



Jornadas de Astronomía Computacional Argentina  
29-30 de Noviembre de 2021

# Astronomía computacional en el IALP y FCAG

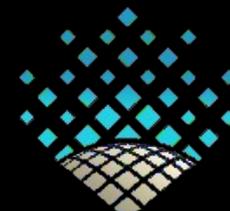
Romina P. Di Sisto  
y colaboración de grupos del  
IALP y FCAG - UNLP



Instituto de Astrofísica de La Plata  
IALP - CONICET - UNLP  
CCT - CONICET - La Plata



CIENCIAS  
PLANETARIAS  
GRUPO ORIGEN



Facultad de Ciencias  
**Astronómicas  
y Geofísicas**  
UNIVERSIDAD NACIONAL DE LA PLATA

# Observatorio Astronómico de La Plata

IALP (CONICET - UNLP)



FCAG - UNLP



**Astronomía**  
**Geofísica**  
**Meteorología**

Varios Institutos y Laboratorios  
(Investigación, transferencia, difusión y docencia)

## **Integrantes del IALP (datos a julio de 2021):**

42 Investigadores (CONICET/CIC)

3 Profesionales (CONICET)

1 Técnico (CONICET)

46 Becarios (CONICET)

7 Docente-investigadores (UNLP)

3 Becarios (UNLP)

1 Administrativo (UNLP)

1 Administrativo (CONICET)

Total: 104

### **FCAG – UNLP**

Planta docente: ~250 docentes o docentes/investigadores

## Grupos

### IALP

|  |       |
|--|-------|
| Agrupaciones estelares, medio interestelar y galaxias activas        | AMIGA |
| Astrofísica de Cúmulos Abiertos                                      | AdeCA |
| Caos en Sistemas Hamiltonianos                                       | CenSH |
| Dinámica de Sistemas Estelares                                       | DdeSE |
| Estrellas Masivas  | EM    |
| Estudios Observacionales de Blazares y Galaxias Peculiares           | EOBGP |
| Evolución Estelar y Pulsaciones                                      | EEyP  |
| Evolución Estelar y Sistemas Binarios                                | EEySB |
| Formación y Evolución de Galaxias                                    | FEG   |
| Grupo de Astrofísica Planetaria                                      | GAP   |
| Grupo de investigación de Estrellas Masivas y Agrupaciones Estelares | GEMAE |
| Grupo Origen - Ciencias Planetarias                                  | GO    |
| Modelos de Estrellas Peculiares                                      | MEP   |
| Sistemas Estelares Extragalácticos y su Contexto Cosmológico         | SEECC |
| Sistemas Estelares: Observaciones y Simulaciones                     | SEOS  |
| Supernova Observations and Simulations                               | SOS   |

### FCAG

Grupo de Gravitación, Astrofísica y Cosmología  
Grupo de Cosmología

....

## Facilidades computacionales en la Facultad

- Granja: Sala con varias computadoras para los alumnos de grado y posgrado
- Clusters de algunos grupos --- obsoletos
- Cluster: Seminare---- apagada--- no se le pued dar soporte-
- Propuesta de algunas Facultades para presentar un Proyecto para hacer un Centro de Cómputos de Alto rendimiento de la UNLP e incorporarlo al sistema de cómputos de alto rendimiento.... Idea: 10 nodos: 640 núcleos, con conectividad moderna o 5 nodos (320 nucleos) dependiendo del presupuesto. La idea es comenzar para luego ir ampliando. Contempla infraestructura para empazarlo y cargos CPA para funcionamiento.

## Facilidades computacionales en el IALP

Servidor: laruidosa

# Facilidades computacionales en el IALP

## Servidor: laruidosa

### System

Host: laruidosa Kernel: 5.0.16-100.fc28.x86\_64 x86\_64 bits: 64  
Distro: Fedora

### Machine

Mobo: Supermicro model: X10DRW-i v: 1.10

### Memory

RAM: total: 251.80 GiB used: 1.27 GiB (0.5%)

### CPU

Topology: 2x 12-Core model: Intel Xeon E5-2650 v4 bits: 64 type: MT MCP SMP  
L2 cache: 60.0 MiB  
Speed: 2904 MHz min/max: 1200/2900 MHz

Los procesadores Intel tienen una tecnología llamada HyperThreading que permite que laruidosa funcione como: 48 cores

