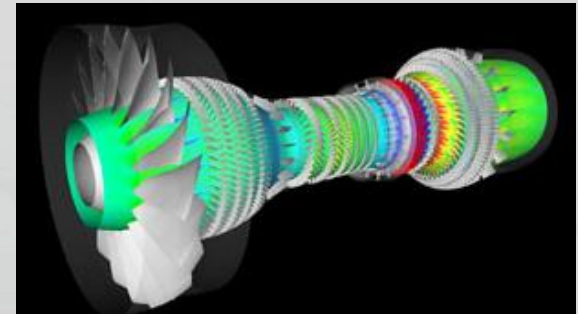
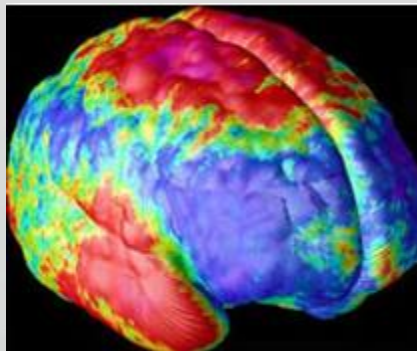
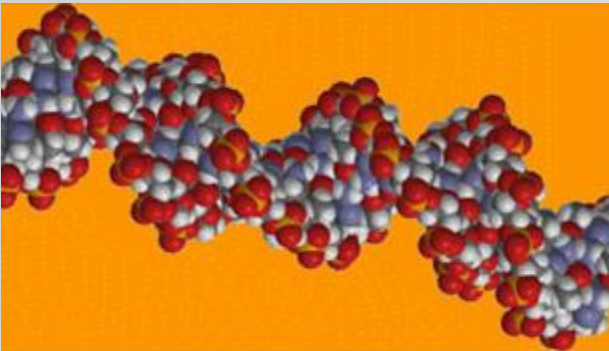
Aplicaciones

- ❑ Diseño y automatización de proyectos de ingeniería.
 - Aerodinámica computacional.
 - Inteligencia artificial y automatización:
 - o Procesado de imágenes, reconocimiento de patrones, comprensión del habla, deducción automática, robótica inteligente, sistemas expertos, ...
 - Aplicaciones de detección remota.
- ❑ Exploración de recursos energéticos: exploración sísmica, modelado de yacimientos, energía de fusión en plasmas, ...



Aplicaciones

- ❑ Investigación médica: tomografía asistida por computador, ingeniería genética, genoma humano, ...
- ❑ Investigación de armamentos y defensa.
- ❑ Problemas de investigación básica: química, mecánica cuántica, dinámica de fluidos, ...





Aplicaciones

(en medianas empresas)

- ❑ Uso de computación de altas prestaciones para el diagnostico temprano de fallos en turbinas de gas.
- ❑ Control de calidad de discos ópticos defectuosos.
 - Control de defectos visual realizado por especialistas.
 - Tiempo por disco: 3 minutos.
 - Control automático sobre un cluster de PC's usando PVM: control de calidad en menos de 1 minuto.





Aplicaciones

(en medianas empresas)

- ❑ Proyecto PARSAR: Procesamiento rápido y barato de imágenes SAR (Radar de apertura sintética).
 - Permite obtener imágenes terrestres desde satélites incluso a través de cielos nublados.
 - Problema: retraso entre la adquisición de datos y su procesamiento (≈ 1 mes).
 - o Complejidad de los algoritmos de procesado.
 - o Distancia entre las estaciones receptoras de los datos y los centros de procesamiento.
 - Solución: uso de paralelismo en las propias estaciones receptoras para minimizar tiempo de proceso.



Aplicaciones

(en medianas empresas)

- ❑ Proyecto PCECOWATER: Modelado eficiente de corrientes en entornos medioambientales.
 - Adaptación de código secuencial para su funcionamiento en redes de ordenadores.
- ❑ Proyecto HIPEROAD: Diseño aerodinámico de coches.
 - Diseño de un coche: especialista propone diseño inicial, construcción de uno o varios modelos a escala, prueba experimental, modificaciones aerodinámicas, construcción modelo a escala natural, prueba en tunel de viento.
 - En HIPEROAD se desarrolló un entorno de simulación para diseñar y optimizar automáticamente la geometría de un coche usando paralelismo.
 - Usado, por ejemplo, en el diseño preliminar de un modelo de coche deportivo similar al Ferrari 550 Maranello.



