Recursos disponibles de Computación de Alto Desempeño

Infraestructura en FCEN y CSC

Centro de Simulación Computacional p/Aplic Tecnológicas
CONICET
Departmento de Computación
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires

23/08/2016

How did I get here?

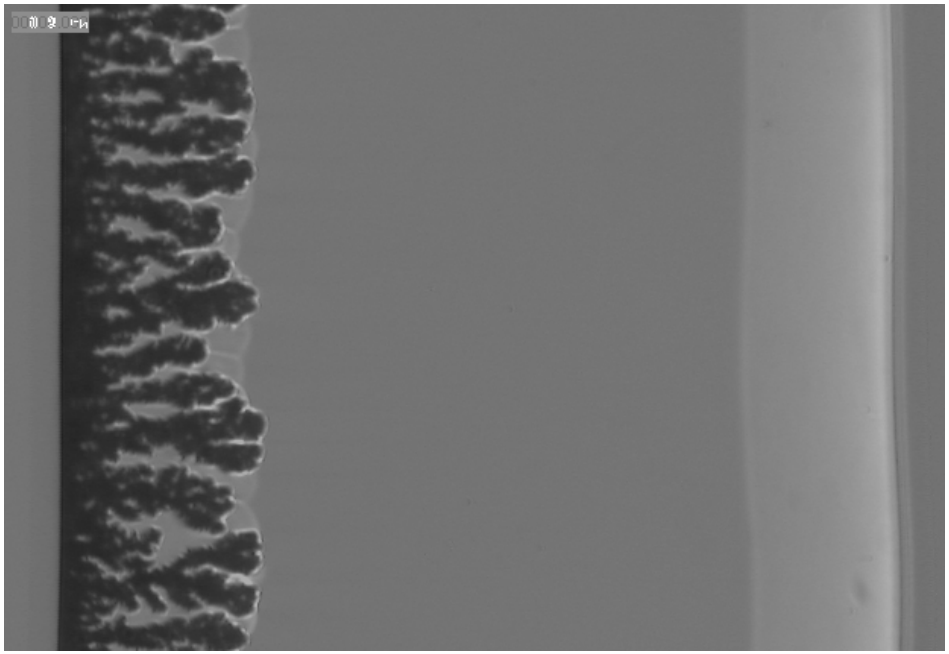
- Graduate Studies: Computer Science. The title is called "Computer Science Licentiate (*Licenciado en ciencias de la computación*). Six years long career including final thesis. Most of the students work and study at the same time.

How did I get here?

- Graduate Studies: Computer Science. The title is called "Computer Science Licentiate (*Licenciado en ciencias de la computación*). Six years long career including final thesis. Most of the students work and study at the same time.
- PhD: also in Computer Science, but focused on visualization, modeling and simulation of Electrochemical Deposition.

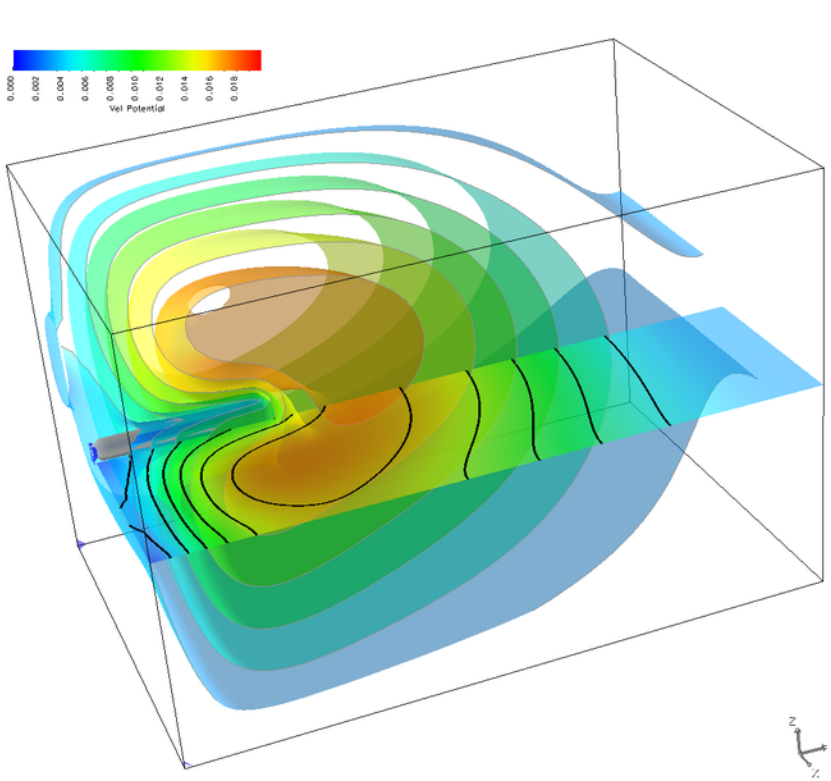
How did I get here?