Almacenamiento (Home): 18 Tb

# Uso de laruidosa

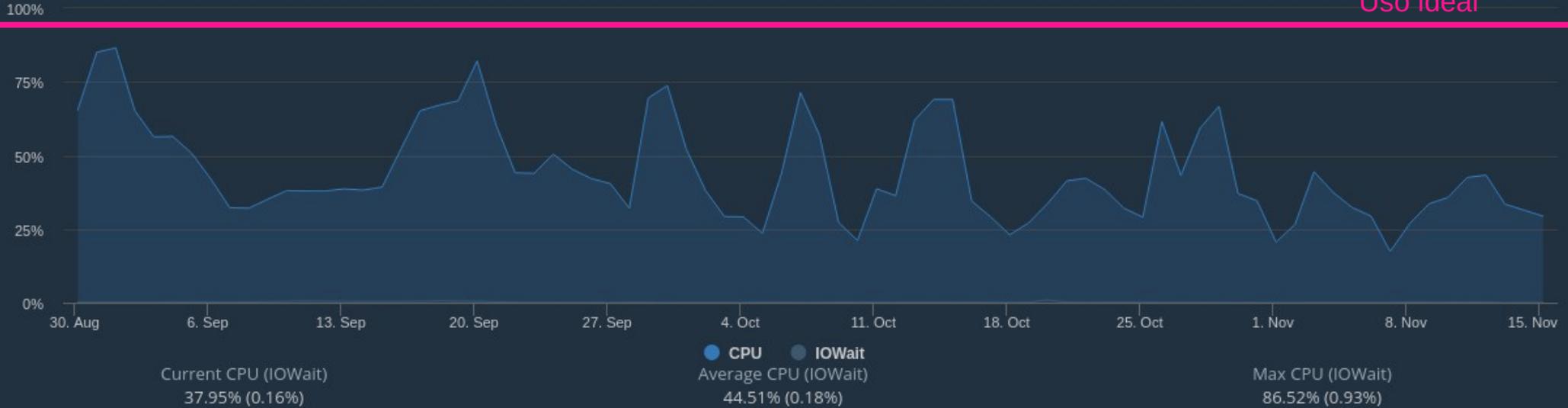
<https://hetrixtools.com/report/uptime/bbb9238188e958faf9dc11a476420865/>

## CPU USAGE

3h 6h 12h 1d 7d 15d 30d 3m 6m 1y

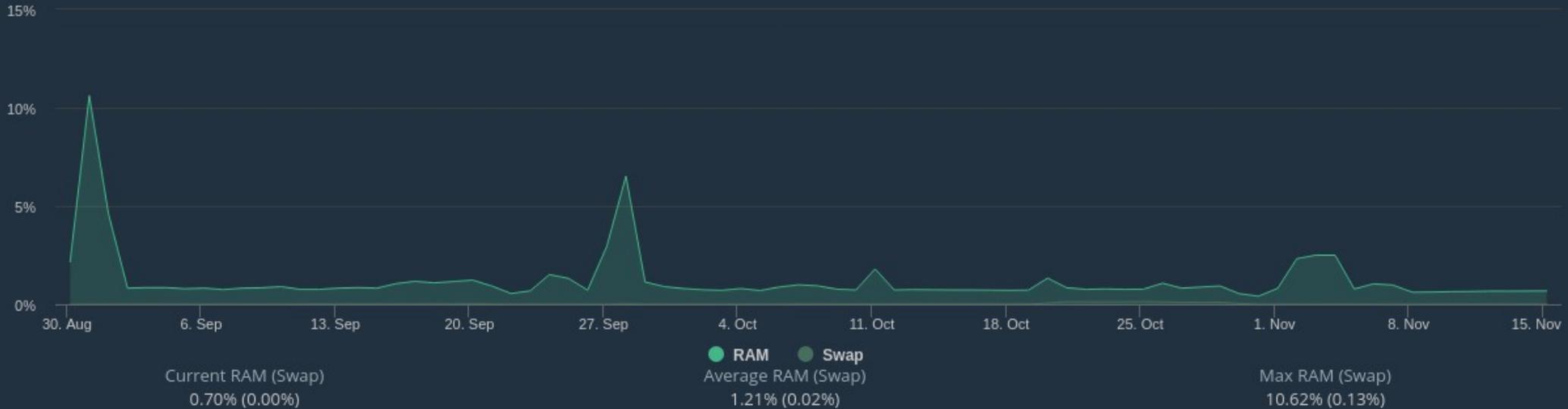


Uso ideal



## RAM USAGE

3h 6h 12h 1d 7d 15d 30d 3m 6m 1y



# DISK USAGE

3h 6h 12h 1d 7d 15d 30d 3m 6m 1y 



Current Disk Usage  
25.15%

 Disk  
Average Disk Usage  
53.05%

Max Disk Usage  
68.40%

| Mount | Size     | Used    | Available | Usage   |
|-------|----------|---------|-----------|---|
| /home | 18.12TiB | 4.68TiB | 13.44TiB  | <div style="width: 25.82%;"><span>25.82%</span></div> |

