

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tunning
  - Job array
- Best practices
- Support



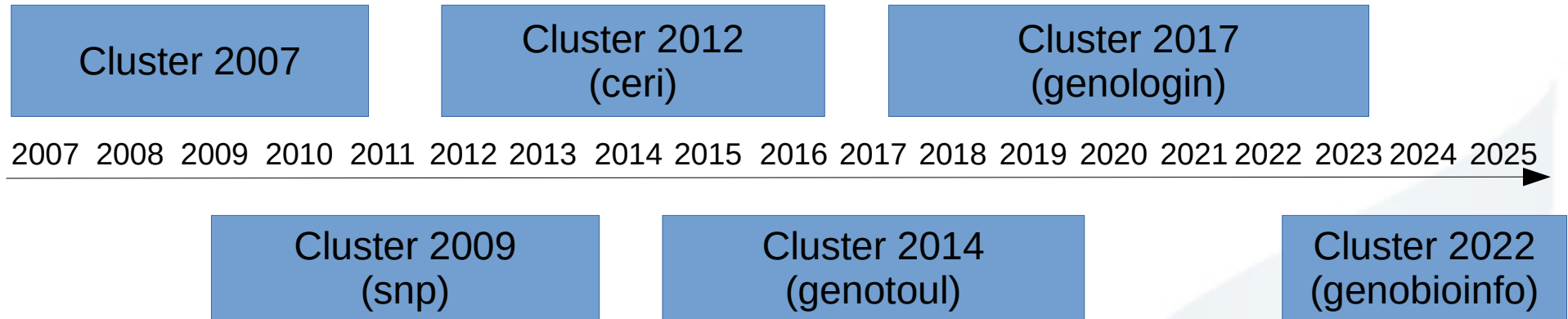
## PRE-REQUISITE : LINUX

- connect to « genobioinfo » server
- Basic command line utilization
- File System Hierarchy
- Useful tools (find, sort, cut, grep)
- Transferring & compressing files

## TODAY

- How to use compute nodes cluster (submit, manage & monitor jobs)
- Objectives : Autonomy, self mastery

# Context renewal strategy



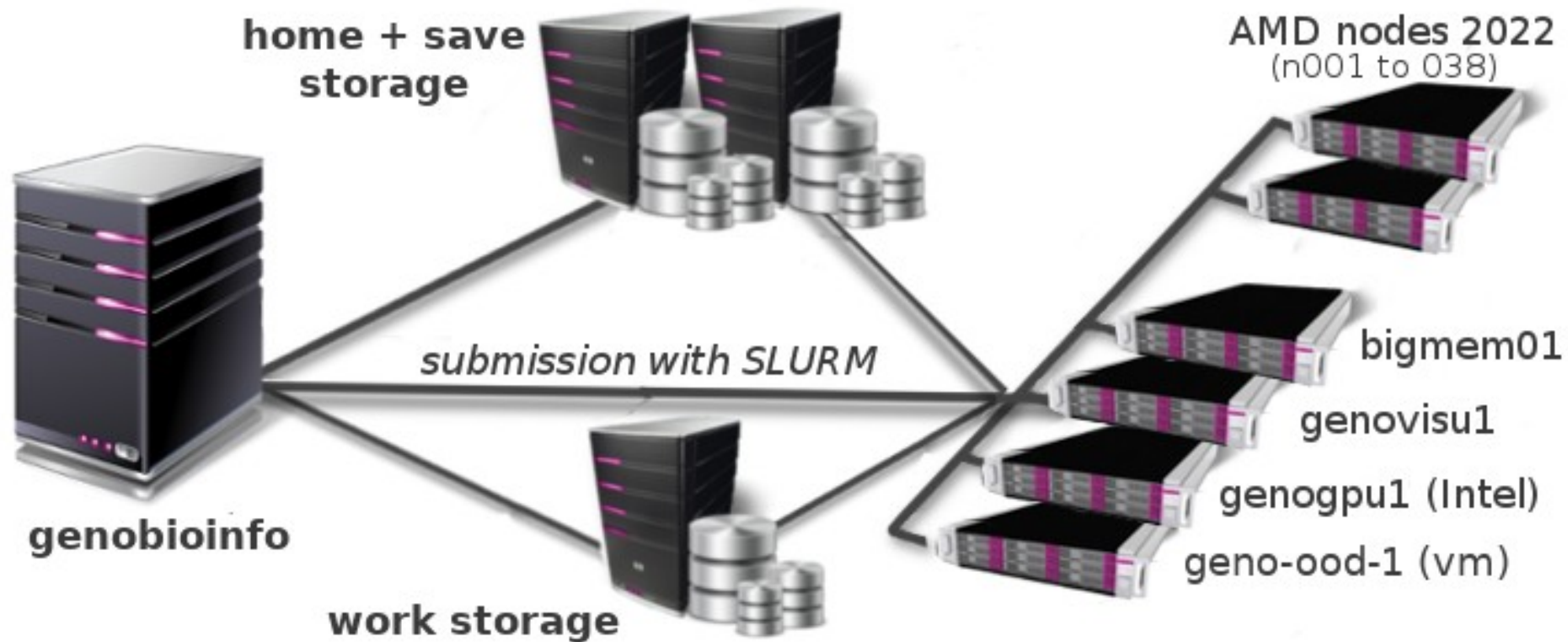
- Overlapping clusters enabling to keep the service active and to renew the machines
- SLURM job scheduler from 2017 (before=SGE)

