



Procesamiento masivo al servicio del análisis de imagen en patología

Daniel Sánchez Morillo

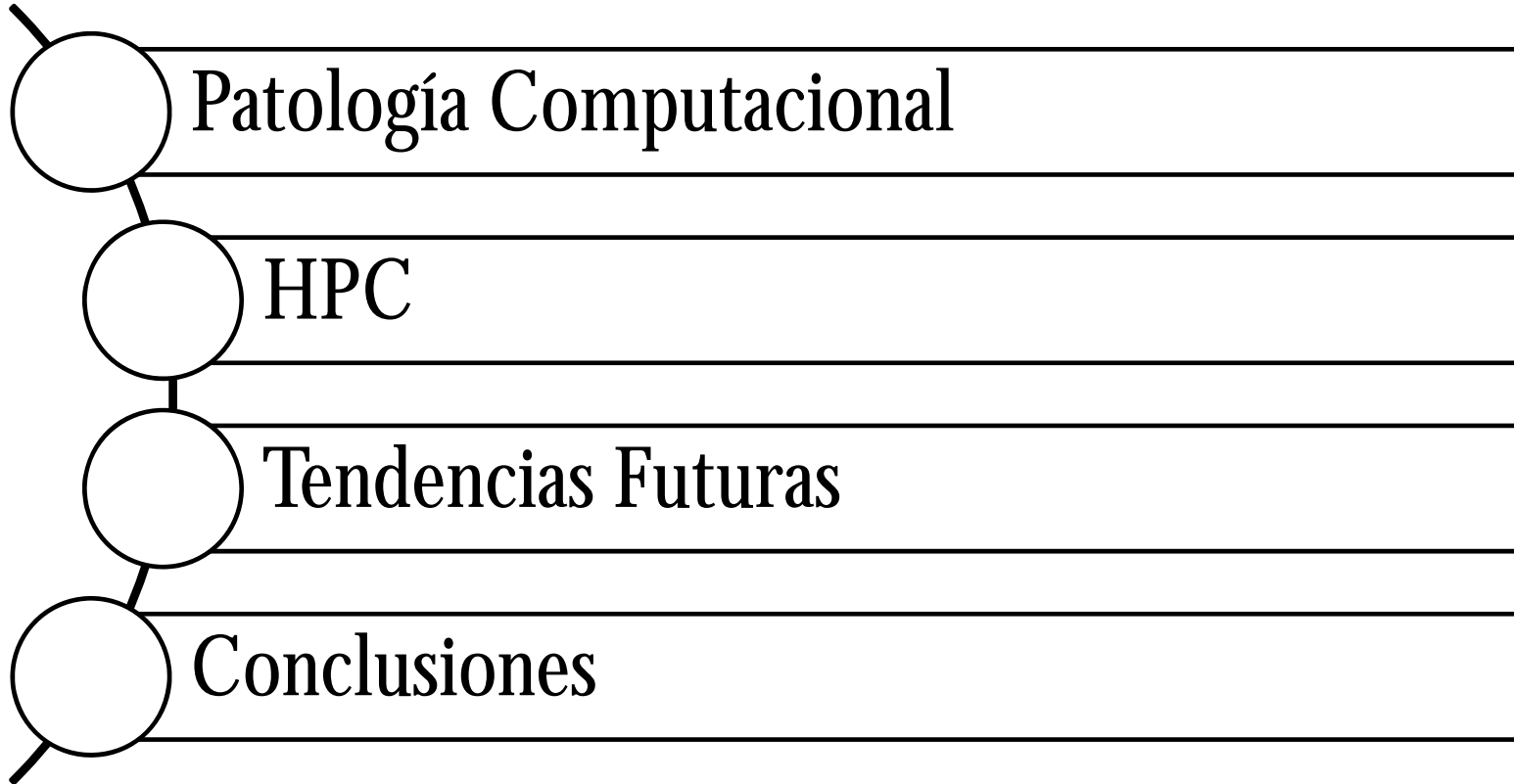
Universidad de Cádiz



Grupo de Investigación de
Ingeniería Biomédica y Telemedicina (IBT)
Octubre 2018

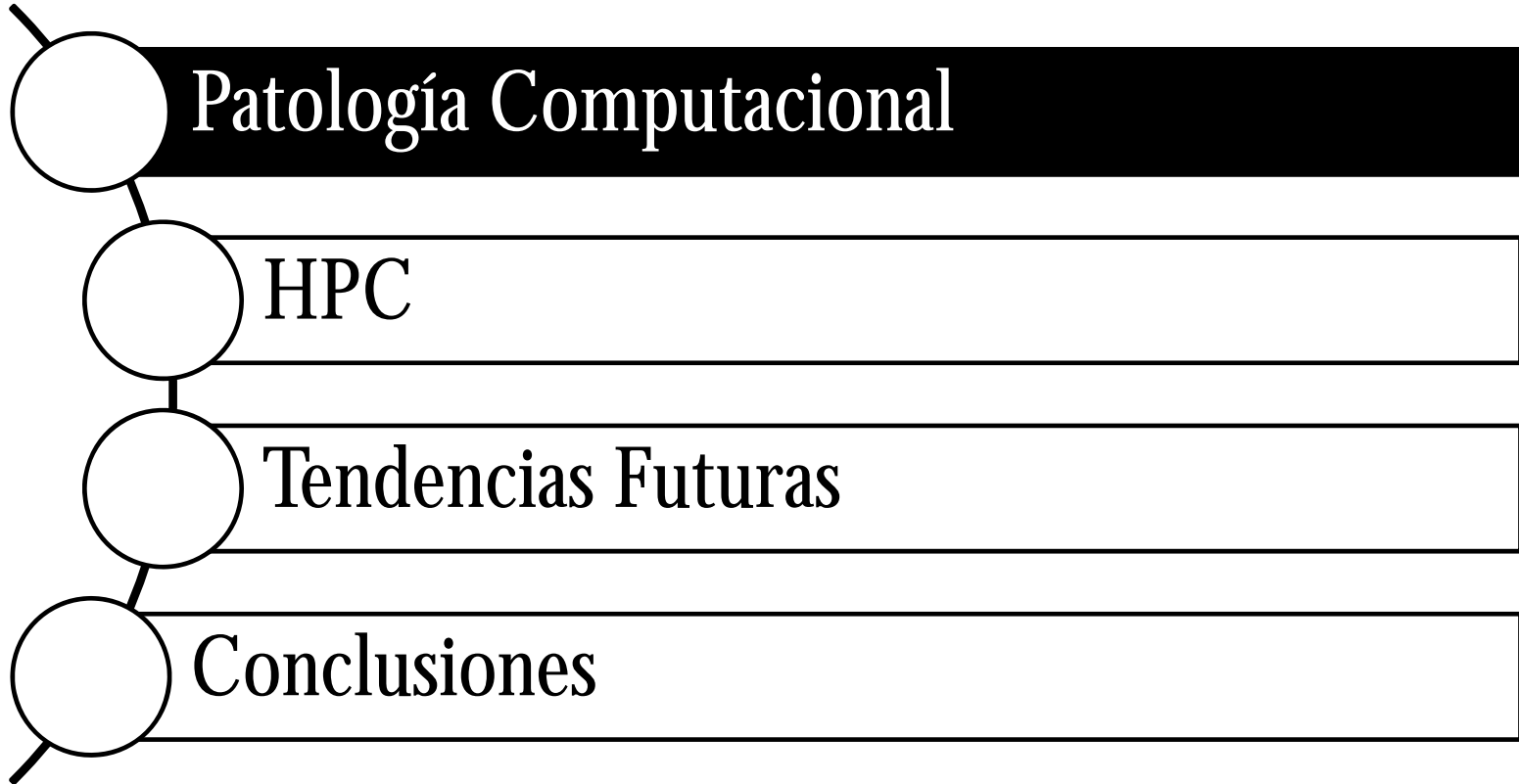
Contenidos

- Patología Computacional
- HPC
- Tendencias Futuras
- Conclusiones



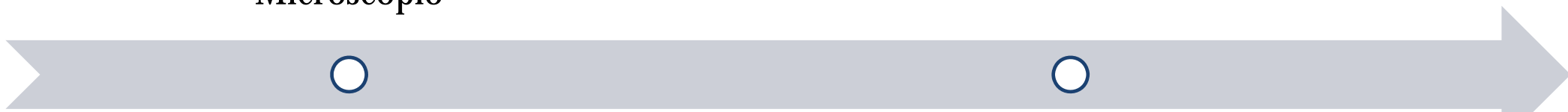
Contenidos