Aplicaciones

(en grandes empresas)



Investigación en tecnologías avanzadas para la exploración de hidrocarburos, modelización de reservas subterráneas y submarinas y flujos de fluidos.



Investigación sobre optimización de
parques eólicos y previsiones de
producción de energía con
aerogeneradores.





Aplicaciones

(en grandes empresas)



Desarrollo de sistemas avanzados de deep learning con aplicaciones a servicios bancarios.



SEAT

Simulaciones para mejorar la comprensión de la física giratoria de las ruedas y su impacto sobre el rendimiento aerodinámico.





Aplicaciones (en grandes empresas)



Métodos estadísticos avanzados para la optimización del mantenimiento, el uso de energía y el control de los procesos de tratamiento y suministro de agua de la ciudad.

ikergune

Investigación sobre detección de datos eficiente, algoritmos para análisis de procesos industriales y visualización de grandes conjuntos de datos industriales.





Técnicas de Inteligencia Artificial y Big Data para mejorar la calidad asistencial y el diagnóstico personalizado.



Pronóstico de tormentas de polvo para mejorar la seguridad de los vuelos.

Aplicaciones

(en grandes empresas)

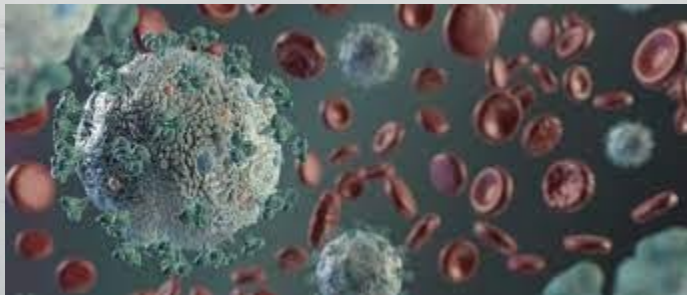


Aplicaciones

(más cercanas)

□ Comprensión del genoma viral:

- Aunque el genoma viral puede secuenciarse, comprender su patología invasiva en tiempo real es difícil debido a sus mutaciones. Pero gracias a la computación de alto rendimiento, las simulaciones innovadoras de estos mecanismos están evolucionando.
- Un ejemplo más reciente de una aplicación de computación de alto rendimiento es la simulación del genoma completo de COVID-19 en una supercomputadora a principios de 2020.



Aplicaciones

(más cercanas)

- ❑ El Centro de Supercomputación de Cataluña (BSC) utiliza la capacidad de cálculo del superordenador MareNostrum en la lucha contra el coronavirus:
 - Almacena y analiza datos clínicos de pacientes de COVID-19 para la creación de herramientas que asistan a los clínicos en el diagnóstico y tratamiento de la enfermedad.
 - Utiliza la capacidad de cómputo del supercomputador MareNostrum 4 (posición 63 en el Top500 de 2021) para analizar la propagación y el impacto social de la epidemia.





Aplicaciones

(más cercanas)

❑ Efectos especiales de Titanic (1997):

- 160 433MHz DEC Alpha systems (Carrera Computers of Newport Beach, California) con una conexión Ethernet a 100Mbps.





Aplicaciones

(más cercanas)

❑ Efectos especiales de Avatar (2009):

- 4.352 servidores HP Proliant BL2x220c G5 Blade con GNU/Linux.
- 40.000 procesadores (refrigerados por agua).
- 104 Terabytes de RAM.
- Para la película se procesaron 12 MB/frame \rightarrow 288 MB/s \rightarrow 17,3 GB/m
- Las 24 horas del día \rightarrow 1,3-1,4 millones de tareas \rightarrow Procesando 708 GB/s.

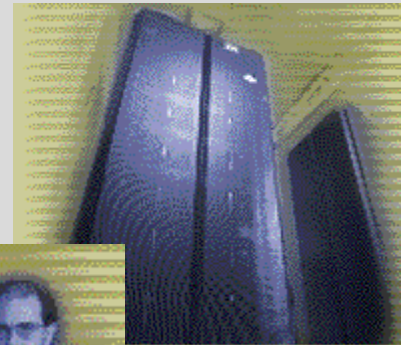




Aplicaciones

(más cercanas)

- ❑ Derrota de Kasparov frente a Deep Blue (1996-1997):
 - 2 frames IBM RS/6000 con 48 procesadores.