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tunning
  - Job array
- Best practices
- Support



# Infrastructure



# Infrastructure

## login & compute nodes

### login nodes ([genobioinfo.toulouse.inrae.fr](http://genobioinfo.toulouse.inrae.fr))

- 2 login nodes : genobioinfo1&2 \* (32 cores, 512 GB RAM)
- Linux based on RedHat-8 distribution
- Hundreds of users simultaneous
- Secured (ssh only)

# Infrastructure

## login & compute nodes

### login nodes ([genobioinfo.toulouse.inrae.fr](http://genobioinfo.toulouse.inrae.fr))

- To serve development environments
- To test his script before data analysis
- To launch jobs on the cluster nodes
- To get data results on the /save directory

**DO NOT run data treatment on login nodes**

## Compute nodes

Cluster available resources

Node type	Name(s)	#Threads <sup>①</sup>	Memory <sup>①</sup>	Processor(s) <sup>①</sup>	GPU <sup>①</sup>
standard (x38)	n001 to n038	128	2 TiB	2 x AMD EPYC 7713 <sup>①</sup>	-
bigmem	bigmem01	128	4 TiB	2 x AMD EPYC 7713 <sup>①</sup>	-
gpu	gpu01	128	1 TiB	2 x Intel Gold 6338 <sup>①</sup>	4 x NVIDIA A100 80G
gpu	gpu02 & gpu03	256	1.5 TiB	2 x AMD EPYC 9554 <sup>①</sup>	4 x NVIDIA L40S 48G
visu	visu01	128	512 GiB	2 x AMD EPYC 7513 <sup>①</sup>	1 x NVIDIA RTX6000 24G

# Infrastructure

## login & compute nodes

### Compute nodes

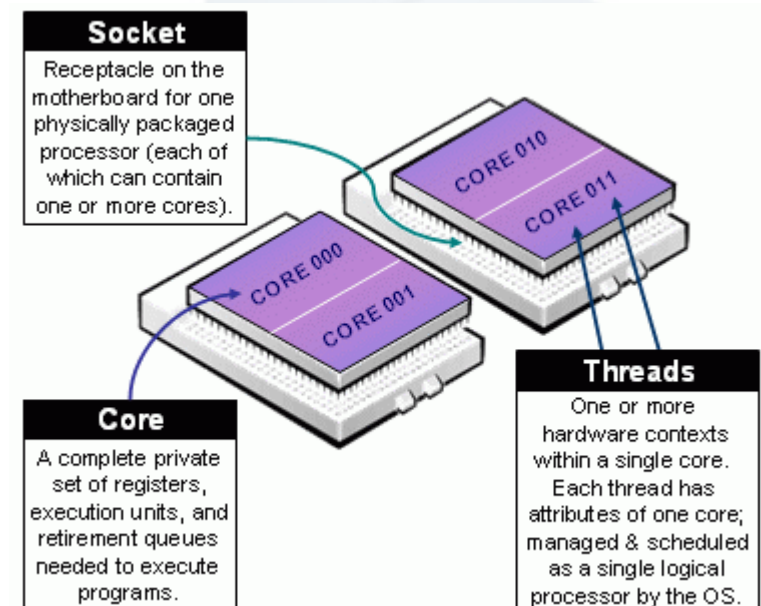
- Low latency & high bandwidth interconnection (100Gb/s)
- Interactive mode : for beginners / for remote display
- Batch access : for intensive usage (most of jobs)
- No direct ssh access to the nodes
- Workspace exactly the same as login nodes (exception read only on /save directory)

