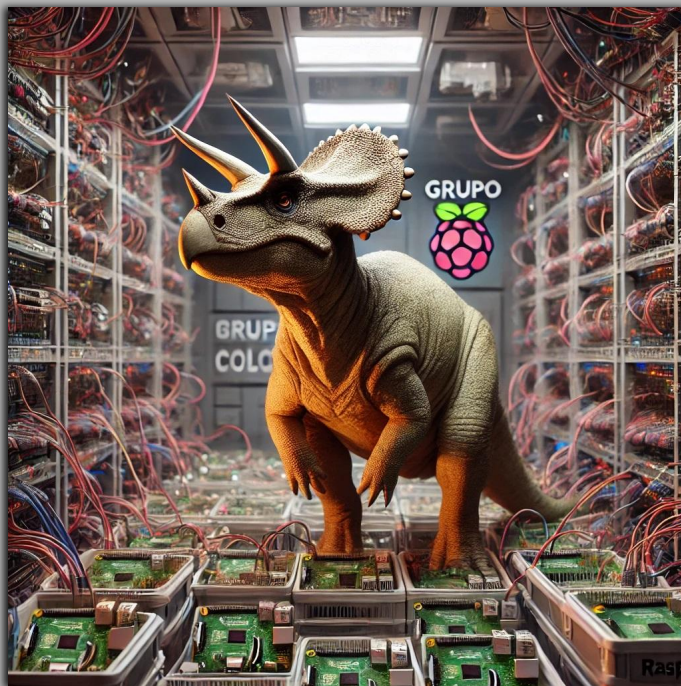




Universidad  
Nacional de  
General  
Sarmiento



ICI

Instituto de Ciencias

# Workshop Introducción al Cómputo Paralelo

Lic. Martha Semken - Lic. Mariano Vargas

Director: Dr. Andrés Rojas Paredes

Grupo de investigación ColosUngs

# Workshop: ¿Por que estamos acá?



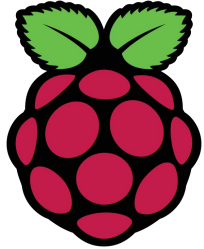
¡Es diciembre, tenemos que dar finales!

... y nosotros tomarlos y corregirlos



# Workshop: Quienes nos auspician

- UNGS
- ICI
- IDEI



y el Grupo ColosUngs, y de paso les cuento ... (¡viene la propaganda!)



**Universidad  
Nacional de  
General  
Sarmiento**

**ICI**

Instituto de Ciencias



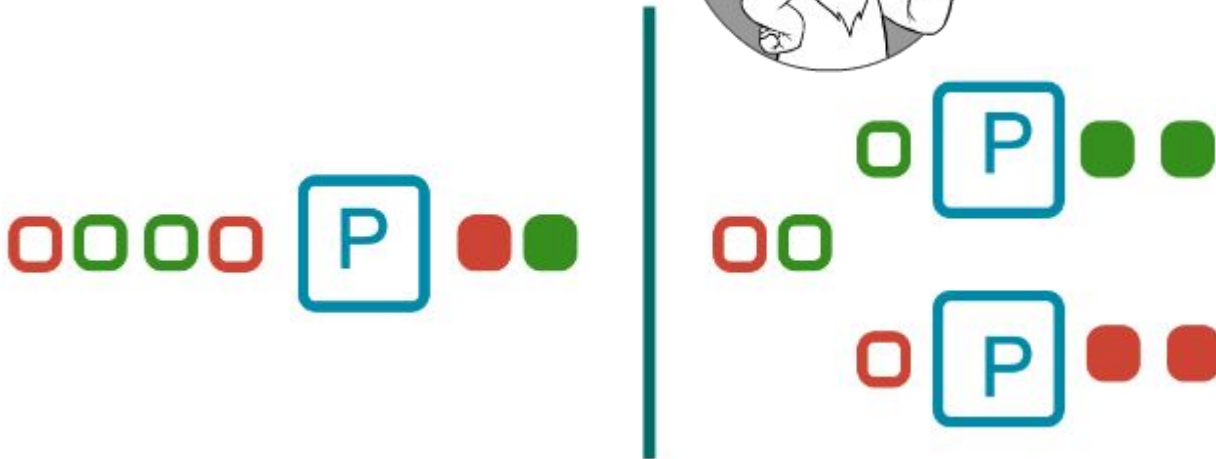
# Workshop: Nosotros

Aunque no lo crean, nosotros también estudiamos y nos tomamos examen!