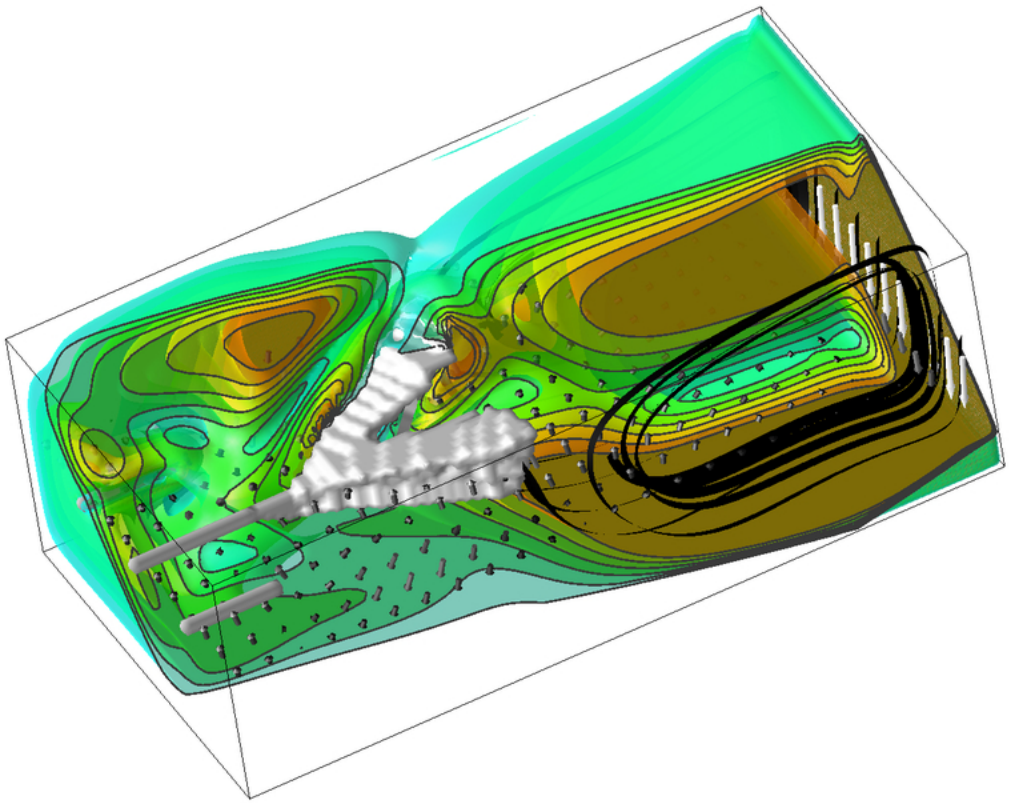
- Graduate Studies: Computer Science. The title is called "Computer Science Licentiate (*Licenciado en ciencias de la computación*). Six years long career including final thesis. Most of the students work and study at the same time.
- PhD: also in Computer Science, but focused on visualization, modeling and simulation of Electrochemical Deposition.



How did I get here?