## Cluster / Node

- Cluster : a set of compute nodes
- Node : a computer with multi-processors and huge memory

## Socket / Core

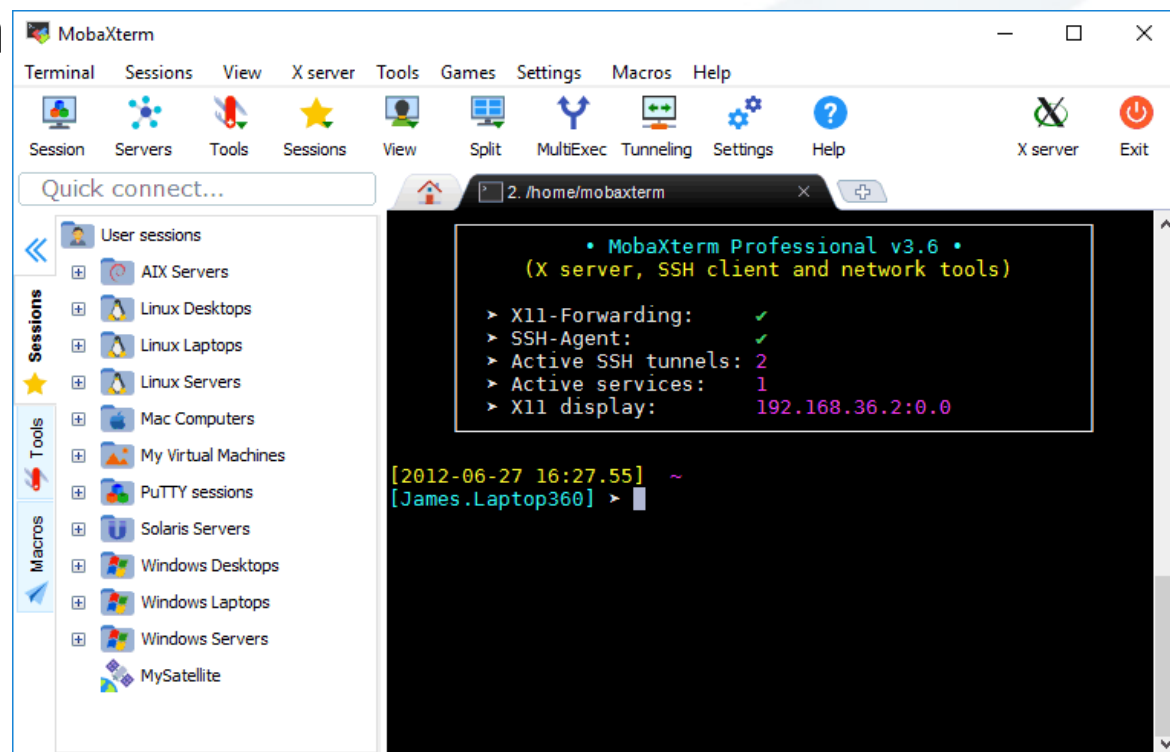
- Socket: Central Processing Unit (CPU)
- Core : multi-core in a CPU
- Thread (multithreading deactivated)  
=> 1 core = 1 thread



# Infrastructure

## User account

- Access to the login servers: **genobioinfo** (1 or 2)
- Linux / Mac : via command line SSH connection  
***ssh <login>@genobioinfo.toulouse.inrae.fr***
- Windows : via MobaXterm