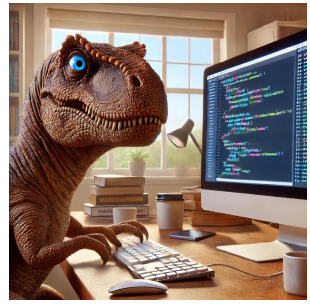




Workshop: Ok, ¿y que hacen?



# Workshop: ¿Para que me sirve?



- ¿Para aprender?
- Y para ganar puntos: Para los/las estudiantes de la Lic. En sistemas que participen de ambos encuentros y aprueben la evaluación, se podrán acreditar dos puntos IDEI y para el público en general que asista a los dos encuentros, se podrá acreditar 1 punto por asistencia.
- Por lo tanto hay que estar presentes en ambos encuentros, y para el/la que quiera esos dos puntos, examen

# Workshop: volvamos a lo nuestro



¿Qué estudiaremos?

# Workshop: Introducción al Cómputo Paralelo HPC

Agenda:

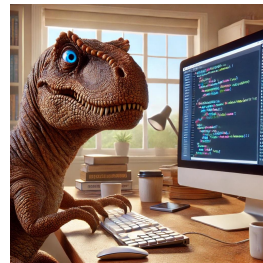
Sábado 7 de diciembre:

- Intro, conceptos generales, arquitectura, capas
- Hardware
- Software
- Modelo de Programación
- cómo acceder a un cluster
- y qué hago





# Workshop: Introducción al Cómputo Paralelo HPC



## Agenda:

Sábado 14 de diciembre:

- OpenMpi
- OpenMp
- Programación Híbrida
- Ejercicios
- Examen

# Workshop: ¿Qué es el cómputo paralelo o HPC?



- HPC: High Performance Computing
- Consiste en aumentar poder de cómputo con el fin de lograr resultados mucho más rápido de los que se logra con una computadora.
- Generalmente involucra grandes problemas en ciencia, ingeniería o negocios,

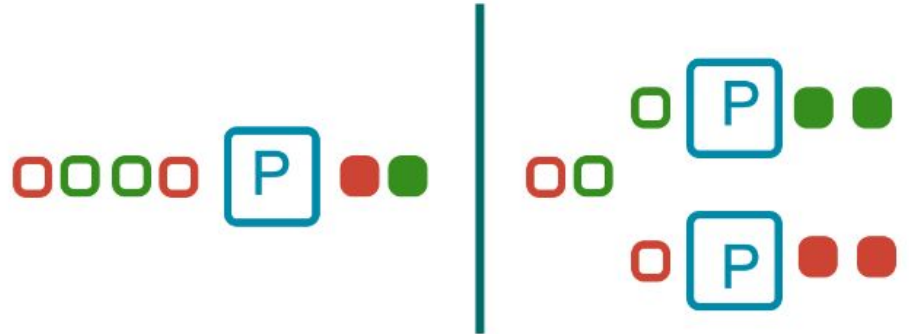
# Workshop: ¿Qué es el cómputo paralelo o HPC?

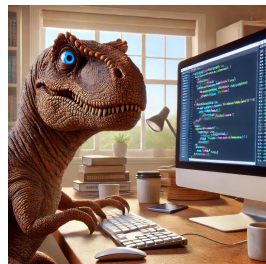


también, El cómputo paralelo es una técnica de programación y ejecución que divide un problema en tareas más pequeñas, distribuyéndolas entre múltiples procesadores para ser resueltas simultáneamente

Objetivo: Aumentar la velocidad de ejecución y/o manejar problemas de gran tamaño, o sea aumentar el poder de cómputo.

Una mirada simple e inocente  
del tema.





# ¿Por qué es Importante el Computación Paralela?

## Motivaciones:

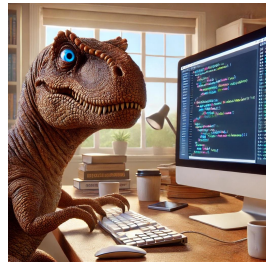
- Acelerar el procesamiento de tareas complejas.
- Manejar grandes volúmenes de datos (big data).
- Utilización eficiente de recursos en clústeres y supercomputadoras.