- Graduate Studies: Computer Science. The title is called "Computer Science Licentiate (*Licenciado en ciencias de la computación*). Six years long career including final thesis. Most of the students work and study at the same time.
- PhD: also in Computer Science, but focused on visualization, modeling and simulation of Electrochemical Deposition.



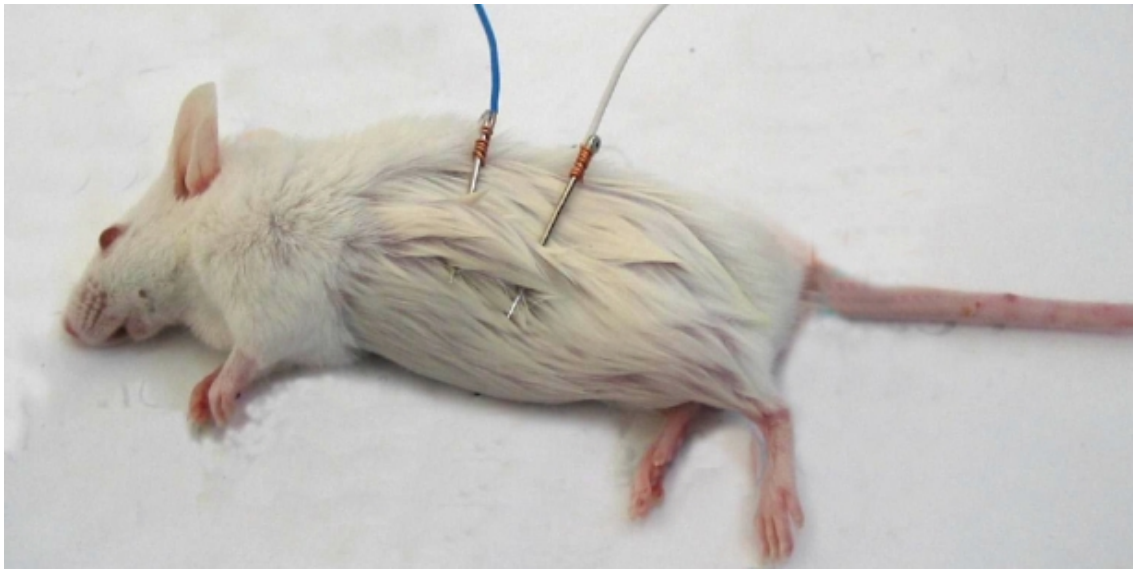


But not only this

Dr. Guillermo Marshall (my PhD tutor) forced his students to start learning Biology, because *"it is the future"*.

But not only this

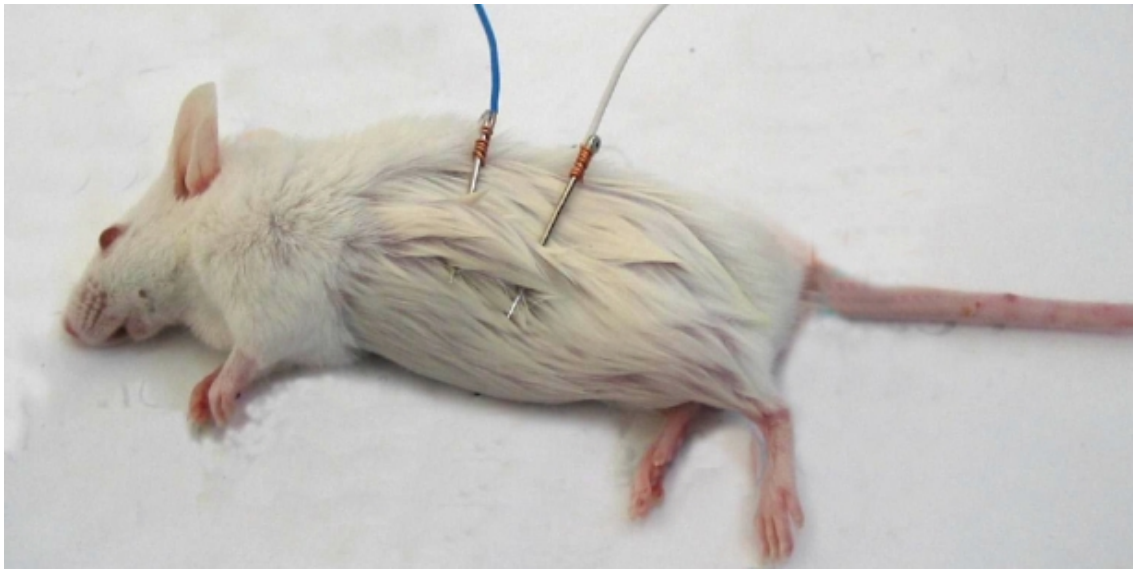
Dr. Guillermo Marshall (my PhD tutor) forced his students to start learning Biology, because *"it is the future"*.