- Patología Computacional
- HPC
- Tendencias Futuras
- Conclusiones

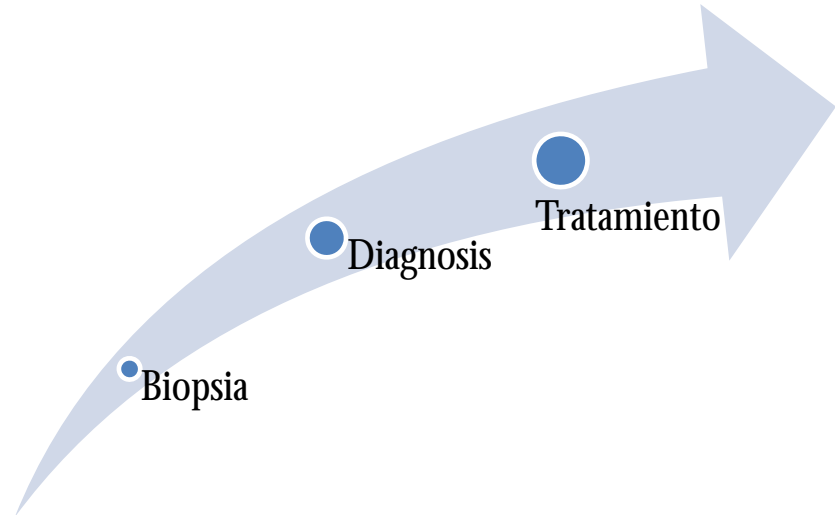


Patología Computacional

Microscopio



Diseño de terapias asociadas a la medicina de precisión



Patología Computacional



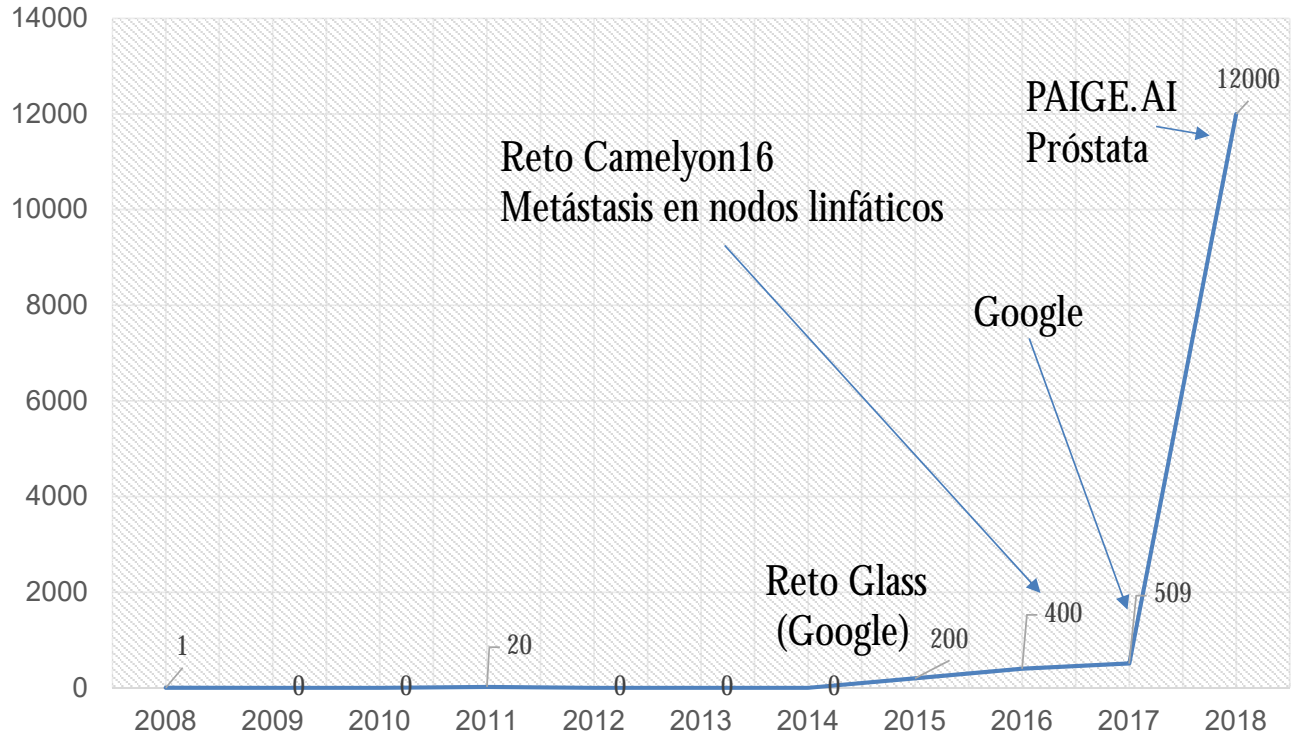
Med Image Comput Comput Assist Interv. 2008;11(Pt 2):1-8.

Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients.

Fuchs TJ¹, Wild PJ, Moch H, Buhmann JM.