## Ejemplos de uso:

- Simulaciones científicas (clima, física, biología).
- Inteligencia Artificial y Machine Learning.
- Procesamiento de imágenes y video.
- **Consulta:** ¿En tu celu tenés la app para ver el pronóstico del tiempo, de dónde pensás que saca esa info? ¿de la nube de Dios?



# Clasificación HPC



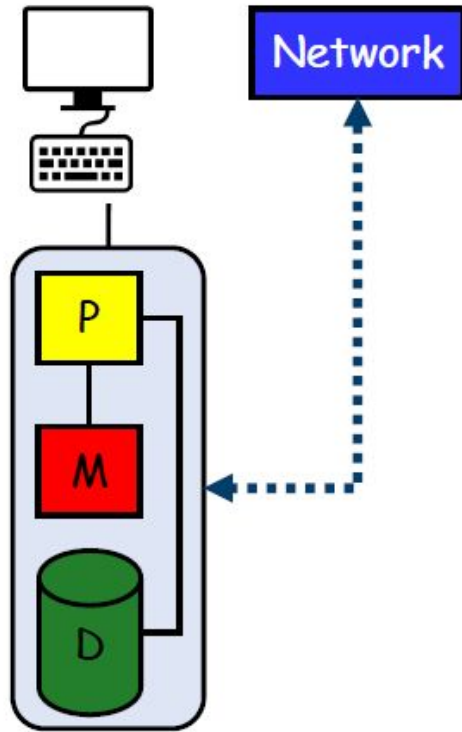
## Niveles de paralelismo:

1. Paralelismo a nivel de datos: Procesar múltiples datos simultáneamente (ej. matrices, imágenes).
2. Paralelismo a nivel de tareas: Ejecutar diferentes tareas independientes en paralelo.
3. Paralelismo combinado (híbrido): Combinación de datos y tareas.

## Arquitecturas:

1. Memoria compartida (múltiples núcleos, misma memoria).
2. Memoria distribuida (nodos con memorias independientes, como en Cronos).
3. Híbrida: Combina ambas.