## Encuesta a los grupos del IALP y FCAG

1- Explícite las necesidades generales computacionales que requiere para su trabajo (RAM, almacenamiento, cómputo, etc), con qué recursos cuenta como grupo, y qué capacidades necesitaría fuera de los recursos con los que cuenta.

2- Si necesita realizar simulaciones numéricas a largo plazo, explícite si lo realiza con recursos propios, con recursos del IALP - FCAG o con recursos de colaboradores externos y porqué. Necesita realizar un número grande de corridas en serie (cientos)?, o realiza simulaciones programadas en paralelo, o ambas?

## **Grupos que no requieren capacidades computacionales extras a lo que tienen (computadoras de escritorio o portátiles actualizadas) o solo mejoras por el momento.**

- Estrellas Masivas (Haguele, Bosch, Fernández Lajús, et al. ) Trabajo mayormente observacional
- Grupo de Gravitación, Astrofísica y Cosmología: (Orsaria et al.) Realizan simulaciones, cientos de corridas, pero no requieren enormes capacidades de cálculo ni RAM o grandes tiempos de cómputo.
- Evolución Estelar y Sistemas Binarios (Benvenuto et al.): Simulaciones de evolución estelar (tiempos cortos y no demasiado almacenamiento), pero esto podría cambiar en el futuro próximo.
- Estudios Observacionales de Blazares y Galaxias Peculiares (Andruchow et al.): Trabajan mayormente con datos observacionales. Requieren:
  - Buena capacidad de memoria que permita almacenar gran número de imágenes.
  - Acceso con muy buena velocidad a internet (bajar imágenes y utilizar bases de datos)... debería mejorarse.
  - Cuentan con computadoras adecuadas pero problemas por los cortes de luz.
  - Podrían necesitar en el futuro capacidad de cómputo para realizar simulaciones.

# Agrupaciones estelares, medio interestelar y galaxias activas (AMIGA)

Baume, Feinstein et al. (Orientación observacional)

## Requerimientos:

- a) Reducción de imágenes para obtener datos fotométricos/astrométricos (eventualmente espectroscópicos o polarimétricos)
- b) Análisis de los datos que surgen de la reducción y procesamiento de las imágenes
- c) Almacenamiento tanto de las imágenes como de los datos reducidos.

- ➔ Volumen de datos es pequeño: computadora de escritorio de altas prestaciones.
- ➔ Volumen de datos grande (relevamientos celestes actuales y los que vendrán) requiere:
  - Alta capacidad de cálculo- Gran cantidad de RAM- Gran capacidad de almacenamiento

"la ruidosa" del IALP satisface de momento estas necesidades, no obstante se pueden pensar las siguientes mejoras:

Opción A:- Ampliación de la capacidad de almacenamiento (varios TeraBytes) - Adición de GPUs (no viable en la ruidosa... armar otro server.

Opción B:- Mejorar la velocidad de la red interna del Observatorio

- Utilizar servicios de "cloud computing". O sea, alquilar espacio y capacidad de cómputo de servicios especializados (Amazon, Google, etc ....)

Esta última alternativa (opción B) tendría la gran ventaja de no necesitar realizar el mantenimiento rutinario de los equipos/software y se evitaria su obsolescencia. De hecho, esta es la tendencia de varios observatorios/proyectos (HST, VVV, LSST, etc...)

# Grupo de Investigación en Sistemas Estelares Extragalácticos y su Contexto Cosmológico

Faifer, Smith Castelli, et al.

Trabajo con imágenes y espectros

Volúmenes relativamente pequeños en el pasado, pero actualmente los instrumentos producen imágenes y espectros que insumen centenares de Mb.

Ejemplo: relevamiento S-Plus: cada campo observado representa 5.5Gb de datos (ya reducidos) pero durante el proceso de reducción se genera información temporaria para lo cual es necesario del orden de los 10Tb.