Patología Computacional

α Carencia de grandes Datasets que evitan la transferencia de resultados al ámbito clínico



Patología Computacional



$100.000 \times 100.000 > 3 \text{ Gbytes}$

Patología Computacional



Preprocesado + Segmentación

Varias Horas en una CPU convencional

Patología Computacional

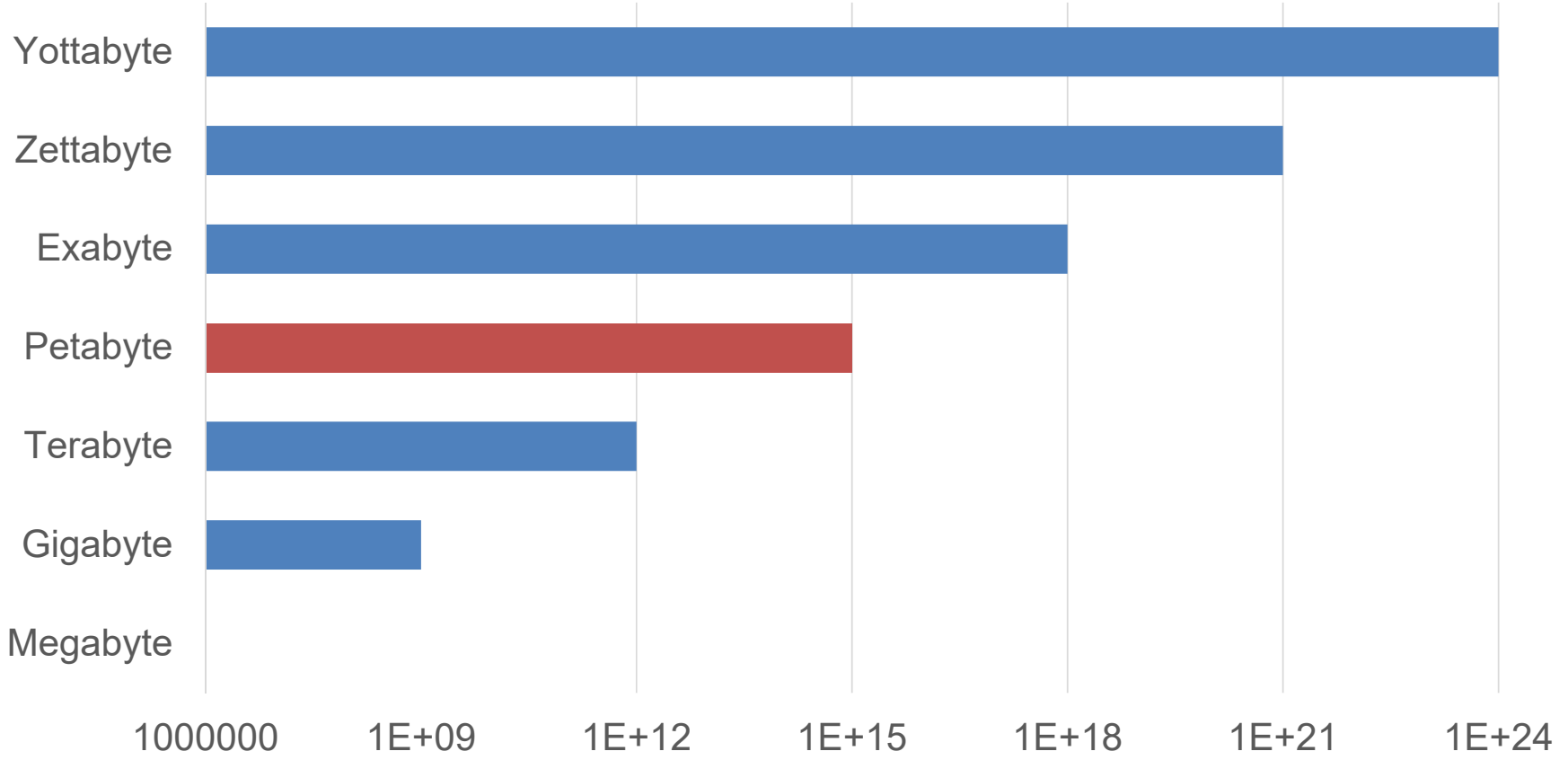


> 350.000 objetos

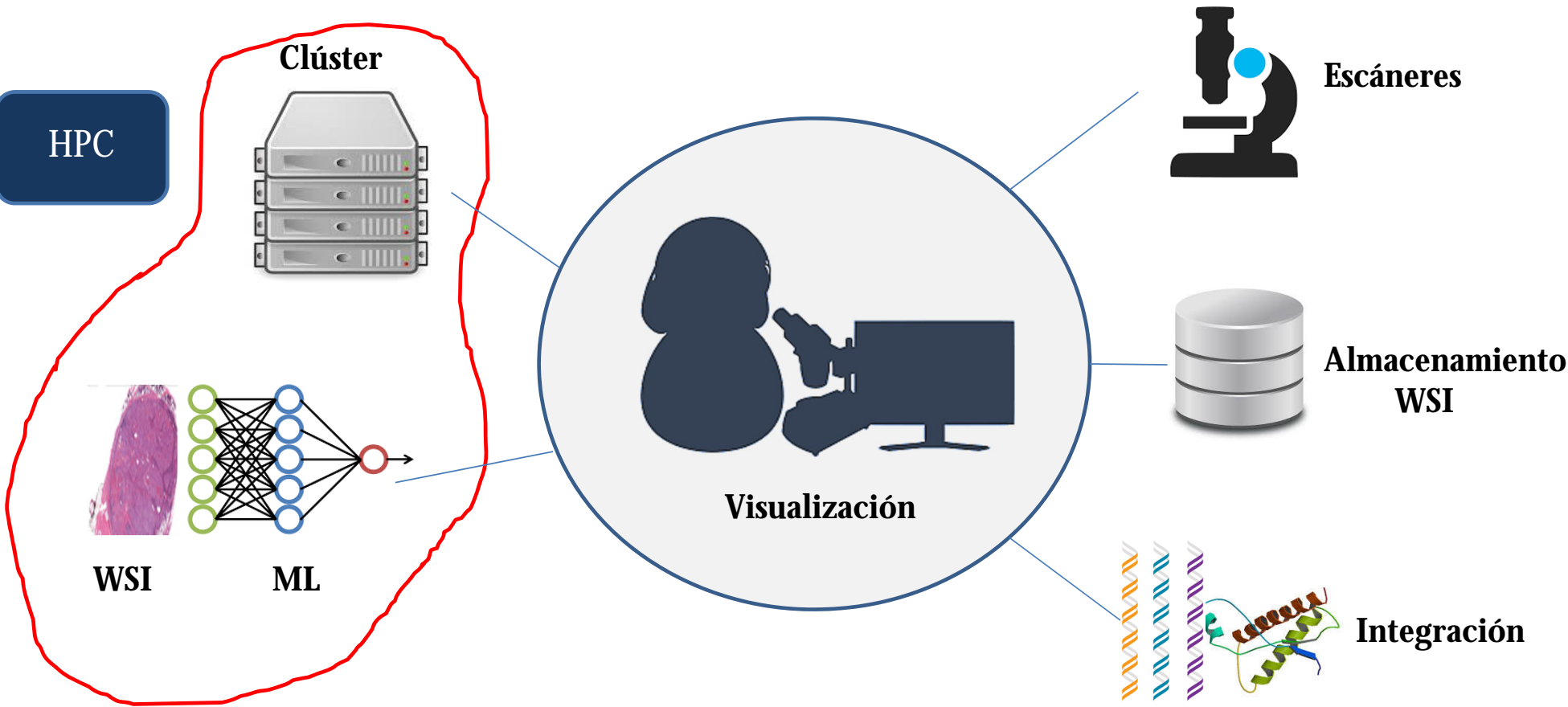
Extraer características – morfológicas, texturas