Aplicaciones

(más cercanas)

- ❑ Watson venció a los campeones de Jeopardy (2011):
 - Un prototipo Blue Gene/Q con 8192 cores.
 - Número 115 en el Top 500 (Noviembre 2010).
 - Más de 100 teraflops (10^{14} flops) de rendimiento pico.
 - 65 teraflops con LINPACK.



Watson, desarrollando la IA

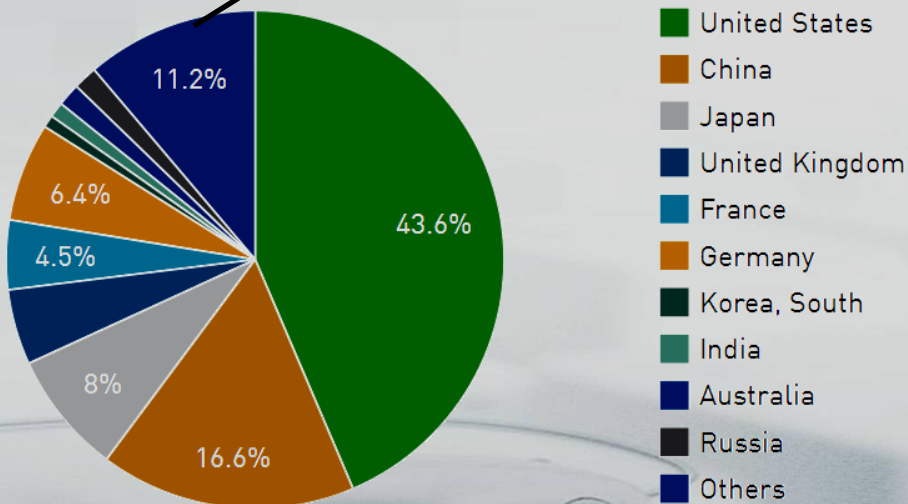
¿Dónde se realiza cómputo paralelo?

□ Fuente: www.top500.org

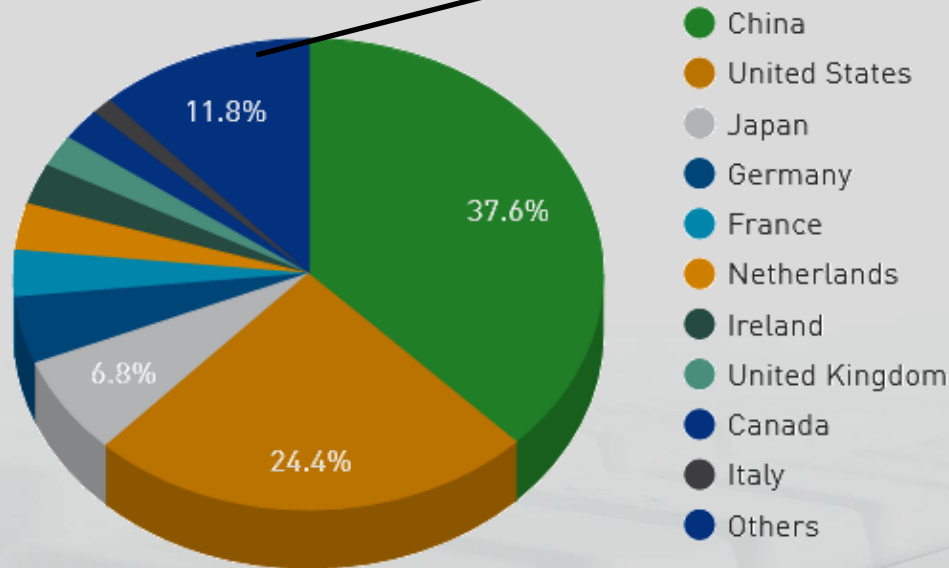
□ Países / rendimiento:

Noviembre 2014

España: 0,4%



España: 0,2%



Junio 2021

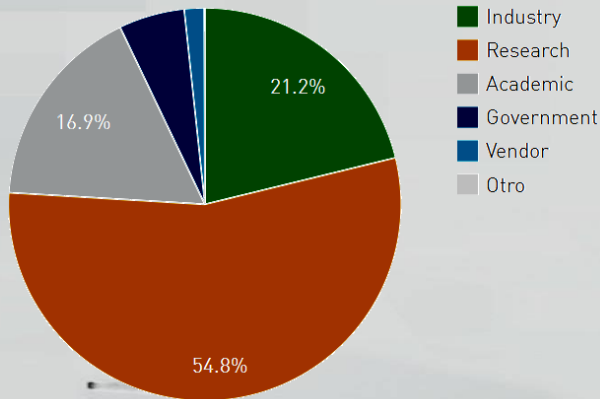
¿Dónde se realiza cómputo paralelo?



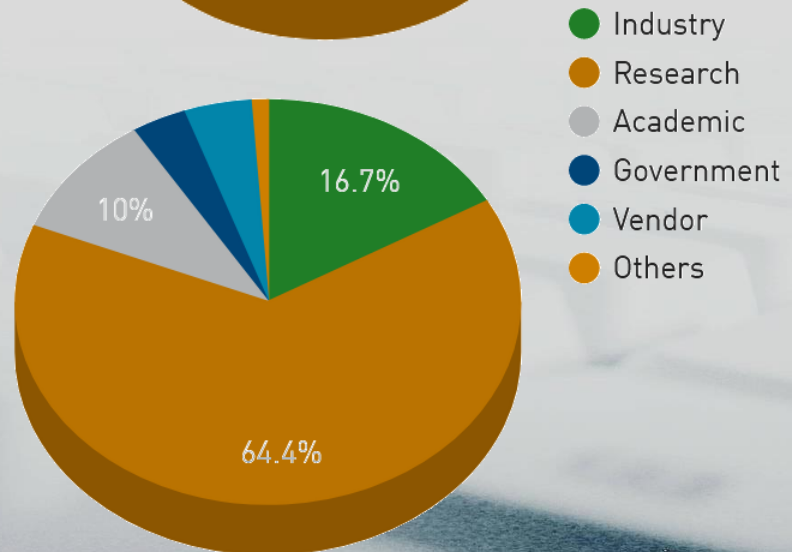
❑ Fuente: www.top500.org

❑ Segmentos / rendimiento:

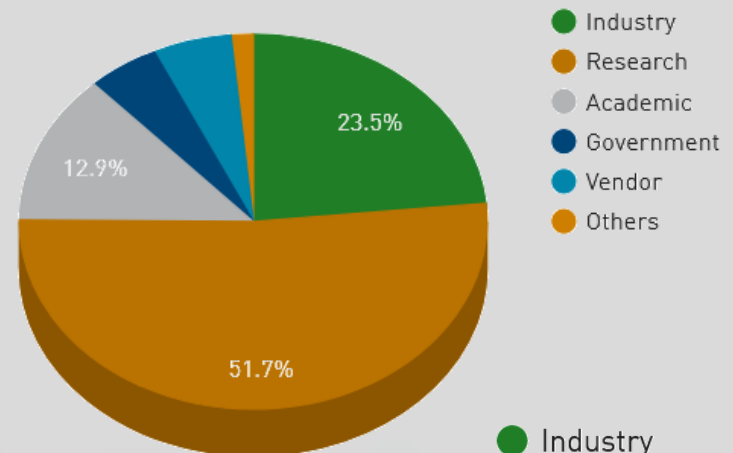
Noviembre 2014



Junio 2022



Junio 2021

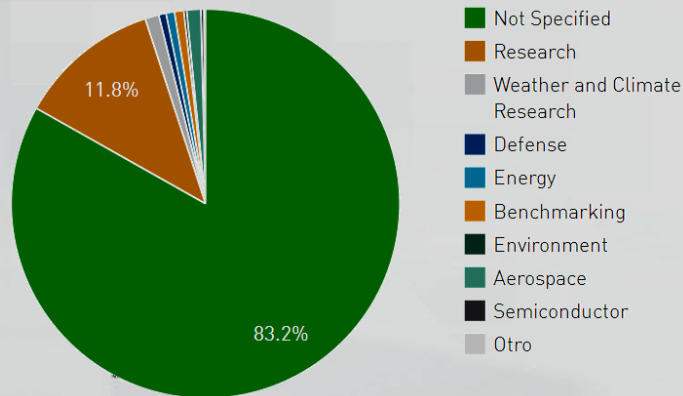


¿Dónde se realiza cómputo paralelo?

