# New computing infrastructure

Séminar 15/05/2023

Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

# A brief introduction

**Matthias Zytnicki**

# Hardware presentation

**Didier Laborie**

# Software utilization

**Marie-Stéphane Trotard**

# Web services

**Patrice Dehais**

Plateforme Bioinfo Occitanie Toulouse

# About us

### Scientific animation

Matthias Zytnicki
(50% FTE)

### Technical animation

Claire Hoede
(100% FTE)

### System administrators

Didier Laborie
(100% FTE)

Marie-Stéphane Trotard
(80% FTE)

Patrice Dehais
(50% FTE)

### Development and data analysis

Christine Gaspin
(30% FTE)

Christophe Klopp
(30% FTE)

Jérôme Mariette
(70% FTE)

Céline Noirot
(80% FTE)

Philippe Bordron
(100% FTE)

Hanae Chouali
(100% FTE)

### Statistical expertise and data analysis

Nathalie Vialaneix (25% FTE)

### Web Development

Julien Touchais (100% FTE)

# Team changes

- The management staff has changed (2023/01/01)
  - operational management: **Claire Hoede**
  - scientific management: **Matthias Zytnicki**

- New arrival:
  - a new IE for Web development: **Philippe Bordron**

- To be hired
  - a **biostatistics** IR: permanent position
  - a **devops** IR: temporarily position (3 years)

L'EUROPE S'ENGAGE L'OCCITANIE AGIT
UNION EUROPÉENNE
La Région Occitanie Pyrénées · Méditerranée

Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

# User support

- # Active users: **1062**
  - INRAE: 439
  - OCCITANIE (not INRAE): 550
  - OTHER: 73
- # Software: > **1000**
- % Availability: **99,5%**
- # Tickets: **1773** (2022)
- # Current projects: **27** (expertise and development)
- # Annual training sessions: **12**

L'EUROPE S'ENGAGE L'OCCITANIE AGIT
UNION EUROPÉENNE
La Région Occitanie Pyrénées · Méditerranée

Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

# Why limited free usage?

*"I can store my data on a hard drive for much less"*

**What we do not charge** academic users for (except. for private companies):

- CPU time
- 1TB work storage + 250GB save storage + 250GB backup

**What we charge for**:

- Storage extension
- VM (Virtual Machine)

- Most acquisitions are paid by projects (OccalDat, Mudis4LS).
- We pay for the rent (electricity costs are increasing).

L'EUROPE S'ENGAGE L'OCCITANIE AGIT
UNION EUROPÉENNE
La Région Occitanie Pyrénées · Méditerranée

Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

# Rules of use

## User account

- Open to the bioinformatics academic community and private companies
- Gives access to:
  - bioinformatics software
  - genomic and post-genomic databases
  - storage facilities (Home, Save, Work)
  - computing cluster (via a login server)
  - support (FAQ, tickets, training)

## Beware

- Password expires 12 months after last password change
- Account is closed after 12 months of inactivity
- Data is purged 3 months after account closure date

# Rules of use (continued)

**User disk spaces**

- **Home** (10GB): configuration files only
- **Save** (250GB): 30-day history + replication to remote site
- **Work** (1TB): 7-day history (only), NO replication

**Computing hours**

- **100,000 h** annual initial quota with possibility of extension
- 3 user groups (contributors, **INRAE / OCCITANIE**, others)

**Exceptional requests:**

- Rental of additional disk space (projects)
- Rental of virtual machines (website, database...)
- Access to **GPU** or **VIEW** resources
- Extension of computing hours
- Access to training accounts

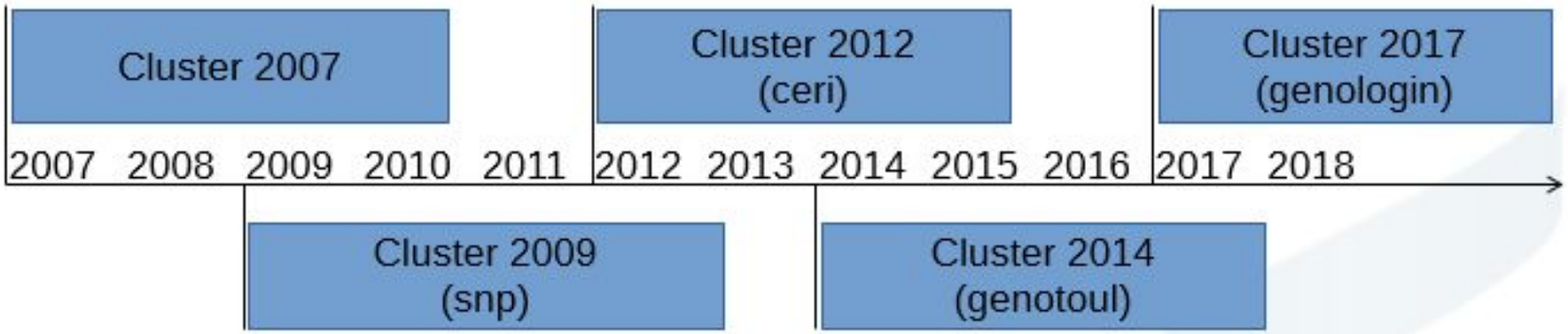
Genotoul Bioinfo

# Hardware presentation

Plateforme Bioinfo Occitanie Toulouse

# New computing cluster

The Bioinfo GenoToul platform has just acquired a new computing infrastructure to replace its obsolete equipment.




L'EUROPE S'ENGAGE L'OCCITANIE AGIT
UNION EUROPÉENNE
La Région Occitanie Pyrénées · Méditerranée

Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

# Timeline of the acquisition

The **timetable** was very **tight** given the shortage of materials, the accommodation at the DROCC and the deadline set by the **REGION** (**FEDER** funding)

- January 2022: drafting of the **specifications**
- February: setting up of the **call for tender procedure**
- March: **publication** of the call for tenders in the **BOAMP** and the **OJEU**
- April: r**eceipt of offers** from suppliers (4): AXIANS, 2CRSI, ATOS and HPE
- May: selection of the successful bidder: **AXIANS** integrator with **LENOVO** equipment
- June: notification of the **public contract**
- October: **delivery** of equipment and start of **installation**
- January 2023: start of INRAE **validation period**

L'EUROPE S'ENGAGE L'OCCITANIE AG|T
UNION EUROPÉENNE
La Région Occitanie Pyrénées · Méditerranée
Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