Capacidad de cálculo y RAM: estos procesos no requieren de una cantidad fuera de lo usual. Esto es, RAM de 32Gb, y procesadores como los de uso extendido en la actualidad (i5-i7) y se realizan simulaciones numéricas y parte del análisis de datos observacionales en la ruidosa .

El grupo cuenta con tres equipos armados con el apoyo del personal de soporte de la FCAG y del IALP. Estos son 2 RAIDS 5, los dos con un volumen de almacenamiento de 14Tb cada uno. El otro equipo es un RAID 6 de 11 Tb de volumen. Cabe destacar que estos equipos cumplen diversas funciones para el grupo, por lo que no están restringidos a una única tarea.

## Necesidades a futuro:

Ampliar capacidades de almacenamiento para reducción y manejo de datos.

La reducción de grandes volúmenes de datos se ve limitada por los equipos que disponemos en la actualidad. Por ello, solo se realizan tareas acotadas, mientras que el grueso del proceso de datos como los del relevamiento S-Plus se hace en equipos fuera del país (Chile y Brasil).

## **Dinámica** (Carpintero, Muzzio)

Necesidades: poder de cómputo, y solo secundariamente RAM y almacenamiento

### Trabajos y recursos disponibles:

- server del IALP (laruidosa), PCs individuales con procesadores multicore para corridas de prueba o secundarias.
- Para simulaciones largas: con recursos del IALP (laruidosa) + recursos de colaboradores externos (IAFE), para la ejecución de dos corridas al mismo tiempo. Corridas en paralelo (una corrida usando N núcleos).
- Cálculos sobre órbitas (exponentes de Lyapunov, mapas de Poincaré, etc.) en corridas múltiples (N núcleos). Requisitos de RAM baja o media y almacenamiento para preservar por un tiempo prudencial los resultados, (2 o 3 TB).

### Recursos necesarios adicionales y mejoras:

- ninguno como para decir que sin ellos se impide el desarrollo de los trabajos.
- Deseable más y mejores procesadores en laruidosa (o un server o cluster nuevo) a efectos de reducir los tiempos de integración, o eventualmente para encarar proyectos más osados.
- Procesadores más veloces ya que se realizan integraciones muy largas.

# Grupo de Astrofísica Planetaria (Guilera, Ronco, Miller Bertolami)

Necesidades: poder de cómputo, RAM media (16–32 Gb) y almacenamiento medio a alto (10Tb)

Necesidad básica: que la energía se corte lo mínimo posible !

## Trabajos y recursos disponibles:

- Equipos actuales: para el cómputo y análisis de las simulaciones: - 2 PCs core i9 (10 cores - 20 hilos), con 32GB de RAM - 2 PCs core i7 (6 cores - 12 hilos), con 16 GB de RAM, cada uno con 2 discos de 2TB y cada integrante tiene PCs de escritorio que utilizamos principalmente como consolas.
- En general las simulaciones a largo plazo (N cuerpos) se realizan con equipos propios, siempre y cuando no impliquen un número alto de diferentes simulaciones (barriendo, por ejemplo, el espacio de parámetros del problema en cuestión).
- Simulaciones que implicaron la realización de un par de miles de simulaciones, solo fueron posibles con el acceso de colaboradores a un cluster en el exterior (Suiza)
- Dependiendo del trabajo se realizan entre decenas/centenas a miles de simulaciones de duración desde horas o días.
- Código principal adaptado para realizar M simulaciones distintas en simultáneo variando los parámetros del modelo, pero si  $M \gg N$  (N = número de hilos libres en un equipo o varios), las simulaciones se corren de a tandas y se estiran los tiempos de cómputo.

## Recursos necesarios adicionales y mejoras:

- Para cómputo: equipos con varias decenas/centenas de cores para poder correr simulaciones en simultáneo y reducir los tiempos de cómputo (estudios de síntesis poblacionales y otros).
- Contar con algún equipo con placas GPU tanto para usar códigos de N cuerpos en GPU (GENGA), o pensar en proyectos que involucren el uso de códigos hidrodinámicos

# Grupo Origen – Ciencias Planetarias (de Elía, Di Sisto et al.)

Necesidades: poder de cómputo, RAM media y almacenamiento medio a alto

Análisis de los Procesos Físicos y Dinámicos asociados a la Formación y Evolución de Planetas alrededor de *Estrellas de Diferentes Tipos Espectrales*

- Propiedades Generales de Planetas de Tipo Terrestre en *Zona de Habitabilidad Circumestelar*
- Evolución Colisional y Dinámica de Reservorios de Cuerpos Menores
- Colisiones y Cráteres en el Sistema Solar