- We finished working at one of the hospitals dependent of UBA (Roffo Hospital, specialized in Oncology).

But not only this

Dr. Guillermo Marshall (my PhD tutor) forced his students to start learning Biology, because *"it is the future"*.



- We finished working at one of the hospitals dependent of UBA (Roffo Hospital, specialized in Oncology).
- We started trying to apply the same ideas behind Electrochemical Deposition as a therapy against cancer.
- Later, also worked in a real Biology wet laboratory Molecular Biology!

HPC... Why? How?

While I was doing my PhD, I needed to use a cluster...

HPC... Why? How?

While I was doing my PhD, I needed to use a cluster...

First PC based cluster

Speedy Gonzalez

- Finished administering the cluster for all the users because I was the only one who could (i.e. was forced to) do it.
- And as I needed, I had to learn a lot about cluster administration.

HPC... Why? How?

While I was doing my PhD, I needed to use a cluster...

First PC based cluster

Speedy Gonzalez

- Finished administering the cluster for all the users because I was the only one who could (i.e. was forced to) do it.
- And as I needed, I had to learn a lot about cluster administration.
- Then, I had to teach the next guys how to do it.

HPC... Why? How?

While I was doing my PhD, I needed to use a cluster...

First PC based cluster

Speedy Gonzalez

- Finished administering the cluster for all the users because I was the only one who could (i.e. was forced to) do it.
- And as I needed, I had to learn a lot about cluster administration.
- Then, I had to teach the next guys how to do it.
- So I became like a reference in HPC, mainly because I know how the users think and how the admin guys *react* to user requests.

CeCAR: shared HPC facility at FCEN-UBA

When a group of professors and researchers from FCEN-UBA thought about establishing a shared HPC facility, I finished as the technical person-in-charge.

CeCAR: shared HPC facility at FCEN-UBA

When a group of professors and researchers from FCEN-UBA thought about establishing a shared HPC facility, I finished as the technical person-in-charge.

2006: First Infiniband based cluster, in that time, no clear info was available...

CeCAR: shared HPC facility at FCEN-UBA

When a group of professors and researchers from FCEN-UBA thought about establishing a shared HPC facility, I finished as the technical person-in-charge.

2006: First Infiniband based cluster, in that time, no clear info was available... Besides, Mellanox has its main plant in Israel... Who remember what happened in 2006?

CeCAR: shared HPC facility at FCEN-UBA

CeCAR

When a group of professors and researchers from FCEN-UBA thought about establishing a shared HPC facility, I finished as the technical person-in-charge.

2006: First Infiniband based cluster, in that time, no clear info was available... Besides, Mellanox has its main plant in Israel... Who remember what happened in 2006? **Operation "Summer Rains"**!!!!

CeCAR

The equipment in this facility was recently upgraded to:

- 18 CPU nodes, 2 AMD Opteron 6320 each (8 cores per chip), 64GB RAM and 300GB of local user space (mounted at /tmp).
- 12 GPU nodes, same as above but with 2 NVidia Tesla K20c per node.

CeCAR

The equipment in this facility was recently upgraded to:

- 18 CPU nodes, 2 AMD Opteron 6320 each (8 cores per chip), 64GB RAM and 300GB of local user space (mounted at /tmp).
- 12 GPU nodes, same as above but with 2 NVidia Tesla K20c per node.
- Headnode, similar to nodes, but with 256GB RAM.
- 2 extra memory nodes (planned for genome assembly): 512GB RAM, 4 AMD Opteron 6320 each.