Entrenamiento + Validación

Patología Computacional



Patología Computacional



Laboratorio de patología = Ambiente Big-Data

Contenidos

- Patología Computacional
- HPC**
- Tendencias futuras
- Conclusiones

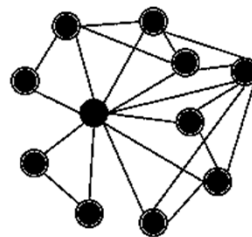
High Performance Computing

Uso de técnicas de procesamiento en paralelo para ejecutar programas de forma eficiente, fiable y rápida

Computación paralela, distribuida, en la nube

High Performance Computing

¿Por qué HPC?



High Performance Computing

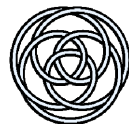
¿Por qué HPC en Patología Computacional?



Imagen = matriz / pixeles – RGB



Escala Tera-Peta



Creciente complejidad de algoritmos



Tiempos de respuesta poco útiles

High Performance Computing

¿Por qué HPC en Patología Computacional?

- Procesado paralelo de miles de imágenes
- Superpíxeles en un escaneado WSI
 - Procesar cada fotograma-campo simultáneamente
 - Acelerar el proceso $k \cdot N$

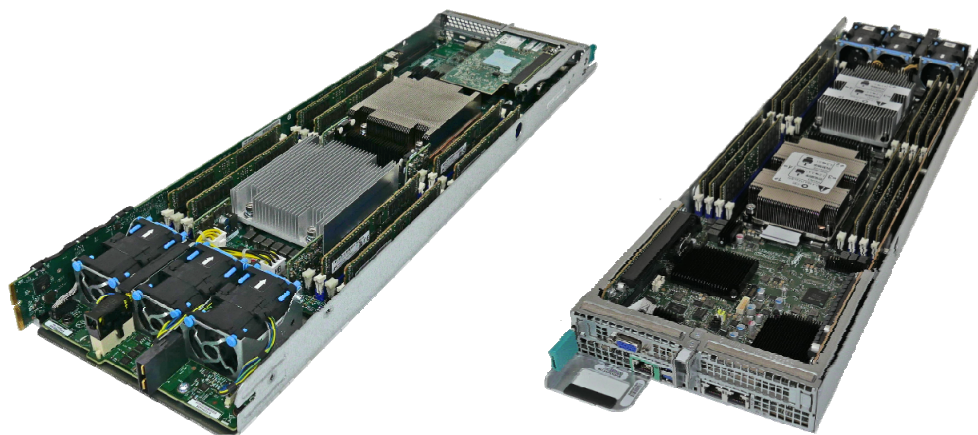
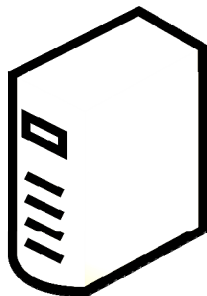
High Performance Computing

Terminología

Nodo = Ordenador individual (CPU, CPU+GPU,...)

Core = Unidad de Procesamiento.

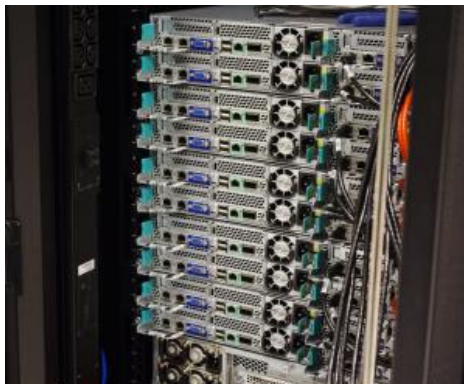
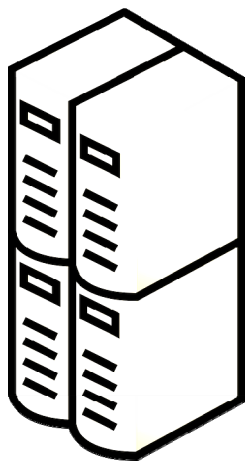
1 Nodo = Varios Cores



High Performance Computing

Terminología

Cluster= Conjunto de nodos, unidos por una red, trabajando de forma conjunta (4,16,64,128,256,...)



High Performance Computing

Un clúster de cálculo HPC y el software de cálculo lo que hacen básicamente es dividir el problema en problemas más pequeños.

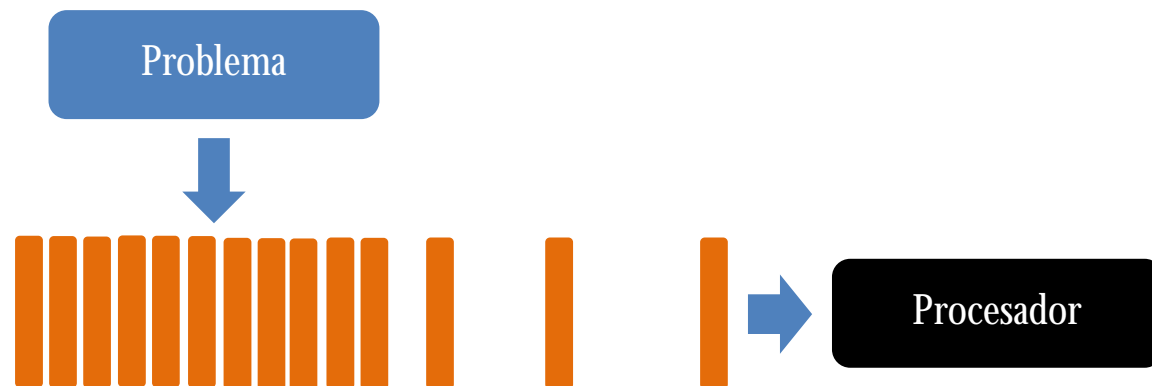
Dichos cálculos se realizan en los diferentes nodos.