# Infrastructure

## Disk spaces

- All of directories are the same between genobioinfo servers & cluster nodes (you don't have to copy anything)
- Examples :
  - /home, /save, /work* : user (or project) directories
  - /bank* : international genomics databanks
  - /usr/local/bioinfo* : Bioinformatics software

# Infrastructure

## User quotas

- **10GB** for **/home** directory (configuration files only)
- **250GB (\*2)** for **/save/user** directory (permanent data, with replication)
- **1TB** for **/work/user** directory (temporary compute disk space)
- **100,000H** annual **calculation time** (500H for private user)  
You could have more time on demand (resource request)

**Refresh your linux knowledge:**

<https://genotoul-bioinfo.pages-forge.inrae.fr/linux-cluster/cluster/tp1.1/>

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tuning
  - Job array
- Best practices
- Support



# Environment Search/Find a soft (Web)

**Website (Resources/Software):** <http://bioinfo.genotoul.fr/index.php/resources-2/software/>

**Resources**

- + Carbon footprint
- + Hardware
- + Monitoring
- + Software**
- + Databanks
- + Pricing

## SOFTWARE

The GenoToul bioinformatics platform provides access to high-performance computing resources with softwares already installed to ease its usage. An exhaustive list is provided hereunder. **Software are updated only upon user request.** If you need any other software or if you need an update, fill the [installation software form](#).

---

Select a category:

Search a software:

---

### All software

Application	Description	Availability/Use
<a href="#">3D-DNA</a>	3D de novo assembly (3D-DNA) pipeline.	Genobioinfo Cluster: <a href="#">How to use</a>
<a href="#">3rdChimeraMiner</a>	Exploration of whole genome amplification generated chimeric sequences in long read sequencing data.	Genobioinfo Cluster: <a href="#">How to use</a>
<a href="#">AAF</a>	This is a package for constructing phylogeny without doing alignment or assembly.	Genobioinfo Cluster: <a href="#">Ask for Install</a>
<a href="#">ABCtoolbox</a>	BCtoolbox is a general purpose program to perform Approximate Bayesian Computation. ABCtoolbox can be used for ABC inference on almost any type of model, including models arising in physics, biology or engineering.	Genobioinfo Cluster: <a href="#">Ask for Install</a>
<a href="#">ABYSS</a>	ABYSS (Assembly By Short Sequences) is a de novo, parallel, paired-end sequence assembler that is designed for short reads.	Genobioinfo Cluster: <a href="#">How to use</a>
<a href="#">AC-DIAMOND</a>	AC-DIAMOND attempts to speed up DIAMOND via better SIMD parallelization and compressed indexing. Experimental results show that AC-DIAMOND was about 6-7 times faster than DIAMOND on aligning DNA reads or contigs while retaining the essentially the similar sensitivity. AC-DIAMOND was developed based on DIAMOND v0.7.9.	Genobioinfo Cluster: <a href="#">Ask for Install</a>

# Environment

## Environment modules

The **Environment Modules package** provides for the dynamic modification of a user's environment via modulefiles.

**module** command alter or set shell environment:

- add command in your PATH
- define specific environment variable
- add path to dependencies
- add path to specific librairies

Modules can be loaded and unloaded dynamically.

Modules are useful in managing different versions of applications.

# Environment Search examples

## module search cutadapt

----- /tools/modulefiles -----

bioinfo/Cutadapt/1.8.3: loads the bioinfo/Cutadapt/1.8.3 environment

bioinfo/Cutadapt/4.3: loads the bioinfo/Cutadapt/4.3 environment

## module search blast

----- /tools/modulefiles -----

bioinfo/NCBI\_Blast+/2.9.0+: loads the bioinfo/NCBI\_Blast+/2.9.0+ environment

bioinfo/NCBI\_Blast+/2.10.0+: loads the bioinfo/NCBI\_Blast+/2.10.0+  
environment

bioinfo/NCBI\_Blast/2.2.26: loads the bioinfo/NCBI\_Blast/2.2.26 environment

bioinfo/RMBlast/2.13.0: loads the bioinfo/RMBlast/2.13.0 environment

bioinfo/WuBlast/2.0a19: loads the bioinfo/WuBlast/2.0a19 environment

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tuning
  - Job array
- Best practices
- Support



# Software usage

## Run a soft

### Run a software

To run a software you need to load the corresponding module.