# De nuestro viejo mundo secuencial al nuevo mundo paralelo

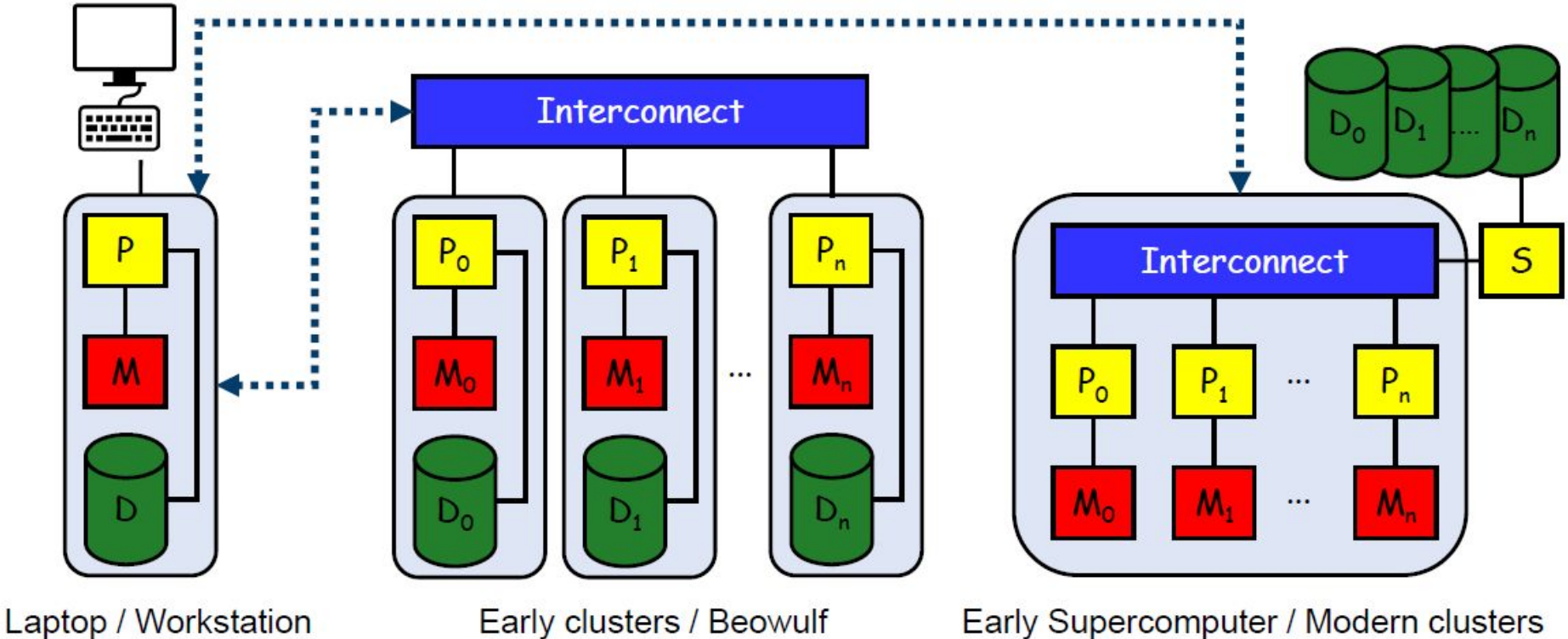


Laptop / Workstation

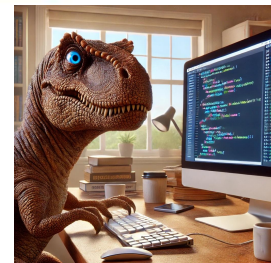
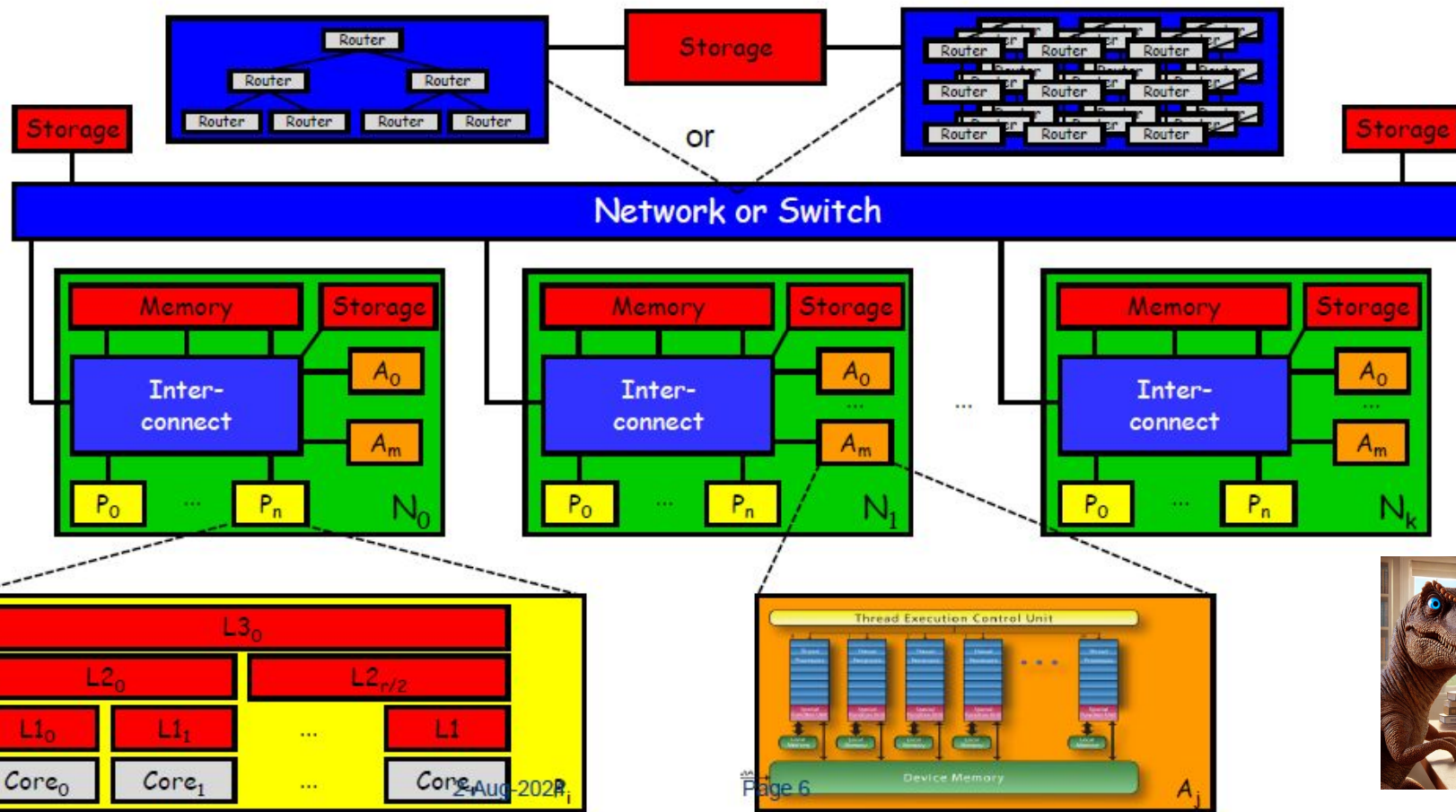
- **Processor**
  - Central Processing Unit (CPU)
    - Multi-core (2-8) vs. Many-core (dozens to hundreds)
- (Main) **Memory**
  - (Dynamic) Random Access Memory (RAM) (DRAM)
  - High Bandwidth Memory (HBM)
- **Disk**
  - Hard Disk Drive (HDD)
  - Solid State Drive (SSD)
- **Network**
  - Router
  - Switch
  - Interconnect