# Migration strategy

**2023 (transition period)**

- From now, all new user accounts will be created on the new cluster
- We plan to migrate all users (data, projects, VMs, etc.) by the end of this year
- ~20 projects will be migrated per month (schedule to come)
- We are going to shut down several old cluster nodes every month

| | **Old cluster (2014, 2017)** | **New cluster (2023)** |
|---|---|---|
| **#NODES** | 78 | 39 |
| **#CORES** | 4424 | 4928 |
| **RAM per core (average)** | 6GB | **16GB** |
| **Specific nodes** | 1 view node | 1 view node + 1 **GPU** node |

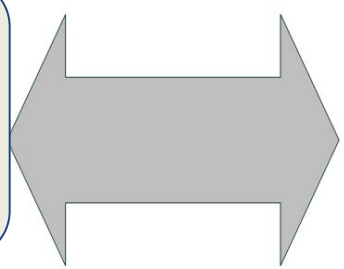
L'EUROPE S'ENGAGE L'OCCITANIE AGIT
UNION EUROPÉENNE
La Région Occitanie Pyrénées · Méditerranée

Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

# New computing cluster

VISU
- Authentification SSH
- Bureau à distance
- HTTPS (OOD)

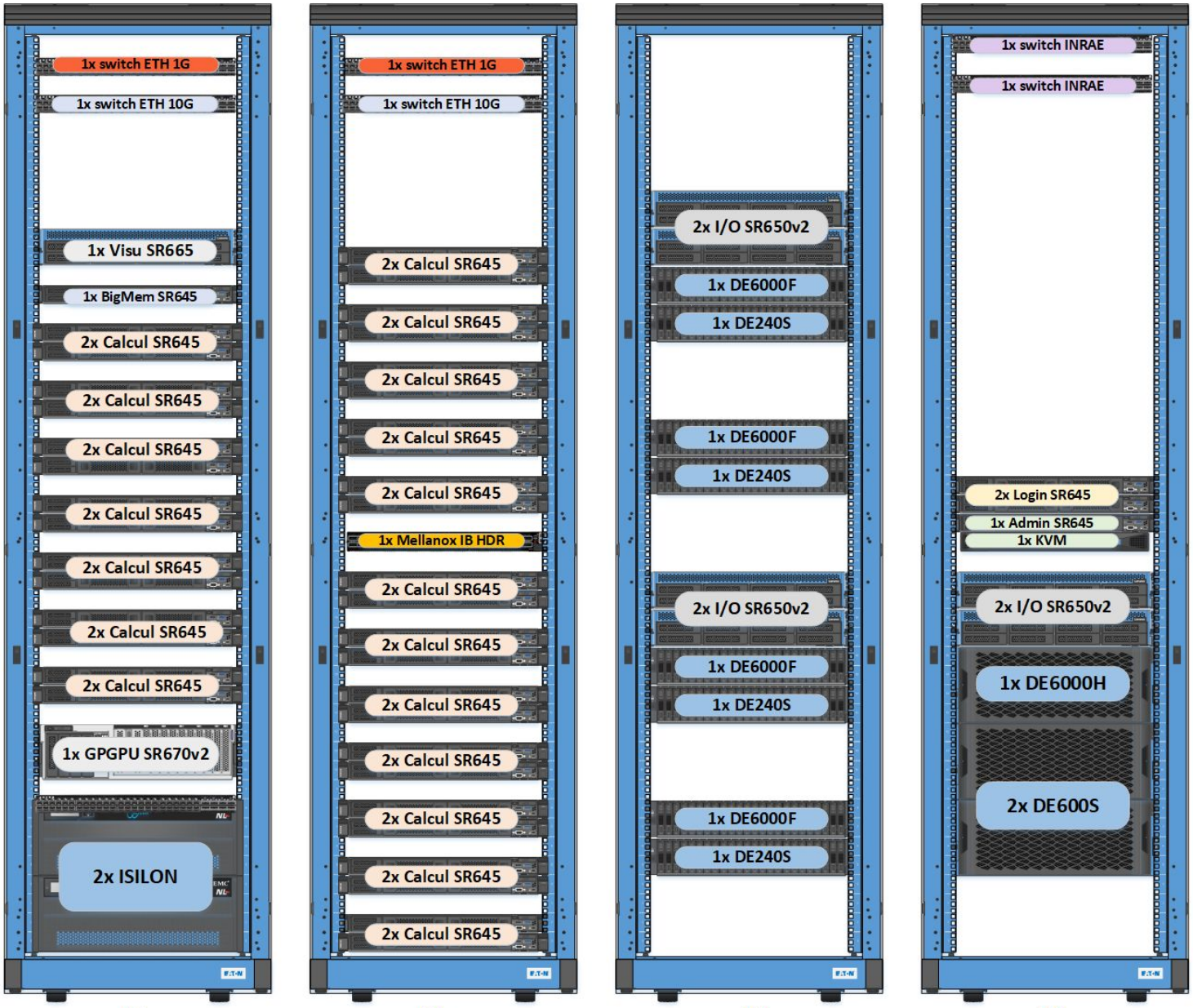
LOGIN (x2)
- Authentification SSH
- Transfert de fichiers
- Compilation
- Soumission de jobs