## Simulaciones y Herramientas Numéricas

### - SIMULACIONES NUMÉRICAS DE N-CUERPOS:

Códigos **MERCURY** y **EVORB** (Ampliamente Utilizados en la Bibliografía)

Código **D3** (Desarrollado en el Seno del Grupo Origen de Ciencias Planetarias)

### - SIMULACIONES DE EVOLUCIÓN COLISIONAL Y CRÁTERES

Código **ACDC** (Desarrollado en el Seno del Grupo Origen de Ciencias Planetarias)

## Recursos Computacionales

- 10 PC, que posee entre 8 y 12 núcleos, 16 a 64 Gb de RAM y discos de 1 a 4 Tb .
- Además, utilizamos Laruidosa, anteriormente, Seminare (FCAG) y otros clústers propios del Grupo de Investigación, que hoy están fuera de servicio.
- Equipos de Cómputo proporcionados por Colaboradores de otros Institutos.

## Principales Limitaciones y Condicionamientos

### - Número de Cuerpos involucrados en las Simulaciones de N-Cuerpos:

El número de cuerpos incrementa el tiempo de CPU. El tiempo de corrida depende del paso de integración. Las simulaciones pueden requerir entre 1 y 4 meses.

### - Distribución Espacial de los Cuerpos Simulados:

Paso de integración ~ fracción del período orbital del cuerpo más interno. Cuerpos más internos, pasos de integración menores, mayor tiempo de CPU requerido.

- **Procesos Físicos Involucrados y su Modelado:** ejemplo: simulaciones de N-cuerpos que incluyen fragmentación de los cuerpos que impactan. La incorporación de un elevado número de fragmentos luego de cada colisión incrementa de manera excesiva el tiempo de CPU requerido. Buscamos un equilibrio entre el modelado del proceso físico y el tiempo de cómputo.

**Número de Simulaciones de N-Cuerpos:** la acreción de cuerpos en la formación y evolución de un sistema planetario resulta ser un proceso de naturaleza estocástica. Se requiere un elevado número de experimentos numéricos, los cuales luego deben ser analizados estadísticamente para entender tendencias generales. Se desarrollan ~50 simulaciones de N-Cuerpos por escenario de trabajo. Más recursos computacionales nos permitiría acceder a una mejor interpretación de los resultados y elaborar una estadística más robusta.

**Recursos necesarios adicionales y mejoras: idem GAP**

## Formación y Evolución de Galaxias (Cora et al. )

La investigación se basa en el uso de simulaciones cosmológicas de gran envergadura, tanto en volumen simulado como en resolución de masa. Consisten en simulaciones de materia oscura acopladas con un modelo semi-analítico de formación de galaxias, y simulaciones hidrodinámicas autoconsistentes.

- Se requieren los tres tipos de recursos: cómputo, almacenamiento y memoria.

Actualmente, según las actividades que realiza el grupo, cobran más relevancia la **capacidad de alocar memoria RAM para el análisis de grandes bases de datos (salidas de simulaciones hidrodinámicas generadas por otros grupos)**, y **memoria de disco para su almacenamiento**. Sin embargo, una buena capacidad de cómputo también acelera el proceso de calibración del código semi-analítico usado y la generación de catálogos de galaxias a partir de su aplicación a simulaciones de materia oscura provista por otros grupos de investigación.

- Actualmente contamos con un servidor HP proliant DL360 Gen9, con dos procesadores Xeon de 8 núcleos (16 threads) cada uno y 128 GB de memoria RAM (con slots para sumar hasta ~ 700 GB) . El Servidor cuenta con 3 Tb de almacenamiento, pero usamos una computadora externa como almacenamiento extra a la que el servidor accede por red (no es óptimo).

### Mejoras necesarias:

- Mejorar las capacidades actuales -----> dispositivo estilo **NAS para almacenamiento**, en lugar del servidor, y con placas de red optimizadas para transferir información. Además, para aprovechar la capacidad del servidor y no tener que limitar el número de galaxias simuladas en nuestras corridas, vendría bien **completar la capacidad de memoria RAM del servidor**.

- Para ampliar las capacidades del grupo y poder correr simulaciones numéricas cosmológicas de materia oscura pura o hidrodinámicas competitivas, necesitaríamos **acceso a un cluster con grandes capacidades de cómputo paralelo**.