CeCAR

The equipment in this facility was recently upgraded to:

- 18 CPU nodes, 2 AMD Opteron 6320 each (8 cores per chip), 64GB RAM and 300GB of local user space (mounted at /tmp).
- 12 GPU nodes, same as above but with 2 NVidia Tesla K20c per node.
- Headnode, similar to nodes, but with 256GB RAM.
- 2 extra memory nodes (planned for genome assembly): 512GB RAM, 4 AMD Opteron 6320 each.
- Xyratex ClusterStor 6000 with 120TB installed capacity mounted at /home using Lustre.
- Infiniband interconnect (flat tree topology), Mellanox.
- Technical staff: part-time student (more or less 4 hours a day).

CeCAR: Who can use it?

- CeCAR is open for anyone belonging to the scientific community.

CeCAR: Who can use it?

- CeCAR is open for anyone belonging to the scientific community.
- Only need to fill an online form with data related to the application to be used and the problem to be solved:
  `http://cecar.fcen.uba.ar/solicitud/`

CeCAR: Who can use it?

- CeCAR is open for anyone belonging to the scientific community.
- Only need to fill an online form with data related to the application to be used and the problem to be solved:
  `http://cecar.fcen.uba.ar/solicitud/`
- Two main types of usage: free and paid.
  - **Free**: any user has an amount of CPU hours in each month. After using those hours, can still use all the idle equipment.
  - **Paid**: you can paid to be a prioritized user. Not necessarily cash, it can be equipment (or other ways that can be established!). You get prioritized access to resources.

CeCAR: Who can use it?

- CeCAR is open for anyone belonging to the scientific community.
- Only need to fill an online form with data related to the application to be used and the problem to be solved:
  `http://cecar.fcen.uba.ar/solicitud/`
- Two main types of usage: free and paid.
  - **Free**: any user has an amount of CPU hours in each month. After using those hours, can still use all the idle equipment.
  - **Paid**: you can paid to be a prioritized user. Not necessarily cash, it can be equipment (or other ways that can be established!). You get prioritized access to resources.
- Services provided by CeCAR:
  - HPC in `x86` cores.
  - HPC in `GPU` boards.
  - Cloud
  - Storage

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: CentOS 6.8 (Carbon).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: CentOS 6.8 (Carbon).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.
- 18 nodos CPU:
  - 2 AMD Opteron 6320 (8 cores cada uno)
  - 64GB de RAM
  - 500GB de rígido montado localmente.
  - /tmp tiene espacio disponible de scratch.

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: CentOS 6.8 (Carbon).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.
- 18 nodos CPU:
  - 2 AMD Opteron 6320 (8 cores cada uno)
  - 64GB de RAM
  - 500GB de rígido montado localmente.
  - /tmp tiene espacio disponible de scratch.
- 12 nodos **GPU**. Idem anterior, con 2 (dos) placas K20c cada uno.

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: CentOS 6.8 (Carbon).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.
- 18 nodos CPU:
  - 2 AMD Opteron 6320 (8 cores cada uno)
  - 64GB de RAM
  - 500GB de rígido montado localmente.
  - /tmp tiene espacio disponible de scratch.
- 12 nodos **GPU**. Idem anterior, con 2 (dos) placas K20c cada uno.
- 2 nodos de *ensamble*:
  - 4 AMD Opteron 6320 (8 cores cada uno)
  - 512GB de RAM

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: CentOS 6.8 (Carbon).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.
- 18 nodos CPU:
  - 2 AMD Opteron 6320 (8 cores cada uno)
  - 64GB de RAM
  - 500GB de rígido montado localmente.
  - /tmp tiene espacio disponible de scratch.
- 12 nodos **GPU**. Idem anterior, con 2 (dos) placas K20c cada uno.
- 2 nodos de *ensamble*:
  - 4 AMD Opteron 6320 (8 cores cada uno)
  - 512GB de RAM
- Storage:
  - Xyratex ClusterStor 6000 (soporte Infiniband nativo).
  - 120TB para uso compartido, montado en /home.

