

 **Programación Paralela y Distribuida**
Introducción a la Programación HPC

 **Índice**


1. Introducción
2. HPC: Top500 y aplicaciones
3. Programación HPC

2

 **Introducción**

- Hay muchas aplicaciones científicas y de ingeniería que tienen tiempos de ejecución muy elevados
 - Dinámica molecular, química cuántica, estudio de nuevos materiales, **bioinformática**, defensa, finanzas, dinámica de fluidos, resistencia de materiales, electrónica, animación, **meteorología**, ...
- Experimentos virtuales vs físicos
- La paralelización es la única forma de alcanzar tiempos de ejecución razonables y poder abordar experimentos más complejos.
- Aplicaciones “insaciable”, escalan más rápido que el Hw

3

 **Introducción**

- HPC: High Performance Computing (Computación de Alto Rendimiento)
- Busca obtener tiempos de ejecución razonables para estas grandes aplicaciones, permitiendo de este modo realizar experimentos que de otro modo serían inviables, por tamaño, complejidad y/o precisión.
- Usa máquinas diseñadas específicamente para este propósito (Centros de Supercomputación: CeSViMa)
- La programación de estas máquinas tiene sus peculiaridades, tanto por el hardware, de cierta complejidad y diversidad, como por las aplicaciones, programadas por científicos e ingenieros


4

IBM High Performance Computing 

What Drives HPC? --- "The Need for Speed..."
Computational Needs of Technical, Scientific, Digital Media and Business Applications
Approach or Exceed the Petaflops Range

 <p>CFD Wing Simulation 512x64x256 Grid (8.3 x 10⁶ mesh points) 5000 FLOPs per mesh point, 5000 time steps/cycles 2.15 x 10¹⁴ FLOPs</p>	 <p>Materials Science Magnetic Materials: Current: 2000 atoms; 2.64 TF/s, 512GB Future: HDD Simulation – 30TF/s, 2 TBs Electronic Structures: Current: 300 atoms; 0.5 TF/s, 100GB Future: 3000 atoms; 50TF/s, 2TB</p>
 <p>CFD Full Plane Simulation 512x64x256 Grid (3.5 x 10¹⁷ mesh points) 5000 FLOPs per mesh point 5000 time steps/cycles 8.7x 10²⁴ FLOPs</p>	 <p>Spare Parts Inventory Planning Modeling the optimized deployment of 10,000 part numbers across 100 parts depots and requires: • 2 x 10¹⁴ FLOPs (12 hours on 10, 650MHz CPUs) • 2.4 PetaFlop/s sust. performance (1 hour turn-around time) Industry trend for rapid, frequent modeling for timely business decision support drives higher sustained performance</p>
 <p>Digital Movies and Special Effects ~ 1E14 FLOPs per frame 50 frames/sec 90 minute movie • 2.7E19 FLOPs • ~ 150 days on 2000 1 GFLOPs CPUs</p>	

Source: A. Johnson, et al. Source: G. Bailey, NERSC Source: Pixar Source: B. Dietrich, IBM

 **Top500**

- Máquinas con una gran potencia de cálculo que se emplean para ejecutar estas aplicaciones
- Las prestaciones se miden resolviendo un sistema de ecuaciones lineales denso, en 64 bits de precisión.
- El esquema general se basa en:
 - Múltiples nodos (memoria distribuida)
 - Cada nodo multiprocesador y cada procesador multicore (memoria compartida)
 - Conectados por una red de altas prestaciones.
- Frecuentemente se apoyan en coprocesadores para acelerar las rutinas más costosas.

6

Ejemplo de cluster HPC

201 TFLOPS en 7 racks
677 MFLOPS por watio (#9 on Green500, Nov 2010)

7

Rank	Site	System	Cores	Nodes	Power (1000W)	Power (1000W)
1	National Super Computer Center in Guangzhou, China	Tianhe-2 (88k/1.3M) - 71k/45k/1.3M Cluster, Intel Xeon E5-2682 10C 2.80GHz, 11 Express-2, Intel Xeon Phi 312P, HPC	1120000	33602.7	34002.4	17000
2	DOE/SC/Oak Ridge National Laboratory, United States	Titan - Cray XK7 - Opteron 6274 10C 2.00GHz, Cray Gemini interconnect, NVIDIA K20x, Cray PE	80540	1739.0	27112.3	8209
3	DOE/ERDC/LNL, United States	Sageblade - BlueGene/Q, Power 620 1.60 GHz, Custom IBM	117264	1717.3	26120.7	7600
4	RICIS Advanced Institute for Computational Science (AICS), Japan	K computer - SPARC64 V8x 2.0GHz, Tofu, Fujitsu	70020	10510.0	11090.4	12640
5	DOE/SC/Argonne National Laboratory, United States	Mira - BlueGene/Q, Power 620 1.60GHz, Custom IBM	78432	8088.0	10063.3	2845
6	Swiss National Supercomputing Centre (CSCS), Switzerland	Piz Daint - Cray XC30, Xeon E5-2670 8C 2.80GHz, Xeon Phi 312P, NVIDIA K20x, Cray PE	110964	6271.0	7788.9	2325
7	Toshiba Advanced Computing Center/Line, United States	Sageblade - PowerEdge C620, Xeon E5-2660 8C 2.80GHz, Infiniband QDR, Intel Xeon Phi 312P, Dell	42342	5168.1	6125.1	4010
8	Forschungszentrum Juelich (FZJ), Germany	JASMIN - BlueGene/Q, Power 620 1.60GHz, Custom Measurement IBM	41870	5058.0	5872.0	2301
9	DOE/ERDC/LNL, United States	Wukon - BlueGene/Q, Power 620 1.60GHz, Custom Measurement IBM	39276	4263.3	5033.2	1912
10	Leibniz Rechenzentrum, Germany	SuperMUC - Catalyst D630M, Xeon E5-2680 8C 2.80GHz, Infiniband FDR, NVIDIA K20x, IBM	18740	2097.0	3185.1	3623
11	GSFC Center, Tokyo Institute of Technology, Japan	T1SHIMANE 2.5 - Cluster Platform SL390s G7, Xeon E5-2670 8C 2.80GHz, Infiniband QDR, NVIDIA K20x, MSCW	7430	2843.0	6008.4	1099
12	National Supercomputing Center in Tianjin, China	Tianhe-1A - HPC 11k MP, Xeon E5-2670 8C 2.80GHz, NVIDIA 200, HPC	18078	2568.0	4701.0	4042