**module load <module\_name>**

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 module\_name** (only one module)

# Software usage

## Usage examples

### Use Bismark-0.24.0

*module search bismark*

*module load bioinfo/Bismark/0.24.0*

*which bismark*

*/usr/local/bioinfo/src/Bismark/Bismark-0.24.0/bismark*

*Bismark --help*

### Use Python-3.7.9

*module search python*

*module load devel/python/Python-3.7.9*

*which python*

*/tools/devel/python/Python-3.7.9/bin/python*

*python --help*

# Software usage

## Module command

**module** : (no arguments) print usage instructions

**module search search\_module** : display available versions for a specific application

**module avail** : list available software module

**module load module\_name** : add a module to your environment

**module unload module\_name** : unload remove a module

**module purge** : remove all modules

**module show module\_name** : show what changes a module will make to your environment

**module help module\_name** : path to the "How\_to\_use\_SLURM\_<soft\_name>" file

For more documentation, see the Environment Module website :

<http://modules.sourceforge.net/>

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tuning
  - Job array
- Best practices
- Support



## Software documentation

- official software documentation in the installation folder `/usr/local/bioinfo/src/<soft_name>/<soft_version>`
- our website Software page: **link to software website.**

## Use on SLURM cluster

- "How\_to\_use\_SLURM\_<soft\_name>" file:

software installation directory `/usr/local/bioinfo/src/<soft_name>`

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

- a basic « **example\_on\_cluster** » directory in the software installation directory

`/usr/local/bioinfo/src/<soft_name>/example_on_cluster`

## example : samtools (1)

LICENSE:

-----

The MIT/Expat License

See software documentation for more informations.

Location: /usr/local/bioinfo/src/samtools

-----

Load binaries and environment:

-----

(include bgzip,tabix,htsfile)

-> Version v0.1.19

module load bioinfo/samtools/0.1.19

-> Version v1.10

module load bioinfo/samtools/1.10

-> Version v1.14

module load bioinfo/samtools/1.14

-> Version v1.18

module load bioinfo/samtools/1.18

Example directory for use on cluster:

-----

/usr/local/bioinfo/src/samtools/example\_on\_cluster

To submit: sbatch test\_samtools-1.14.sh

# Help

## example : samtools (2)

**ls /usr/local/bioinfo/src/samtools/**

example\_on\_cluster

How\_to\_use\_SLURM\_samtools

Install

samtools-0.1.19

samtools-1.10

samtools-1.14

samtools-1.18

- Find "**How\_to\_use\_SLURM\_<soft\_name>**" file path

### **module help bioinfo/samtools**

Module Specific Help for /tools/modulefiles/bioinfo/samtools/1.18:

See How\_to\_use file:

/usr/local/bioinfo/src/samtools/How\_to\_use\_SLURM\_samtools

- **Browse all "How\_to\_use\_SLURM\_<soft\_name>" files** (in your web browser)

[https://web-genobioinfo.toulouse.inrae.fr/How\\_to\\_Softs/](https://web-genobioinfo.toulouse.inrae.fr/How_to_Softs/) (new cluster)

- **Updated FAQ:** <http://bioinfo.genotoul.fr/index.php/faq/>

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- **SLURM**
  - Basics
  - Tuning
  - Job array
- Best practices
- Support



# SLURM System

## SLURM

- Simple Linux Utility for Resource management
- Adopted by the academic community
- Supported by IT providers
- New features
- **<https://slurm.schedmd.com/>**

## RedHat-8

- Leader on enterprise operating systems
- Supported by IT providers
- Cgroups (Control Groups) compatible

# SLURM

## Submission commands

### Job submission

#### [INTERACTIVE]

- **srun --pty bash** : submit an interactive session with a compute node (default workq partition).
- **srun --x11 --pty bash** : submit an interactive session with X11 forwarding (default workq partition)