# El nuevo mundo paralelo



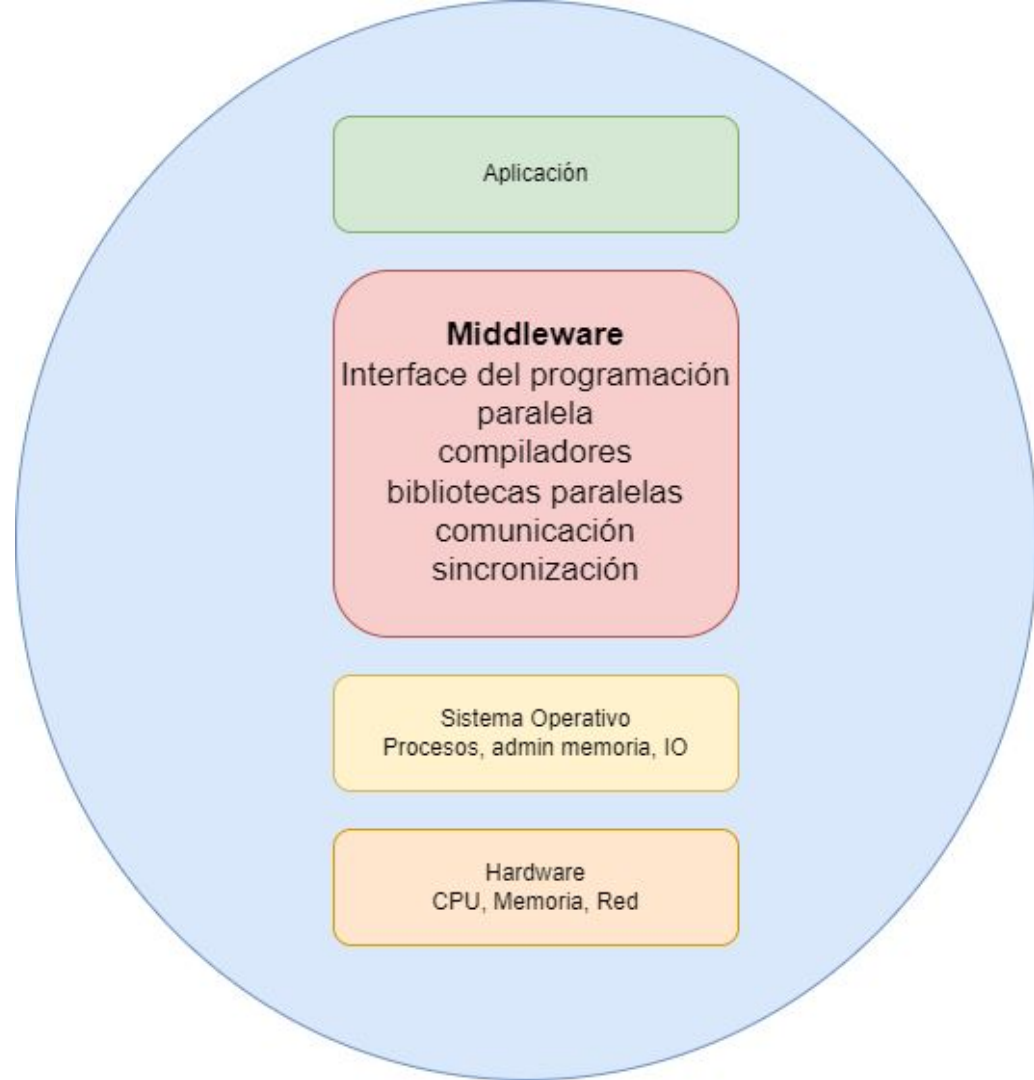
# PARALLEL ARCHITECTURES: STATE OF THE ART ~~MESS~~