8

Green 500

- El consumo energético es muy importante
- Por coste y por la potencia (PFlop/s) alcanzada
- Los primeros dan unos 2 GFlop/w
- Un Kwh (10-12 céntimos) permite 7,2 PFlop ¡¡7200 billones de flop!!
- Importancia creciente porque se quiere alcanzar el Exaflops (1000 Petaflops):
 - ¡¡500 MW es inaceptable!!

9

Green 500

Green500 Rank	MFLOPS/W	Site*	Computer*	Total Power (kW)
1	4,503.17	GSFC Center, Tokyo Institute of Technology	T1SHIMANE KFC - LX 11-4GPU104Re-10 Cluster, Intel Xeon E5-2620v2 8C 2.100GHz, Infiniband FDR, NVIDIA K20x	27.78
2	3,631.86	Cambridge University	Wilkes - Dell T620 Cluster, Intel Xeon E5-2630v2 6C 2.80GHz, Infiniband FDR, NVIDIA K20	52.62
3	3,517.84	Center for Computational Sciences, University of Tsukuba	WALPACS TCA - Cray 382534-SM Cluster, Intel Xeon E5-2680v2 10C 2.80GHz, Infiniband QDR, NVIDIA K20x	78.77
4	3,185.91	Swiss National Supercomputing Centre (CSCS)	Piz Daint - Cray XC30, Xeon E5-2670 8C 2.80GHz, Xeon Phi 312P, NVIDIA K20x, Level 3 measurement data available	1,753.66
5	3,130.95	ROMEO HPC Center - Champagne-Ardenne	romeo - Bull RA21-E3 Cluster, Intel Xeon E5-2650v2 8C 2.80GHz, Infiniband FDR, NVIDIA K20x	81.41
6	3,068.71	GSFC Center, Tokyo Institute of Technology	T1SHIMANE 2.5 - Cluster Platform SL390s G7, Xeon X5670 6C 2.80GHz, Infiniband QDR, NVIDIA K20x	922.54
7	2,702.16	University of Arizona	iDataPlex D1308M4, Intel Xeon E5-2680v2 8C 2.80GHz, Infiniband FDR14, NVIDIA K20x	53.62
8	2,629.10	Max-Planck-Gesellschaft MP19PP	iDataPlex D1308M4, Intel Xeon E5-2680v2 10C 2.80GHz, Infiniband, NVIDIA K20x	269.94
9	2,629.10	Financial Institution	iDataPlex D1308M4, Intel Xeon E5-2680v2 10C 2.80GHz, Infiniband, NVIDIA K20x	55.62
10	2,358.69	CSIRO	CSIRO GPU Cluster - Nitro G16 3GPU, Xeon E5-2650 8C 2.80GHz, Infiniband FDR, Nvidia K20m	71.01

10

Programación HPC

- La arquitectura subyacente tiene una gran importancia para maximizar el rendimiento
- Uso de herramientas de análisis (*profilers*) para saber dónde optimizar y paralelizar.
- Memoria distribuida:
 - Se programa con MPI (paso de mensajes)
 - Hw: Destaca la red de interconexión
- Memoria compartida:
 - Se programa con OpenMP (*threads*)
 - Hw: Destaca sincronización y coherencia caches

11

Programación HPC

- Coprocesador Intel Xeon Phi:
 - Se programa con OpenMP y Vectorización
 - Hw: Destaca uso de directorios y unidades vectoriales
- Coprocesador GPU (Nvidia y AMD):
 - Se programa con Cuda (a veces tb OpenCL)
 - Hw: Destaca la jerarquía de memoria y ejecución síncrona (SIMT)
- Se combinan todos entre sí: MPI+OpenMP, MPI+Copro, OpenMP+Copro, MPI+OpenMP+Copro

12