Genotoul Bioinfo

# ECA hosting

# Physical location



# Some vocabulary

**Cluster:** all compute nodes

**Node:** a server dedicated to high performance computing (HPC)

**Processor:** physical computing units

**Core**: logical computing unit sharing cache memory

**Thread**: number of running processes
**HT**: Hyper Threading (INTEL)
**SMT**: Simultaneous Multi-Threading (AMD)

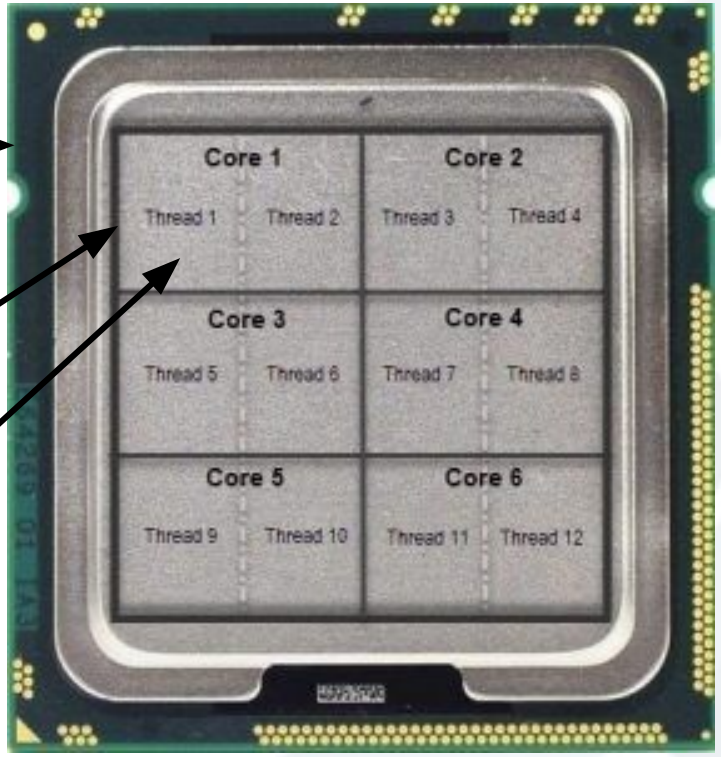



image source:
https://phoenixnap.com/kb/single-vs-dual-processors-server

# Compute nodes (x39)

| **Server type** | **LENOVO SR645** |
|---|---|
| Processor | **2*AMD 7713**<br>**(2*64c)** |
| Frequency | **2.0 / 3.7 GHz (turbo)** |
| RAM memory | **2048G / 4096G (bigmem01)** |
| Ethernet network<br>(Main communication & storage) | **10GbE**<br>latency **10 µs**<br>Throughput **1,2GB/s** |
| Infiniband network<br>(MPI interconnection & work storage) | **HDR-100**<br>latency **1,2 µs**<br>Throughput **12,3GB/s** |

# AMD ZEN3 (MILAN) 7NM

The most efficient choice today:

- Smaller *process node size* (less energy)
- Higher density (more cores, less servers)
- Best performance for bioinformatics computing

# VIEW & GPU nodes

| | VIEW | GPU |
|---|---|---|
| **Server** | **LENOVO SR665** | **LENOVO SR670 v2** |
| Processor | **2*AMD 7513 (2*32c)** | **2*INTEL Gold 6338 (2*32c)** |
| Frequency | 2.6 GHz | 2.0 GHz |
| RAM memory | 512G | 1024G |
| Local disks | 2*SSD 2T (Raid 1) | **4*NVMe 4T (Raid 5)** |
| GPU | **RTX 6000 24G** | **4*A100 80G** |
| Networks | 2*10GbE + HDR-100 | 10GbE + 2*HDR-100 |

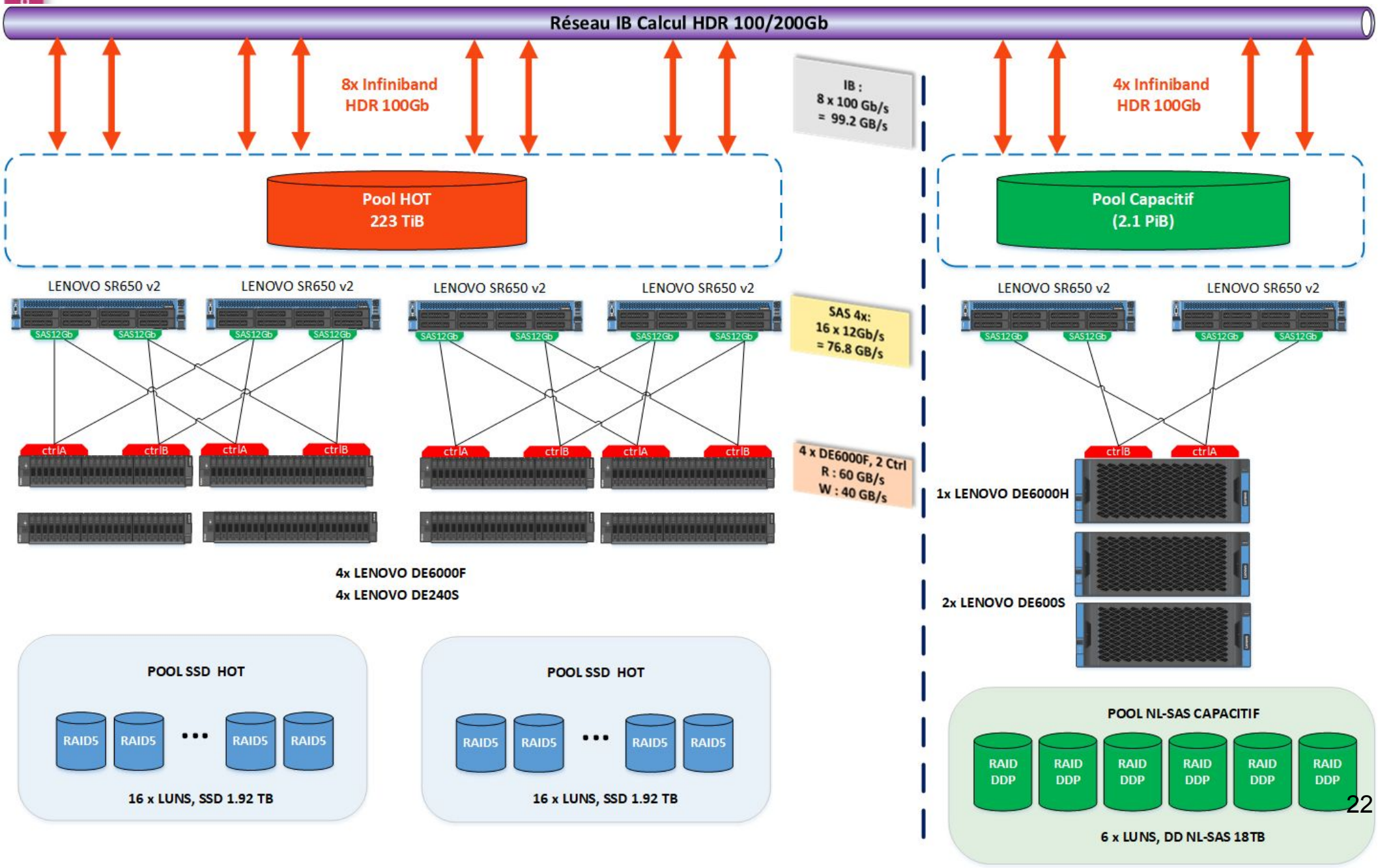
# ADMIN and LOGIN nodes

| | **ADMIN** | **LOGIN (*2)** |
|---|---|---|
| **Server** | **LENOVO SR645** | **LENOVO SR645** |
| Processor | **2*AMD 7313 (2*16c)** | **2*AMD 7513 (2*32c)** |
| Frequency | 3.0 GHz | 2.6 GHz |
| RAM memory | 512G | 512G |
| Local disks | 2*SSD 2T (Raid 1) | 2*SSD 2T (Raid 1) |
| Networks | 2*10GbE + HDR-100 | 10GbE + 2*HDR-100 |