A este proceso se le denomina **paralelización**.

High Performance Computing

α Computación serie

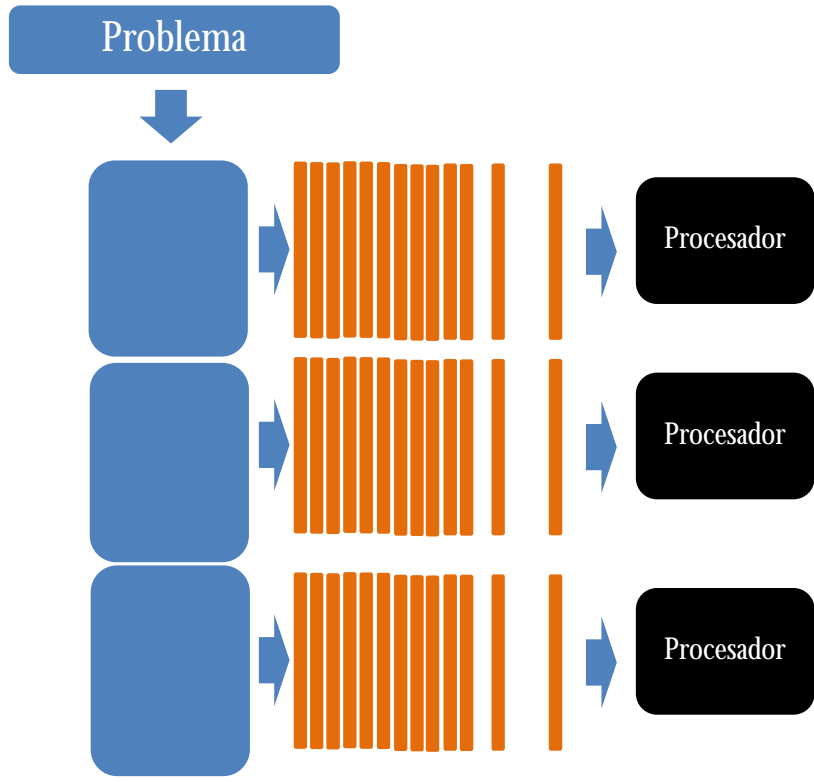
- Serie de Instrucciones discretas
- Ejecución secuencial - Único procesador
- Monopolio CPU



High Performance Computing

α Computación paralela

- Ejecución de múltiples instrucciones en un mismo instante
- Mejor rendimiento en tiempo de ejecución



High Performance Computing

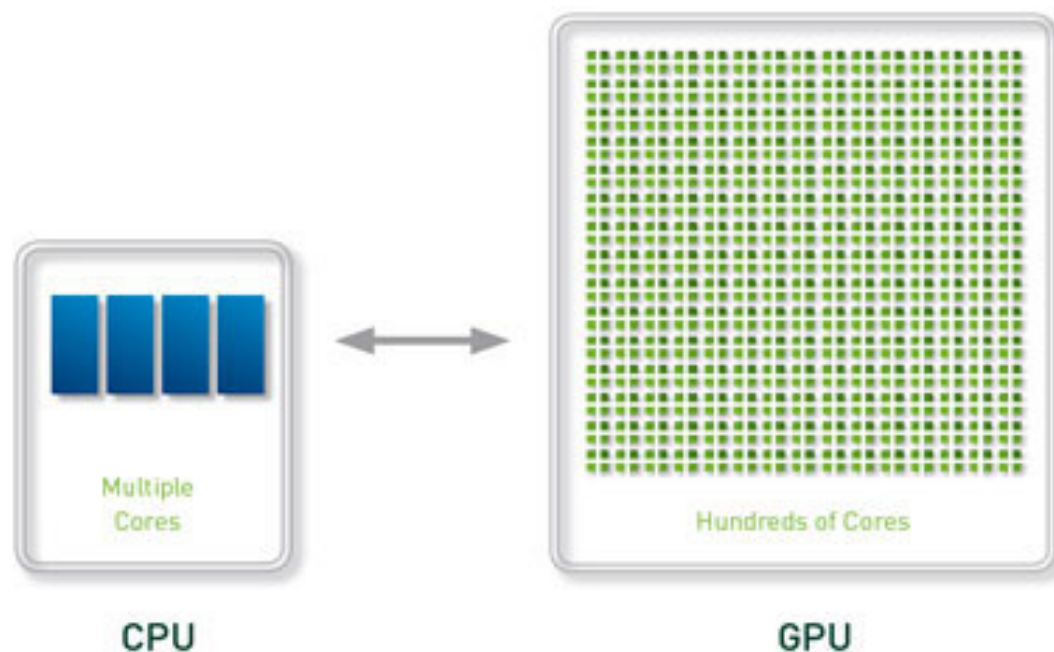
α Componentes de un sistema HPC

1. Nodos (cabecera, computación, acceso)
2. Memoria masiva
3. Sistema de ficheros
4. Red de comunicación alta velocidad
5. Sistema Operativo
6. Software de paralelización

High Performance Computing

α Hardware

- CPU Multi-núcleo
- GP-GPU (CUDA)

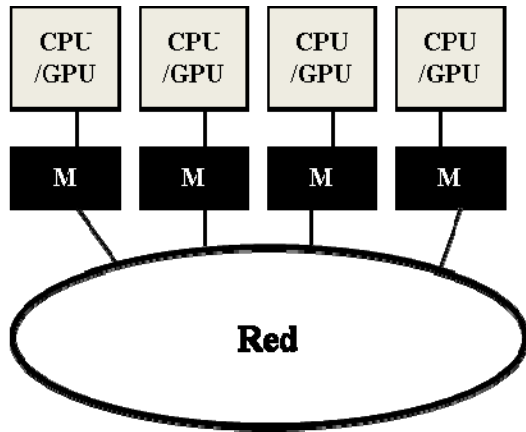


Suavizado imagen → arquitectura
CUDA 10 veces más rápida CPU

High Performance Computing

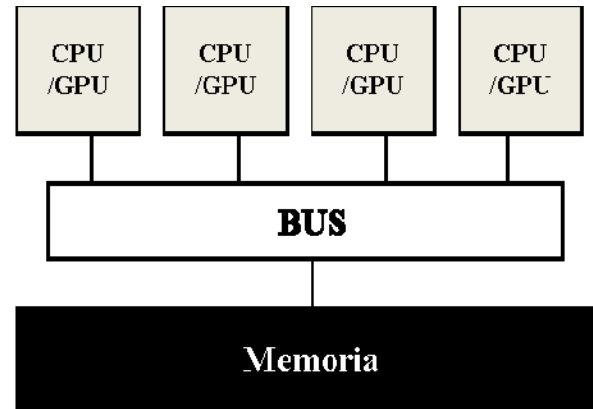
α Programación

– Depende de la arquitectura



Memoria Distribuida.