# Alta complejidad en sistemas paralelos y distribuidos

## 4 CAPAS



# Beneficios y Desafíos del Cómputo Paralelo



## Beneficios:

- Reducción del tiempo de ejecución.
- Escalabilidad: manejar problemas más grandes al agregar más recursos.
- Resolución de problemas que no podrían manejarse de manera secuencial.

## Desafíos:

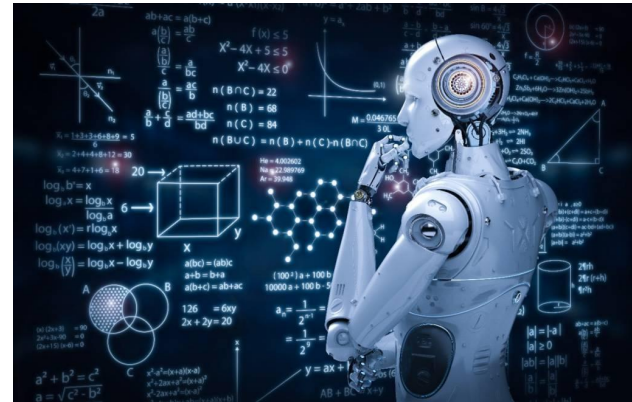
- División de tareas: Cómo dividir el problema para que sea eficiente.
- Sincronización: Coordinar tareas para evitar bloqueos o esperas.
- Comunicación: Overhead en sistemas distribuidos (como en Cronos).
- Hardware limitado: Procesadores y memoria.

# Aplicaciones Reales del Cómputo Paralelo



## Ejemplos destacados:

- Supercomputadoras: simulaciones climáticas y genómicas.
- Machine Learning: entrenamiento de modelos.
- Juegos y gráficos: simulación y rendering.
- Procesamiento de datos en tiempo real: motores de búsqueda, transacciones bancarias.



# Ahora si, ¿Que es un cluster?

**Definición:** Un clúster es un conjunto de computadoras interconectadas que trabajan juntas como si fueran una sola unidad de cómputo.

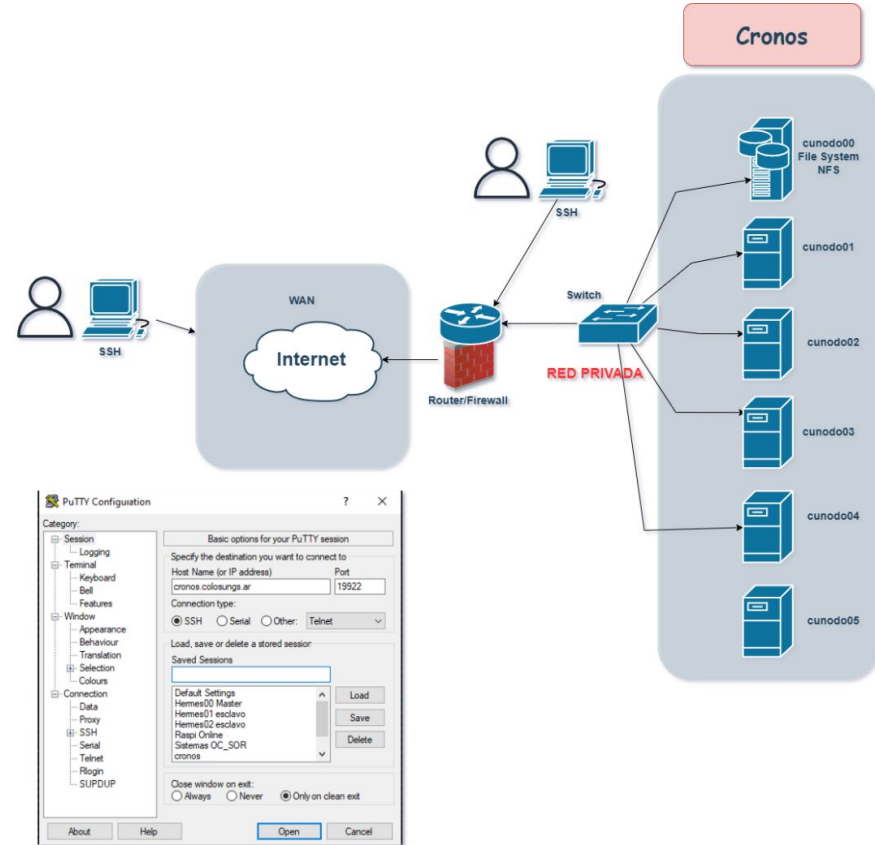
Componentes de un clúster:

- Nodo maestro: Controla y distribuye las tareas.
- Nodos de cómputo: Ejecutan las tareas distribuidas.
- Red: Conecta los nodos.
- Arquitectura Beowulf:
  - Nuestro clúster Cronos sigue este modelo.



# Arquitectura de Cronos

- Arquitectura: Memoria distribuida
- Basado en Raspberry Pi (Nodo Maestro: RPI 4 8GB, esclavos RPI 4 4GB)
- Red Ethernet: 1 Switch 16 bocas para tráfico aislado (100 Mbps)
- Router WiFi, red privada



# Arquitectura de Cronos

## Software:

- SLURM, MUNGE, Open MPI, OpenMP.

## Ventajas del clúster Raspberry Pi:

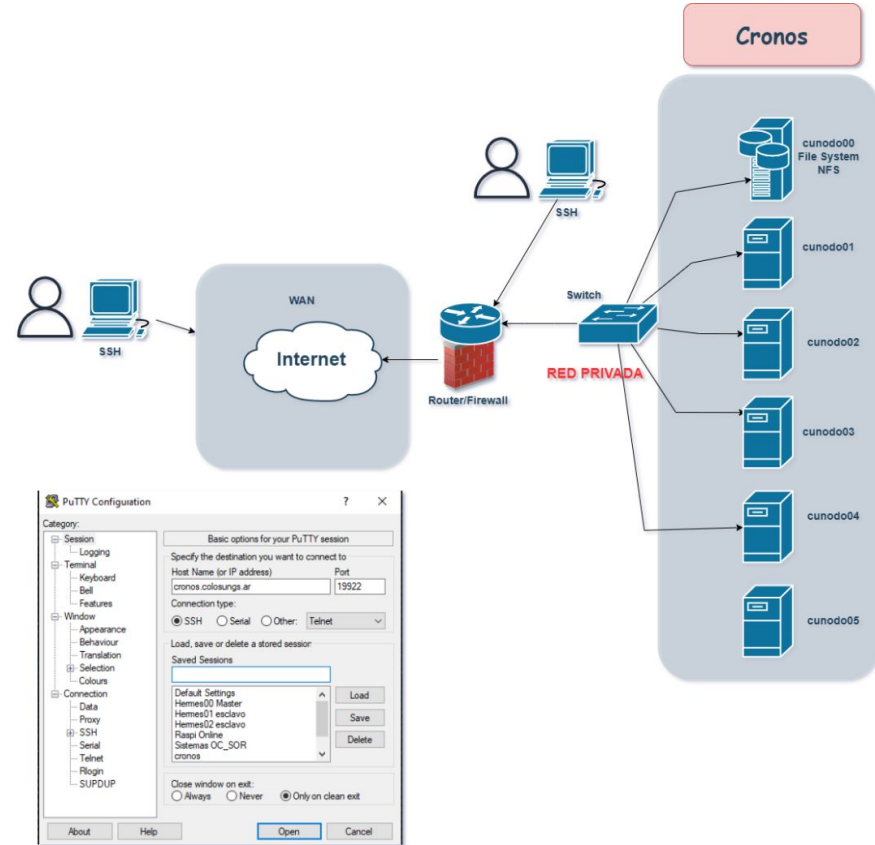
- Bajo costo, accesibilidad.
- Ideal para aprendizaje y pruebas.