# High Performant storage (`/work`)

- Provided by the "**integrator**"
- Parallel File System **IBM Spectrum Scale**
- Integration of servers + disk array
- **Hot data**: **SSD 1,92 TB** (192 disks)
- **Cold data**: **NL-SAS 18 TB** (160 disks)
- **2,3 PB usable**, **40 GB/s** bandwidth (R/W)
- NFS export to virtual machines
- Snapshots, quotas (users, groups)

# Performant storage /work



Genotoul Bioinfo

# Software utilization

# Genobioinfo: new environment

**New environment**
=> changes between the 2 clusters.
=> user access.

**Software usage**
=> reminder: run a soft.

**Help**
=> reminder: software help, other help.

**Others services**
=> monitoring, account information.

**Best practices**
=> security.
=> shared resources and environmental constraint.
=> infrastructure.

L'EUROPE S'ENGAGE L'OCCITANIE AGIT
UNION EUROPÉENNE
La Région Occitanie Pyrénées · Méditerranée

Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

# New environment
## Changes

| | | | Current cluster | New cluster |
|---|---|---|---|---|
| **Login server name** (reconfigure MobaXerm, Filezilla, etc.) | | | **genologin.toulouse.inra.fr** | **genobioinfo.toulouse.inrae.fr** |
| **Location** | | | INRAE Auzeville | ECA Toulouse |
| **Authentification** | | | **same accounts** | |
| **Data** | **Home** | Storage | current | New (from scratch) |
| | | Path | /home/<login> | |
| | **Work** | Storage | current | New (from scratch) |
| | | Path | /work/<login> | /work/**user**/<login> |
| | **Save** | Storage | **Shared (current)** | |
| | | Path | /save/<login> | /save/**user**/<login> |
| | **Banks** | Storage | current → replicated | New |

# New environment
## Changes

| | | Current cluster | New cluster |
|---|---|---|---|
| **CPU** (you may need to recompile your binaries) | | Intel Broadwell | AMD Zen3 Milan (except gpu node) |
| **OS** | | CentOS 7.9 | Red Hat 8.6 |
| **Software*** | | <category>/<soft>-<version> | <category>/<soft>/<version> |
| **Scheduler** | | slurm 21.08.8-2 | slurm 22.05.6 |
| **Accounting** (including cpu hours) | | current | New (from scratch) |
| **Queues** | *unlimitq* | 180 days | **90 days** |

* Some categories have changed, some (not bioinfo) softwares don't belong to the same categories in current and new cluster

# New environment

## User access

**Access to the platform**: via a command line SSH connection or MobaXterm for Windows

New frontal/login servers: genobioinfo1 & genobioinfo2

New hostname for the connection:
**genobioinfo.toulouse.inrae.fr**

Example: $ ssh <login>@genobioinfo.toulouse.inrae.fr

L'EUROPE S'ENGAGE L'OCCITANIE AGIT
UNION EUROPÉENNE
La Région Occitanie Pyrénées · Méditerranée

Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

# Software usage
## Reminder: Run a soft

**“Environment Modules package”**

“Modules” provides a dynamic modification of a user's environment via modulefiles (add command in your PATH, define specific environment variable, add paths to dependencies, add paths to specific libraries).

**(**see Appendix 1 & 2 and https://modules.sourceforge.net/ )

**Run a software**

To run a software you need to load the corresponding module.
**$ module load <modulename>**

To run a software with others software dependencies, you need to load all required modules.

**Best practices**

Check modules already loaded: **$ module list**

Purge modules already loaded if not needed:
**$ module purge** (all modules)
**$ module unload <modulename>** (only one module)

# New environment
## Software installation

### Genologin to Genobioinfo cluster: new installations

- ~ 100 software already installed on Genobioinfo cluster.
- To ask for new software on Genobioinfo cluster:

on demand by filling form on our website (Ask for/Software installation)
https://bioinfo.genotoul.fr/index.php/ask-for/install-soft/

### Installation and update

Default installation on **Genobioinfo** (if not specified in the request).

Reminder : updates only upon user request

Plateforme Bioinfo Occitanie Toulouse

# New environment
## Search/Find a soft (Web)

**Website** (Resources/Software):
https://bioinfo.genotoul.fr/index.php/resources-2/softwares/

**Not installed on Genobioinfo Cluster- link to ask for**

**All software**

| Application | Description | Avaibility/Use |
|---|---|---|
| ARKS | Scaffolding genome sequence assemblies using 10X Genomics GemCode/Chromium data. This project is a new kmer-based (alignment free) implementation of ARCS. It provides improved runtime performance over the original ARCS implementation by removing the requirement to perform alignments with bwa mem. | **Genologin Cluster:** How to use<br>**Genobioinfo Cluster (soon available):** Ask for Install |
| bwa | Burrows-Wheeler Aligner (BWA) is an efficient program that aligns relatively short nucleotide sequences against a long reference sequence such as the human genome. It implements two algorithms, bwa-short and BWA-SW. The former works for query sequences shorter than 200bp and the latter for longer sequences up to around 100kbp. Both algorithms do gapped alignment. They are usually more accurate and faster on queries with low error rates. | **Genologin Cluster:** How to use<br>**Genobioinfo Cluster (soon available):** How to use |

**Link to soft website**

**Installed on Genobioinfo Cluster- link to help**

# New environment

## Search/Find a soft (Command line)

### New installation paths (new storage spaces)

Bioinfo: /usr/local/bioinfo/src/
Compilers: /tools/compilers
Libraries: /tools/lib
Languages (Python, R , Java, Singularity..): /tools/statistics,/tools/containers,tools/devel
Other system tools : /tools/others_tools
Useful scripts: /tools/bin (sarray, squota_cpu, saccount_info…). In user's default PATH.

### Commands

- with useful scripts:

**$ search_module <soft_name>**: display available versions for a specific application (case insensitive)
**$ search_R_package <package_name>**: find if R package already installed (case insensitive)
**$ search_Python_package <package_name>**: find if Python package already installed (case insensitive)

- with **module** or **ls** command: see our FAQ

https://bioinfo.genotoul.fr/index.php/faq/software_faq/
-> Is a software is already installed on the cluster?