Cada procesador tiene memoria local.
Message passing (MPI) para intercambio de datos entre procesadores.



Memoria compartida.

Un único espacio al que se acceden los distintos procesadores →



GPU

CUDA (Nvidia) - 2006
OpenCL

Arquitecturas híbridas
MPI + CUDA + ...

High Performance Computing

Factores de diseño de un Cluster HPC

1. Hw Nodos + red de comunicaciones
 - Gigabit
 - Infiniband (100Gbit/s)



High Performance Computing

Factores de diseño de un Cluster HPC

2. Físicos

- Red eléctrica
- HVAC
- Espacio físico

High Performance Computing

Factores de diseño de un Cluster HPC

3. Software

- SO: Linux (99%), Windows
- Compiladores
- Librerías Paralelización (abiertas)
- **El coste de desarrollo en la programación paralela es mayor**



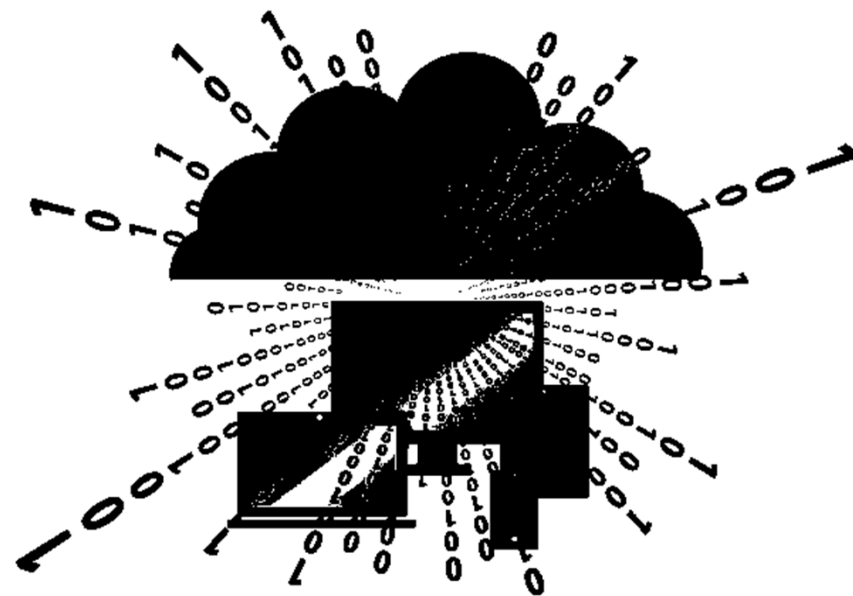
High Performance Computing

α GRIDs

- Computación distribuida
- Hardware heterogéneo
- Geográficamente disperso

α HPCC

- Configuración bajo demanda
- ``Hosted``
 - Provisión de servidores CPU, GPU y FPGA
 - Menor inversión inicial
 - AMAZON



High Performance Computing

Límites en HPC

α Ley de **Amdahls**

– S_n = speedup

$$S_n = \frac{1}{(1 - f) + \frac{f}{s}}$$

f=parte paralela

s=speedup de la parte paralela

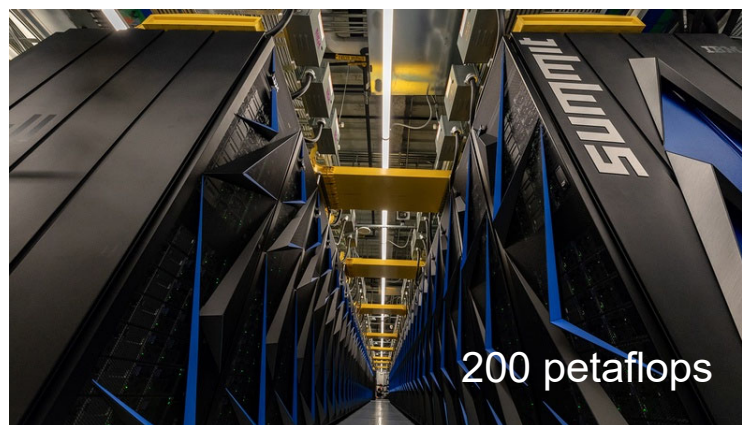


El speedup global está limitado por la fracción del programa que se puede mejorar.

High Performance Computing

1. Summit, IBM-built supercomputer

- Oak Ridge National Laboratory (Tennessee, EEUU)
- 4,356 nodos cada uno con
 - 2 CPUs Power9 de 22-cores y 6 NVIDIA Tesla V100