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: CentOS 6.8 (Carbon).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.
- 18 nodos CPU:
  - 2 AMD Opteron 6320 (8 cores cada uno)
  - 64GB de RAM
  - 500GB de rígido montado localmente.
  - /tmp tiene espacio disponible de scratch.
- 12 nodos **GPU**. Idem anterior, con 2 (dos) placas K20c cada uno.
- 2 nodos de *ensamble*:
  - 4 AMD Opteron 6320 (8 cores cada uno)
  - 512GB de RAM
- Storage:
  - Xyratex ClusterStor 6000 (soporte Infiniband nativo).
  - 120TB para uso compartido, montado en /home.
- Tareas de soporte organizadas por medio de un sistema de tickets.
- Acceso **sin** password, se pide clave pública al registrarse como usuario.

1 Background

2 CeCAR

3 CSC-CONICET

Center for Computational Simulation for Technological Applications (CSC)

One of the newest research centers is the Center for Computational Simulation for Technological Applications (CSC).

The main lines of work are:

- High Performance Computing
- Computacional Fluid Dynamics
- High Performance Computing
- Wireless Communications

Center for Computational Simulation for Technological Applications (CSC)

One of the newest research centers is the Center for Computational Simulation for Technological Applications (CSC).

The main lines of work are:

- High Performance Computing
- Computacional Fluid Dynamics
- High Performance Computing
- Wireless Communications

Goals:

- CSC is a center devoted to create technology.
- This technology should be conceived in the context of a *real* problem.
- Papers are always good, but something more is needed.

Center for Computational Simulation for Technological Applications (CSC)

One of the newest research centers is the Center for Computational Simulation for Technological Applications (CSC).

The main lines of work are:

- High Performance Computing
- Computacional Fluid Dynamics
- High Performance Computing
- Wireless Communications

Goals:

- CSC is a center devoted to create technology.
- This technology should be conceived in the context of a *real* problem.
- Papers are always good, but something more is needed.
- Key: adding value to products created by Argentina Industry.
- Problems are presented by Industry, solutions are proposed by CSC.
- Ideally: associate to partners sharing this view of world, trying to enforce this win-win situation.

Computational Power

A new cluster: TUPAC.

Tupac Amaru II: Last direct descendant of Inca royal blood. He raised against Spanish authorities during colonial time. He was captured and condemned to have his tongue cut out, after watching the execution of his family, and *to have his hands and feet tied to four horses who will then be driven at once toward the four corners of the plaza*.

Computational Power

A new cluster: TUPAC.

Tupac Amaru II: Last direct descendant of Inca royal blood. He raised against Spanish authorities during colonial time. He was captured and condemned to have his tongue cut out, after watching the execution of his family, and *to have his hands and feet tied to four horses who will then be driven at once toward the four corners of the plaza*.

TUPAC

- 4200 AMD Opteron 6276 cores.
- 32 nVidia Tesla GPU.
- 18 TB of DDR3 RAM.
- 72 TB of storage.
- Backup library.
- UPS (critical components)
- QDR Infiniband.
- Separate administrative ethernet networks.
- Near 48TFLOPS.

Heterogeneous Computing

- 8 nodes connected to GPU boards for computing.
- Each node can address four boards (the application should allow it).
- The connection between the boards and the nodes is proprietary from Dell.

Network

- The standard HPC applications usually need low latency and high bandwidth.
- 40Gbps Infiniband 4xQDR for the compute nodes with **less than 5μs** of latency (Grid Director 4036).
- There is an additional administration network (all the components can be independently managed).
- The storage has its own fiber-based network.

Storage Details

- Separated storage connected to 10Gb Ethernet.
- 72TB raw capacity.
- It's assembled using 120 hard disks (15Krpm), 600GB each.
- Its purpose is storing user data, but also administrative data (logs, system configuration, etc.).
- Allows an easy upgrade to support 200 disks.