# Help
## Software help (1/2)

**Software documentation**

- official software documentation in the installation folder: /usr/local/bioinfo/src/<soft_name>/<soft_version>

- our website Software page: https://bioinfo.genotoul.fr/index.php/resources-2/softwares/

**Use on SLURM cluster**: "How_to_use_SLURM_<soft_name>" file

- on our website Resources/Software page (Availability/Use column, click on Genobioinfo cluster link).

- in software installation directory /usr/local/bioinfo/src/<soft_name>

- find "How_to_use_SLURM_<soft_name>" file path : $ module help <modulename>

- a basic « example_on_cluster » directory in the software installation directory: /usr/local/bioinfo/src/<soft_name>/example_on_cluster

Genotoul Bioinfo

# Help
## Software help (2/2)

HOW TO USE ON GENOBIOINFO CLUSTER

```
SOFT : bwa
------
Site du soft: https://github.com/lh3/bwa
-------------

LICENSE:
--------

GPL-v3

See software documentation for more informations.

Location: /usr/local/bioinfo/src/bwa
---------
```

Software informations

```
Load binaries and environment:
------------------------------

-> Version v0.7.17

module load bioinfo/bwa/0.7.17
```

Usage and versions

```
Example directory for use on cluster:
-------------------------------------

/usr/local/bioinfo/src/bwa/example_on_cluster

To submit: sbatch test bwa-0.7.17.sh
```

Example

# Help
## Usage examples

## Use STAR-v2.7.5a:

**$ search_module Star**
bioinfo/STAR/2.7.10b
bioinfo/STAR/2.7.5a

**$ module help bioinfo/STAR/2.7.5a**
---
Module Specific Help for
/tools/modulefiles/bioinfo/STAR/2.7.5a:

See How_to_use file:
/usr/local/bioinfo/src/STAR/How_to_use_SLURM_STAR
---

**$ module load bioinfo/STAR/2.7.5a**
**$ which STAR**
/usr/local/bioinfo/src/STAR/STAR-2.7.5a/source/STAR
**$ STAR --version**
2.7.5a

**$ module unload bioinfo/STAR/2.7.5a**
**$ STAR --version**
-bash: STAR: command not found

# Help
## Other

**Browse all "How_to_use_SLURM_<soft_name>" files** (in your web browser)

https://web-genobioinfo.toulouse.inrae.fr/How_to_Softs/

**Useful scripts (already in your default path or /tools/bin)**
saccount_info,  sq_long or sq_debug (squeue long format), sa_debug (sacct long format), sarray (wrapper for sbatch to create a job array that will execute in parallel SHELL lines from a file), squota_cpu to see your CPU time limit, sq_veille (summary of current jobs, running, pending...), sq_run or sq_pend for running or pending job,last_soft.sh (last softs installed)...

**FAQ:** https://bioinfo.genotoul.fr/index.php/faq/

**Newsletters:** https://bioinfo.genotoul.fr/index.php/about-us/newsletters/

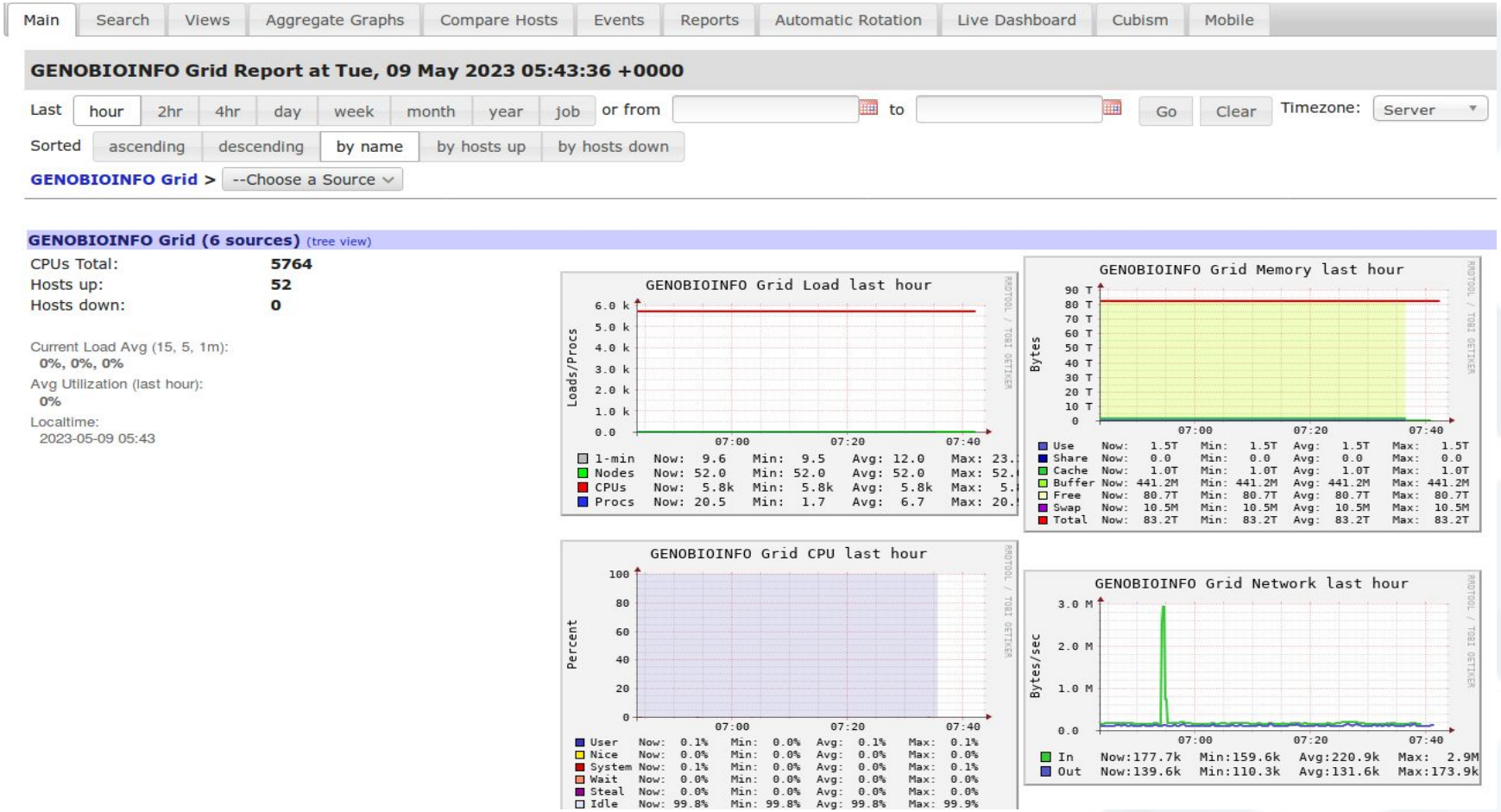
**Training slides:**
Linux (for Genologin cluster):
https://genoweb.toulouse.inra.fr/~formation/unix/doc/Formation_LINUX_GenoToul.pdf