# Formación y Evolución de Galaxias (Cora et al. )

Los proyectos realizados y en curso fueron y son posibles gracias a la colaboración con grupos de investigación en el extranjero que posibilitaron el acceso a sus recursos computacionales :

- Australia: OzStar en la Universidad de Swinburne: <https://supercomputing.swin.edu.au/>
- España/Inglaterra: se integra la colaboración The Three Hundred Project ( <https://www.nottingham.ac.uk/~ppzfrp/The300/index.php> ), y se tiene acceso a tiempo de cómputo en un servidor remoto (Taurus) que tiene 130 TB de memoria, 256 núcleos y 1 TB de RAM .
- Acceso a las simulaciones hidrodinámicas IllustrisTNG. Datos públicos con acceso por una plataforma (colaboración Jupyter Lab) y a partir de una cuenta solicitada por el Max Planck Partner group de la Universidad de La Serena.

## **Equipos utilizados para trabajos recientes y necesarios para continuar con el trabajo:**

Información extraída de la Tesis Doctoral del Dr. Cristian Vega-Martínez, quien desarrolló su trabajo haciendo uso del centro de cómputo de la Universidad Católica de Chile (Geryon y Geryon2), generando catálogos de galaxias que sirvieron de base para muchos trabajos realizados desde 2017 a la fecha. La generación de esas salidas en su versión más simplificada (sin generar espectros sintéticos para cada galaxia) tardó aproximadamente un mes. Ese cluster fue reemplazado por otro mejor al cual ya no se tiene acceso.

- Geryon. Consistente en un arreglo de 64 nodos con un total de 128 procesadores de tipo Intel Xeon Quad-Core (512 núcleos) de la familia E5405 de 2.00 GHz, alcanzando 1.0 Terabyte de memoria RAM en total. Los nodos usan sistemas Linux con arquitectura de 64-bit y se comunican entre sí mediante una red tipo Ethernet con velocidad de hasta 10 Gbit/s.

- Geryon2. Consistente en un arreglo de 15 nodos con un total de 584 procesadores de tipo Intel Xeon de la familia E7, modelos 4820 o 4850 dependiendo del nodo, alcanzando 5.47 Terabytes de memoria RAM en total. Al igual que el anterior, los nodos usan sistemas Linux con arquitectura de 64-bit, pero están conectados mediante una red Infiniband con una tasa de transferencia de datos de hasta 50 Gbit/s.

Esta supercomputadora cuenta además con un arreglo de discos de estado sólido SSD de alta velocidad de 58 Terabytes, los cuales están destinados específicamente a minimizar las latencias producidas por los cálculos que tienen una alta demanda de lectura y/o escritura de datos.

Adicionalmente, ambos recursos comparten una capacidad de almacenamiento en discos duros de hasta 250 Terabytes.

## Conclusiones:

Básico:

- que no se corte la luz...
- organización del uso para su optimización, cuotas, mantenimiento, actualización del equipo
- infraestructura para la instalación y funcionamiento de los equipos.

### 1. Grupos asociados a grandes surveys observacionales:

Por el momento, trabajan con equipos altas prestaciones (i5-i7) y RAIDS para almacenamiento.

Pero:

Volumen de datos grande (relevamientos celestes actuales y los que vendrán) requiere:

- Alta capacidad de cálculo- Gran cantidad de RAM- Gran capacidad de almacenamiento y/o
- Mejorar la velocidad de la red de internet para utilizar servicios de "cloud computing".

### 2. Grupos que requieren simulaciones numéricas necesidades medias:

- Equipo con varias decenas/centenas de cores para poder correr simulaciones en simultáneo y reducir los tiempos de cómputo o muchas simulaciones en serie a largo plazo o utilizar varios cores para simulaciones en paralelo.
- Equipo con placas GPU (para códigos de N cuerpos o códigos hidrodinámicos)
- RAM: necesidades medias
- Almacenamiento: alto: > 100 Tb para todos los grupos.

### 3. Grupos que requieren simulaciones numéricas necesidades altas:

- Cluster de alto rendimiento con grandes capacidades de cómputo paralelo.
- Equipos con centenares de cores
- Varios Tb de RAM
- Almacenamiento: alto: discos de estado sólido SSD de alta velocidad de decenas de Tb, y discos rígidos > 200 Tb.
- Red o sistema de alta velocidad para comunicación y transferencia de datos entre nodos: decenas de Gbit/s

.....

Necesidad común:

**Equipo para respaldo y almacenamiento de datos y corridas**

Muchas gracias

**IT CAN ONLY BE ATTRIBUTABLE**

**TO**

**HUMAN ERROR**