Superordenador Summit.
Fuente: Oak Ridge National Laboratory



Un Centro HPC de Excelencia



Memorial Sloan Kettering
Cancer Center



MSKCC's HPC Cluster

Fuente: www.mskcc.org

Un Centro HPC de Excelencia



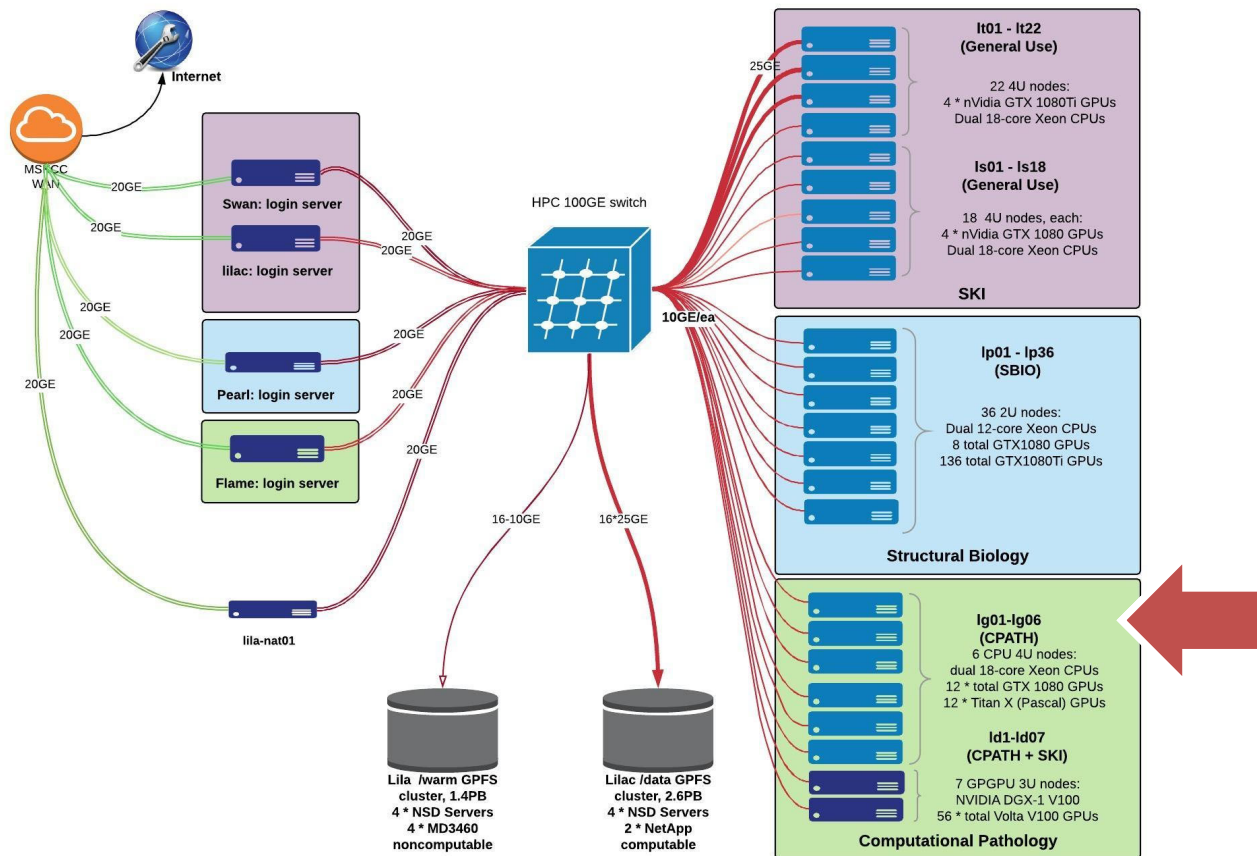
Memorial Sloan Kettering
Cancer Center

- El mayor instituto de investigación en cáncer de los EEUU
- 25M WSI histopatológicas de cáncer

PAIGE

Paige.AI es el acrónimo de
“Pathology AI Guidance Engine.”
Cortesía de Paige.AI.

Un Centro HPC de Excelencia



Fuente: www.mskcc.org

Un Centro HPC de Excelencia

α 6 nodos CPU dual 18-core Xeon

- GTX 1080
- Titan X

α 7 nodos **NVIDIA DGX-1**

- Diseñados para DeepLearning
- Rendimiento > 400 servidores convencionales

α Cada Sistema tiene

- 8 GPUs **Tesla V100** – Librería OpenSlide
- CPU Dual 20-core
- **PyTorch** usado para cargar los datos, construir modelos y entrenar



**Memorial Sloan Kettering
Cancer Center**



Un Centro HPC de Excelencia

α IA

- **CNN** para clasificación y segmentación
- **RNN** (recurrent neural networks) para extracción de información para informes de patología
- **GAN** (Generative adversarial networks) para aprender de imágenes sin anotar



Memorial Sloan Kettering
Cancer Center

Un Centro HPC de Excelencia

Fuchs → conjunto de datos de dimensiones sin precedentes en la patología computacional

- 12160 WSI -> biopsias de aguja de próstata
- 20X → 2424(+) 9736(-)
- 70% + 15% + 15%
- Deep Learning, AUC = 0.98
- FN=4.8% Test set 1.824 WSI

arXiv.org > cs > arXiv:1805.06983

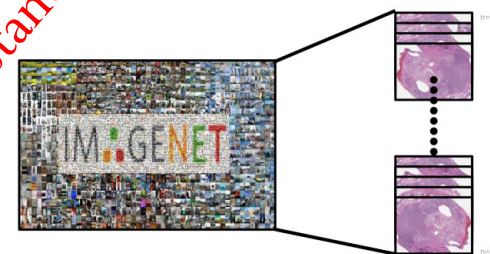
Computer Science > Computer Vision and Pattern Recognition

Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology

Gabriele Campanella, Vitor Werneck Krauss Silva, Thomas J. Fuchs

(Submitted on 17 May 2018 (v1), last revised 27 Sep 2018 (this version, v2))