Cluster (for Genologin cluster):
https://genoweb.toulouse.inra.fr/~formation/cluster/doc/Formation_cluster_SLURM.pdf

# Others services

## Monitoring

**Ganglia:** https://web-genobioinfo.toulouse.inrae.fr/ganglia/
(or our website : Resources/Monitoring)

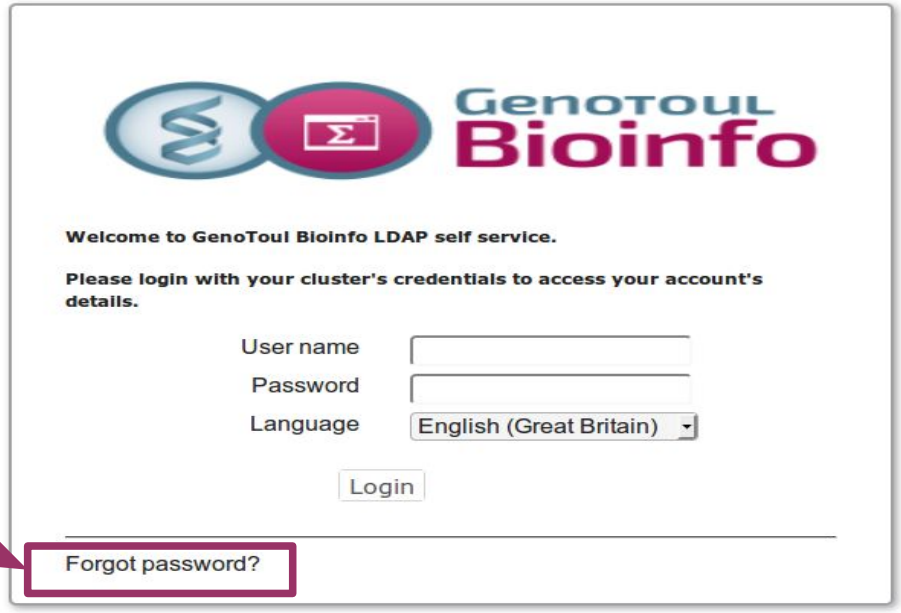



Some help to understand Ganglia view here.

# Others services

## Account information, password change

**Self Service →** https://selfservice.bioinfo.genotoul.fr




**Change your password (every year)**

**CLI command: saccount_info <login>**

- account expiration date and last password change date
- your primary Linux group
- your secondary Linux groups if you have any
- status of your Linux primary group in Slurm (contributors, inrae, region or others)
- your groups' members
- some Slurm limitations of your account : cpu and memory limit, CPU Time ...

# Best practices

## Security

**One user = one account**
You are responsible of the damage caused by your login.

**Default permissions directories**
**- home:** `drwxr-x--x` **R**ead, **W**rite, e**X**ecution for the owner, **R**ead and e**X**ecution for the group members, e**X**ecution for all.
**- save and work:** `drwxr-x---` **R**ead, **W**rite, e**X**ecution for user, **R**ead and **E**xecution for your group members, no permissions for all.

To change permissions: **chmod** command or ACL

https://bioinfo.genotoul.fr/index.php/faq/linux_faq/
-> How to change permissions on file or folder?

# Best practices

## Shared resources and environmental concerns (1/2)

**Cluster is a shared and power-consuming resource,**
**so ... think about others and environmental preservation.**

**Objectives:** Go faster and save time + Free up resources for others + Calculate less, better and therefore consume less energy: different objectives but same means.

**Means:** ask yourself the right questions.

### In which case to use a cluster?

**Cluster use ≠ faster process**

It is useful to use a computing cluster rather than a personal computer when:

- you need more resources than available on your PC (RAM, CPU, disk space),
- the same job needs to be run several times on different data (possibility to run it in parallel on several nodes),
- the program you want to run would benefit from being "multi-threaded" via MPI or OpenMP.

Resources (CPU, memory, I/O…) are shared on a node with other users and this can create latencies.

# Best practices

## Shared resources and environmental concerns (2/2)

**How to request resources adapted to your needs?** (including in workflows as nf_core …)

The smaller a job is (cpu, memory, time), the faster it will be launched. You don't save time by overbooking or misbooking; you stay in the queue longer and you block the others.

A best practice is to run a **test** on a sample **and adjust** your resources before launching identical jobs massively:

- choose the **right queue**: `unlimitq` is very limited in resources. Only use `unlimitq` when a job requires more than 4 days.
- adjust the **memory**: first do a **test and check** the memory consumed by the job in COMPLETED state with the `seff` command: `$ seff <jobid>`
- reserve several **cpus** only if the program used can do multithreading. The number of cpu reserved with Slurm must match the number of cpus set in the software parameters.
- adjust the **time** option `-t`

# Best practices

## Infrastructure

**Help in keeping the infrastructure and the support to run smoothly.**

**- DO NOT run jobs on frontal servers (tmux, screen, R, snakemake)**:

Why ?
- overloading frontal servers slows down everyone.
- overloading frontal servers can crash them and block everyone.

Check your processes on frontal servers before exit: $ pstree -u <login>
Kill them and close Screen/Tmux session when they are no longer useful

**Any job launched on the servers "genobioinfo" will be immediately killed by the system administrators.**

**- Configure your antispam** to accept these e-mails: support.bioinfo.genotoul@inrae.fr,support.bioinfo.genotoul@inra.fr,bioinfo.genotoul@groupes.renater.fr,ldap_admin@genobull.toulouse.inra.fr (changement de mot de passe)

- **Inform us of a departure:** the number of accounts is limited so when we can delete some and leave room for others, it's better.

- **Respect our charter.**

- **Be precise in support requests support** (and/or use our forms to give the right information)

# Support

## Infrastructure

**Bioinfo genotoul website**
https://bioinfo.genotoul.fr/

**Bioinfo Genotoul Chart**
https://bioinfo.genotoul.fr/wp-content/uploads/ChartPFBioinfoGenoToul.pdf

**FAQ**
https://bioinfo.genotoul.fr/index.php/faq/

**Support**
Mail: support.bioinfo.genotoul@inrae.fr
Fill form (best for us): https://bioinfo.genotoul.fr/index.php/ask-for/support/

Genotoul Bioinfo

# What about:

## Web services ?

# Web services

- New infrastructure implies:
  - New computing facilities access (new Slurm scheduler)
  - New storage locations (home & work)
- **Migration is required for web services/servers that use:**
  - **Current cluster (2017) facilities**
  - **Current storage for home or save**
- Impacted services/servers are (among other):
  - https://genoweb.toulouse.inra.fr (public_html)
  - https://vm-galaxy-prod.toulouse.inrae.fr (Galaxy)
- => New servers will be set up, and users will be guided through the migration steps.
- => Web admin will be contacted and assisted as well.
- Migration will be as smooth as possible, and public URL will stay unchanged.