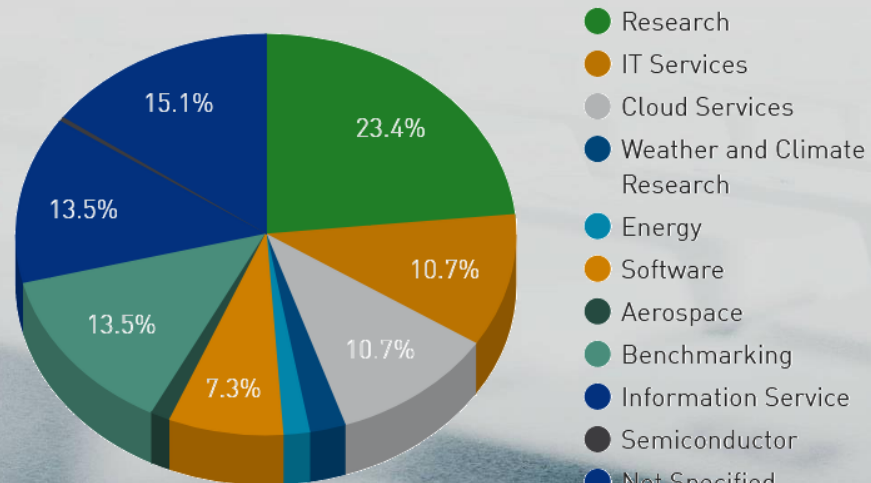
❑ Fuente: www.top500.org

❑ Áreas de aplicación / rendimiento:

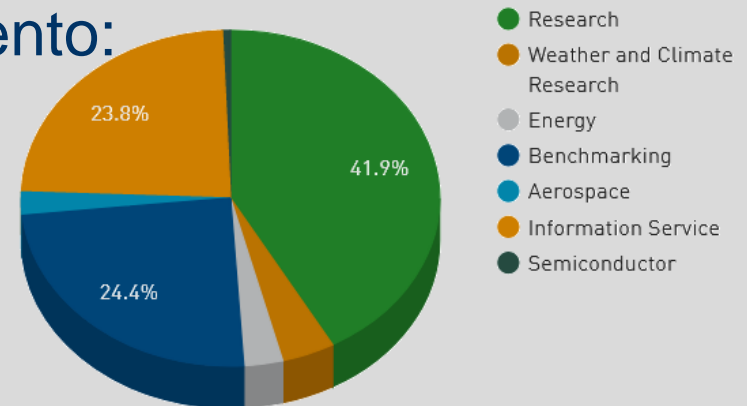
Noviembre 2014



Junio 2022



Junio 2021





Un ejemplo del TOP500

❑ Supercomputador Joliot-Curie

- Potencia computacional máxima de 22 petaflops¹.
- Es la supercomputadora más poderosa de Francia dedicada a la investigación abierta académica e industrial, y la tercera computadora de investigación más poderosa de Europa (según el TOP500 ranking publicado el 22 de junio de 2020).
- Amplia potencia informática con más de 440.000 núcleos x86 de alto rendimiento en casi 5000 nodos HPC.
- 18 proyectos relacionados con la COVID-19: proporcionando resultados utilizados para la lucha contra la COVID-19 y acelerar el desarrollo de un tratamiento basado en un conocimiento óptimo del virus.

1. petaflop = 1000 teraflops = 1000^2 gigaflops = 1.000.000.000.000.000 flops
flops: operaciones de coma flotante por segundo



TOP500 2022: los 5 primeros

Rango	Nombre	Ubicación	Cores	PetaFlops
1	Frontier	DOE/SC/Oak Ridge National Laboratory, Estados Unidos	8.730.112	1.685,65
2	Supercomputer Fugaku	RIKEN Center for Computational Science, Japón	7.630.848	537,21
3	Lumi	HPE EuroHPC/CSC Finland	1.110.144	214,35
4	Summit	DOE/SC/Oak Ridge National Laboratory, Estados Unidos	2.414.592	200,79
5	Sierra	DOE/NNSA/LLNL Estados Unidos	1.572.480	125,71