Multiple-instance binary classification

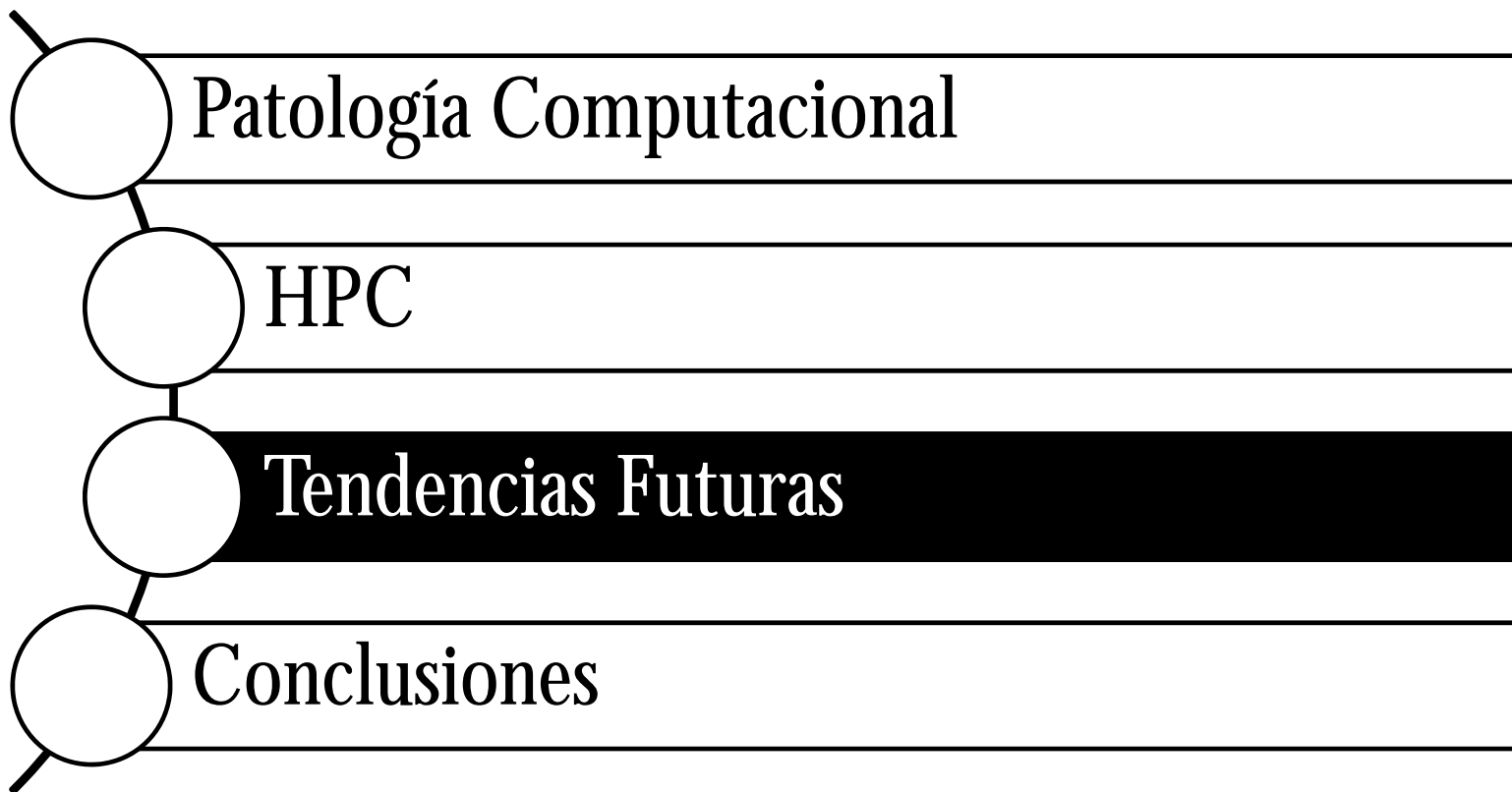


All of ImageNet
 482 x 415 * 14,197,122
 = 2.8 trillion pixels

474 Whole Slides
 100,000 x 60,000 * 474
 = 2.8 trillion pixels

Terabyte scale

Contenidos



Tendencias Futuras

α Report ASCAC, Dic. 2017



1. Lógica reconfigurable (FGPA).

- Mapear aplicaciones específicas en FPGAs, con ventajas en consumo, rendimiento en niveles exascale
- Convertir software en hardware
- En 2-5 años, convivencia de CPUS con FPGA
 - Grandes posibilidades para desarrollos en la nube (AMAZON)

2. Tecnologías ópticas (Silicon Photonics)

- Anchos de banda ultra-altos
- Eficiencia energética

Post-Exascale (2020's)

Tendencias Futuras

3. Computación neuromórfica

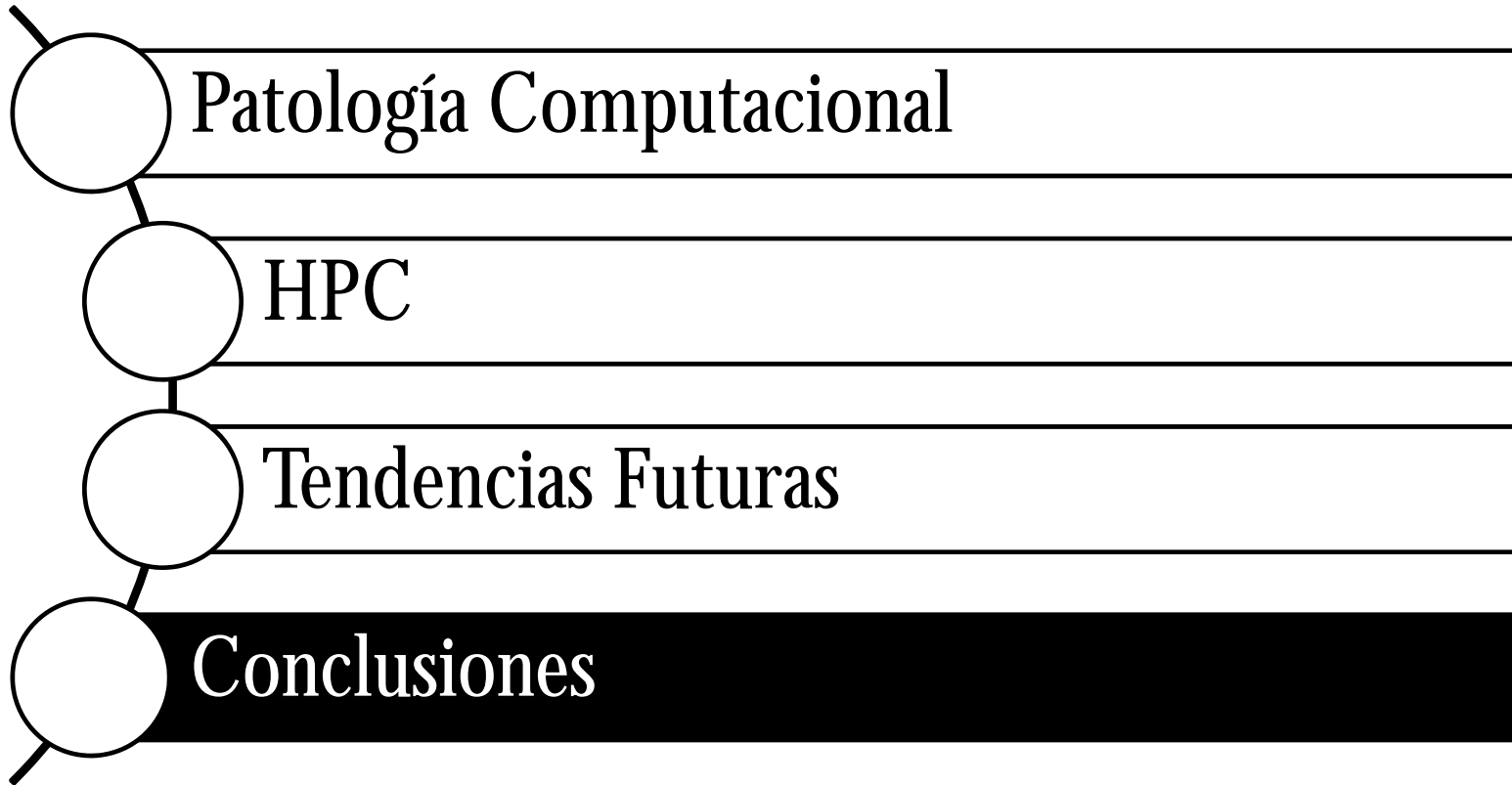
- Implantación hardware de circuitos emulando el comportamiento de neuronas y sinapsis
- Ideales para DNN
- IBM, Google, Microsoft, NVIDIA
- No competitiva todavía

4. Computación Cuántica y Computación Analógica

- Cubits
- Computador cuántico de 30 cúbits = 10 teraflops convencionales
- Google -> prototipo ordenador cuántico 'Bristlecone' (2018)

Post-Moore (2030's)

Contenidos



Conclusiones

1. HPC imprescindible con la llegada de las grandes BBDD en PD
2. IA , Big-Data
3. Inversión estratégica
4. Configuración escalable
5. RRHH

¡Gracias por su atención!

**“El edificio de la medicina descansa
en el diagnóstico del patólogo”
(Tomas Fuchs, 2018)**

Más datos, mejores decisiones

daniel.morillo@uca.es