GenoToul Bioinfo

# Thank you for your attention

Any questions ?

L'EUROPE S'ENGAGE L'OCCITANIE AGIT

UNION EUROPÉENNE

La Région Occitanie Pyrénées · Méditerranée

Projet cofinancé par le Fonds Européen de Développement Régional
Financement dans le cadre de la réponse de l'Union à la pandémie de COVID-19

# WORK Data Migration


From the new genobioinfo node (on new cluster):
Replace "**username**" by your own login and type:

"rsync -avh /oldwork/**username** /work/user/**username**"

# Appendix 1 :Software usage

## Create your own environment

### Create your modulefile

Basic example  for the modulefile  ~/save/my_own_module :

```
#%Module1.0#####################################################################
##

module load bioinfo/STAR/2.7.5a
prepend-path PATH /save/<login>/my_binaries/
```

### Load your module

**$ module load -f ~/save/my_own_module**

### Verify your environment

**$ which STAR**
/usr/local/bioinfo/src/STAR/STAR-2.7.5a/source/STAR

### Unload one module

**$ module unload bioinfo/STAR/2.7.5a**

### Unload all module and specific variable

**$ module purge**

# Appendix 2: Software usage
## Environment modules

**Basic command to use module:**

module : (no arguments) print usage instructions
module avail : list available software module
module load modulename : add a module to your environment
module unload modulename : unload remove a module
module purge : remove all modules
module show modulename : show what changes a module will make to your environment
module help modulename : path to the hel ("How_to_use_SLURM_<soft_name>" file for us)

For more documentation, see the Environment Module website : https://modules.sourceforge.net/

# Appendix 3: hot or cold data?

**How to know if file or directory is on hot or cold work space storage?**

$ mmlsattr -L <file or directory>

Example:
$ mmlsattr -L scripts_test

```
file name:            scripts_test
metadata replication: 1 max 2
immutable:            no
appendOnly:           no
flags:
storage pool name:    system
fileset name:         user
snapshot name:
creation time:        Wed Aug 18 14:09:57 2021
Misc attributes:      DIRECTORY
Encrypted:            no
```

**system=hot space**
**capacitive=cold space**

# Storage benches (`/work`)

| Caractéristique | Par baie | Solution proposée |
|---|---|---|
| Max Random IOPS Read (4KB) | 1 M | 4M |
| Max Random IOPS Write (4KB) | 160 000 | 640 000 |
| Max Sequential Reads (1MB) | 15 GB/s | 60 GB/s |
| Max Sequential Writes (1MB) *cache mirroring disabled* | 14 GB/s | 56 GB/s |
| Max Sequential Writes (1MB) *cache mirroring enabled* | 10 GB/s | 40 GB/s |

# Storage benches (/work)

| | Per disk array | Our solution (4 disk arrays) |
|---|---|---|
| Max random IOPS Read (4 KB) | 1M | 4M |
| Max random IOPS Write (4 KB) | 160 000 | 640 000 |
| Max Sequential Reads (1 MB) | 15 GB/s | 60 GB/s |
| Max Sequential Write (1 MB)<br>*Cache mirroring disabled* | 14 GB/s | 56 GB/s |
| Max Sequential Write (1 MB)<br>*Cache mirroring enable* | 10 GB/s | 40 GB/s |

# Compute benches

- Bench HPL (39 nodes) : **136 TFlops/s**
- Bench HPL GPU : 55 TFlops/s

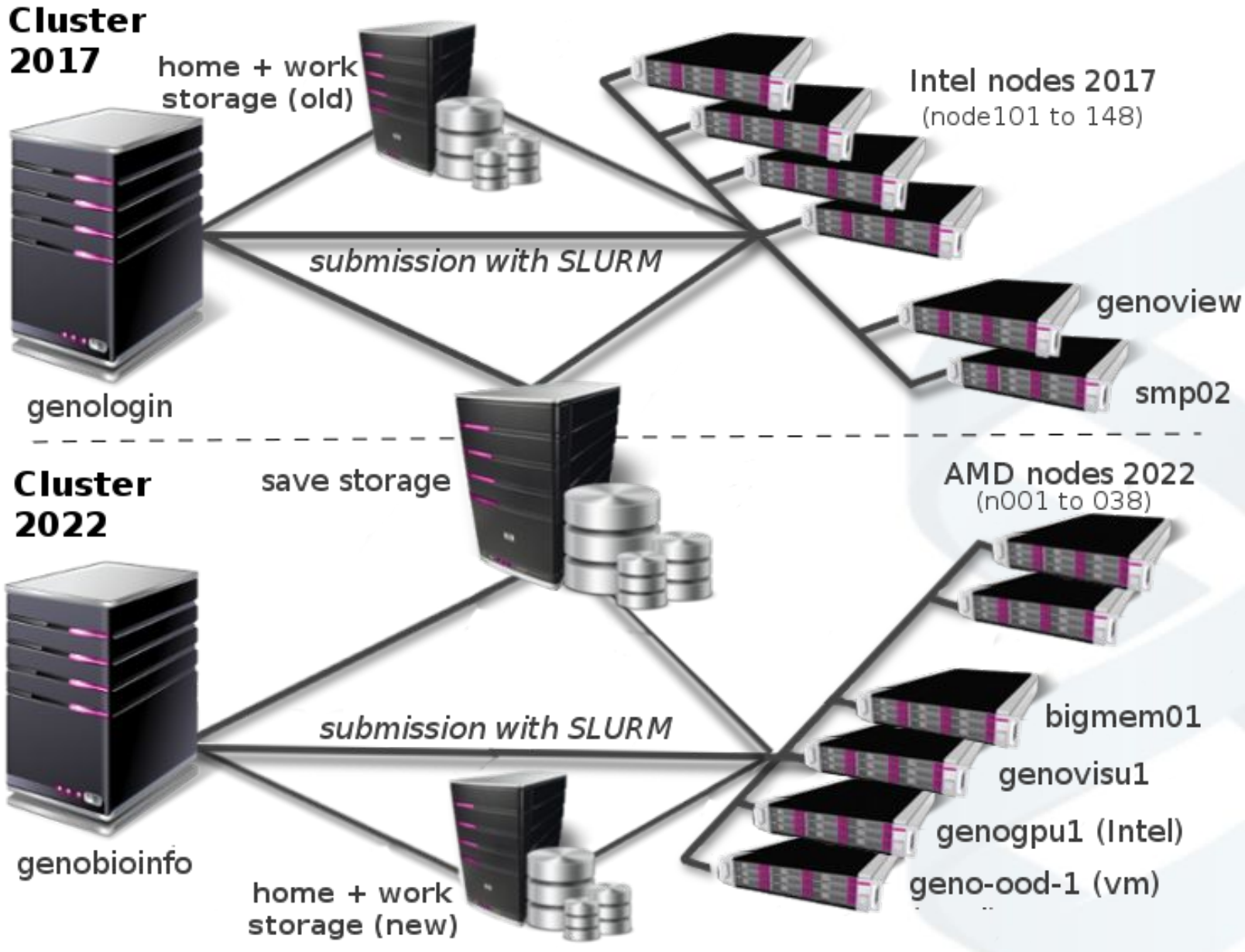
- Bench STREAM memory (1 nodes) : de l'ordre de 300GB/S