Backup system

- It includes a backup library.
- Directly connected to 10Gb Ethernet network, does not disturb low latency network.
- The library can backup **72 TB** in 24 available slots.
- Bacula backup system is implemented using a mix full/incremental backup policies to ensure protect user and system data.

Assembling view

A nicer view

TUPAC: Who can use it?

- TUPAC is open for anyone belonging to the scientific or technological community.

TUPAC: Who can use it?

- TUPAC is open for anyone belonging to the scientific or technological community.
- Only need to fill an online form with data related to the application to be used and the problem to be solved:
  `http://tupac.conicet.gov.ar/formulario/`

TUPAC: Who can use it?

- TUPAC is open for anyone belonging to the scientific or technological community.
- Only need to fill an online form with data related to the application to be used and the problem to be solved:
  `http://tupac.conicet.gov.ar/formulario/`
- Two main types of usage: free and paid.
  - **Free**: any user has an amount of CPU hours in each month. After using those hours, can still use all the idle equipment.
  - **Paid**: you can paid to be a prioritized user. Not necessarily cash, it can be equipment (or other ways that can be established!). You get prioritized access to resources.

TUPAC: Who can use it?

- TUPAC is open for anyone belonging to the scientific or technological community.
- Only need to fill an online form with data related to the application to be used and the problem to be solved:
  `http://tupac.conicet.gov.ar/formulario/`
- Two main types of usage: free and paid.
  - **Free**: any user has an amount of CPU hours in each month. After using those hours, can still use all the idle equipment.
  - **Paid**: you can paid to be a prioritized user. Not necessarily cash, it can be equipment (or other ways that can be established!). You get prioritized access to resources.

Focus

The main objective of TUPAC is supporting the development of technological applications, the rest is served as *best effort*.

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: RedHat 6.7 (Santiago).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: RedHat 6.7 (Santiago).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.
- 60 nodos CPU:
  - 4 AMD Opteron 6276 (16 cores cada uno)
  - 128GB de RAM
  - 300GB de rígido montado localmente.
  - /tmp tiene espacio disponible de scratch.

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: RedHat 6.7 (Santiago).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.
- 60 nodos CPU:
  - 4 AMD Opteron 6276 (16 cores cada uno)
  - 128GB de RAM
  - 300GB de rígido montado localmente.
  - /tmp tiene espacio disponible de scratch.
- 8 nodos GPU. Cada uno tiene acceso a 4 placas 2050 cada uno (por ahora, en el futuro, distribuiremos más las placas).

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: RedHat 6.7 (Santiago).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.
- 60 nodos CPU:
  - 4 AMD Opteron 6276 (16 cores cada uno)
  - 128GB de RAM
  - 300GB de rígido montado localmente.
  - /tmp tiene espacio disponible de scratch.
- 8 nodos GPU. Cada uno tiene acceso a 4 placas 2050 cada uno (por ahora, en el futuro, distribuiremos más las placas).
- Storage:
  - Límite de 15TB por unidad lógica.
  - Interconectado con los nodos usando dos servidores NFS
  - Internamente interconectado con fibra.

Technical Specs

- Red InfiniBand de baja latencia
- Sistema Operativo: RedHat 6.7 (Santiago).
- OpenMPI y bibliotecas de funciones standard.
- SLURM como administrador de recursos.
- 60 nodos CPU:
  - 4 AMD Opteron 6276 (16 cores cada uno)
  - 128GB de RAM
  - 300GB de rígido montado localmente.
  - /tmp tiene espacio disponible de scratch.
- 8 nodos GPU. Cada uno tiene acceso a 4 placas 2050 cada uno (por ahora, en el futuro, distribuiremos más las placas).
- Storage:
  - Límite de 15TB por unidad lógica.
  - Interconectado con los nodos usando dos servidores NFS
  - Internamente interconectado con fibra.
- Tareas de soporte organizadas por medio de un sistema de tickets.

I know, too fast, too much...

Thanks!