## Limitaciones:

- - Menor potencia comparada con clústeres HPC.

## Green Computing:

¿Qué es?



# Green Computing



# Green Computing

**Green computing**, también conocido como "green IT" o "Tecnologías de la Información Verde", es un enfoque que se centra en la utilización eficiente y sostenible de la tecnología de la información y las comunicaciones (TIC) con el objetivo de reducir su impacto ambiental y promover prácticas más respetuosas con el medio ambiente en la industria de la tecnología y la informática. El término "green" en este contexto se refiere a la sostenibilidad y la preocupación por el medio ambiente.





# Green Computing

## Green computing:

- Eficiencia Energética
- Reciclaje y Reutilización
- Virtualización
- Evaluación de Ciclo de Vida
- Fuentes de Energía Renovable
- Educación y Concienciación



## SIMPLX LINE OF RESOURCE MANAGEMENT

## SIMPLX LINE OF RESOURCE MANAGEMENT

## Simple Linux Utility for Resource

## SLURM Controller Nonce



## Login Node

Login node

## Slurm Controller

Controlled

## Compute Nodes

**Jodes**

Assimite  
to somute  
resomute

## Compute Nodes

## JOBS

Simple Jobs  
to available compute  
compute nodes  
measurable

Joint Nodes

## Compute Nodes

## Compute Nodes

## Compute Nodes

# Herramientas Principales: SLURM y MUNGE



## MUNGE:

- Sistema de autenticación segura entre nodos.
- Uso: permite que los nodos confíen entre ellos.

## SLURM: (TOP 500)

- Sistema de gestión de trabajos en el clúster.
- Permite asignar tareas a nodos, gestionar recursos, y monitorizar trabajos.

<https://top500.org/lists/green500/2024/11/>

# Slurm



El sistema que se encarga de gestionar y asignar los recursos del servidor (tiempo de cálculo o memoria) a cada usuario es SLURM, un sistema de gestión de tareas y de clústeres. Con esto se consigue un mejor aprovechamiento de los recursos y evita que un solo usuario monopolice toda la máquina cuando haya otros esperando para ejecutar sus trabajos.

Slurm (Simple Linux Utility for Resources Management), es un sistema de gestión de trabajos de alta calidad utilizado en entornos de computación de alto rendimiento (HPC) y clústeres de computadoras. Slurm se utiliza para programar trabajos en un clúster y administrar recursos como nodos, procesadores y memoria. Permite a los usuarios solicitar recursos de manera eficiente y equitativa para sus trabajos, y asegura que los recursos se utilicen de manera óptima en el clúster. Slurm como sistema de gestión tiene tres tareas claves:

1. Asignar a los usuarios acceso exclusivo o no exclusivo a nodos de cómputo durante un tiempo determinado para que puedan ejecutar sus tareas.
2. Proporciona un framework que permite iniciar, ejecutar y supervisar el trabajo.
3. Se encarga de arbitrar la disputa de recursos, administrando una cola de tareas pendientes.

# Slurm

Comandos básicos:

En el pizarrón



# Slurm



Los comandos de slurm son necesarios para la iniciación y configuración de los cluster. Tienen una estructura definida donde deben cumplirse ciertas reglas:

- Empezar el comando con #SBATCH, que describen los recursos que requiere un trabajo.
- Cuando se encuentra un comando unix, el resto de las líneas #SBATCH son **IGNORADAS (ojo con esto, SI-SO-SE)**.



# Slurm



## Recursos

Se deben enumerar los recursos requeridos por el trabajo.

## Cola de ejecución

Los distintos recursos de cómputo del cluster Cronos son accesibles a través de colas de trabajos, que son líneas de espera en las que trabajos con requerimientos similares se ubican hasta que les toque su turno y los recursos requeridos estén disponibles. Las colas disponibles y sus límites dependen del cluster.



## Procesos (tasks)

- Cada trabajo en ejecución consiste en una cantidad de procesos corriendo simultáneamente.
- Los procesos se distribuyen en uno o más nodos de cómputo (computadoras del cluster) según la disponibilidad de recursos.
- En las aplicaciones MPI, la cantidad de procesos es el parámetro `-np` pasado a `mpirun` o `mpiexec`.
- En el script, la cantidad `P` de procesos del trabajo se define

¿Preguntas?