- Bench **BWA**, **SAMTOOLS** (1 instance on 1 node):  100s
- Bench BWA, SAMTOOLS (128 instances on 1 node):  250s

- Bench **STAR** (1 instance * 128 threads) : 193s
- Bench STAR (16 instances * 16 threads) : 1775s

Genotoul Bioinfo

# New environment
## For one year, 2 clusters



# Open On Demand portal

**=> Work In Progress (currently being deployed)**

Launch **interactive sessions** on the virtualization node through the Web browser:

Tools :

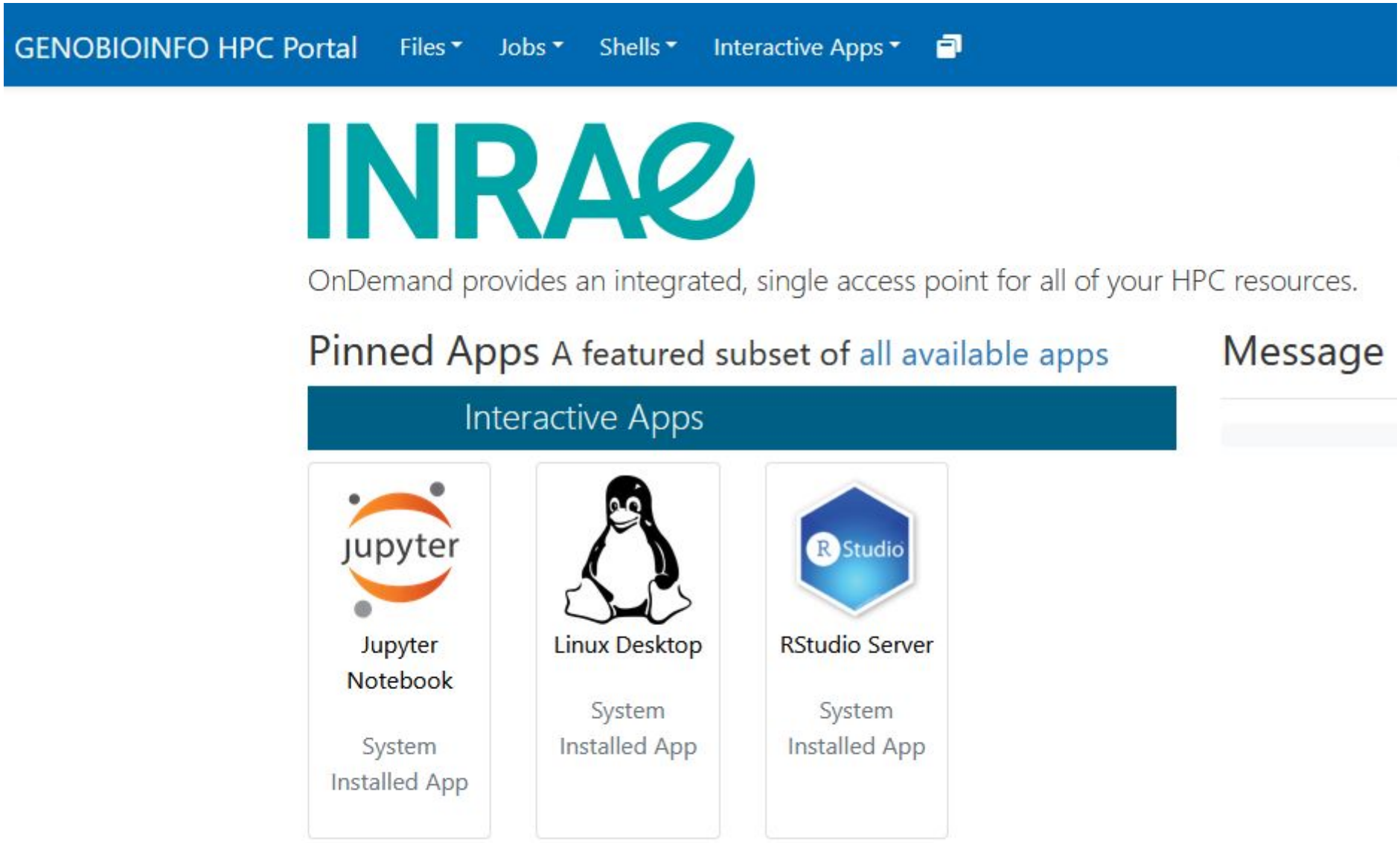
- **Linux 3D desktop**,
- **Jupyter Notebook**,
- **Rstudio**

**Submit jobs** (SLURM) on the compute cluster

**Browse the files**

**Transferts** from/to the user desktop

# Portal Open On Demand



Genotoul Bioinfo

# What about:

## https://vm-galaxy-prod.toulouse.inrae.fr

# Galaxy

- Current Galaxy server uses your account for:
  - job submission on current cluster (2017)
  - temp processing files in the work storage
- A new Galaxy server will be installed this year:
  - With latest version of Galaxy server
  - Submission on new cluster (still with your account)
  - Use of new storage locations
- Waiting for this new server for every body:
  - Old one will still be available for current **and new** users
  - Migration support will be provided (dataset and tools)
- Old server will be stopped at the end of the year