#### [BATCH]

- **sbatch** : submit a batch script to slurm.
- **sarray** : submit a batch job-array to slurm
- **scancel** : kill the specified job

# SLURM

## Monitoring commands

### Job monitoring

- **sinfo** : display nodes, partitions, reservations
- **squeue** : display jobs and state
- **scontrol show** : get informations on jobs, nodes, partitions
- **sview** : graphical user interface
- **sacct** : display accounting data
- **seff** : display consumed ressources (time, cpu, ram)

**Submit your first job:**

<https://genotoul-bioinfo.pages-forge.inrae.fr/linux-cluster/cluster/tp1.2/>

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tuning
  - Job array
- Best practices
- Support



# SLURM

## Default parameters

- **workq partition**
- **1 node**
- **1 thread = 1 core**
- **2GB RAM memory per core = 2GB here**
- 100KH annually compute time (more on demand)
- 10000: max jobs number for all users
- 2500: max jobs number per user
- 2500 : max tasks number in a job array

# SLURM

## sbatch script

```

# !/bin/bash

#SBATCH --time=00:10:00          #job time limit
#SBATCH -J testjob              #job name
#SBATCH -o output.out          #output file name
#SBATCH -e error.out           #error file name
#SBATCH --mem=8G               #memory reservation
#SBATCH --cpus-per-task=4      #ncpu on the same node
#SBATCH --mail-type=BEGIN,END,FAIL #event notification
#SBATCH --mail-user=firstname.lastname@inrae.fr #user email

module purge                    #Purge any previous modules
module load bioinfo/ncbi-blast-2.2.29+ #Load the application
blastall ...                    # My command line I want to run on the cluster

```

# SLURM

## Directives (1/2)

<b>-p workq</b>	partition name
<b>--time=00:10:00</b>	job time limit
<b>-J testjob</b>	job name
<b>-o output.out</b>	output file
<b>-e error.out</b>	error file name
<b>--mem=8G or --mem-per-cpu</b>	memory size

# SLURM

## Directives (2/2)

<b>--cpus-per-task=4</b>	ncpu on the same node
<b>--mail-type=[events]</b>	event notification
<b>--mail-user=[address]</b>	user email
<b>--export=[ALL NONE variables]</b>	copy environment
<b>--workdir=[dir_name]</b>	working directory
<b>--wrap="command"</b>	With sbatch to submit directly one command"

# SLURM

## Partitions

- Each job is submitted to a specific partition (the default one is the workq).
- Each partition has a different priority considering the maximum time of execution allowed.

Partitions (queues)	Access	Nb Nodes	Max time	Max threads
workq	everyone	39	4 days (96h)	4992
unlimitq	everyone	39	90 days	4992
interq (VISU)	<del>on demand</del> everyone	1	12h	128
gpuq (GPU)	on demand	3	8 days	640

# SLURM

## Ressources

- There are job limitations on users + group of users (slurm\_info\_limit)
- It depends on your linux group : contributors / INRAe or OCCITANIE / others
- It is the same thing for the RAM memory (1 thread = 16GB)

Partition / max threads	workq (group)	workq (user)	unlimitq (all)	unlimitq (user)
Contributors	5760	750	<b>780</b>	500
INRAe or Occitanie	4608	500	<b>780</b>	376
Others	1440	250	<b>780</b>	100

### gpujob max :

cpu=128, mem=512G  
 gres/gpu:nvidia\_a100=2,  
 gres/gpu:nvidia\_l40s=2,

## Multithreading

<https://genotoul-bioinfo.pages-forge.inrae.fr/linux-cluster/cluster/tp2/>



# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tuning
  - Job array
- Best practices
- Support



# SLURM

## Job arrays

### **sbatch -a | array=<indexes>**

Submit a job array, multiple jobs to be executed with identical parameters.

Multiple values may be specified using a comma separated list and/or a range of values with a « - » separator.

#### Example :

--array=1-10

--array=0,6,16-32

--array=0-16:4 #a step of 4

--array=1-10%2 #a maximum of 2 simultaneously running task

SLURM_ARRAY_TASK_ID	Job array ID (index) number
SLURM_ARRAY_JOB_ID	Job array's master job ID number
SLURM_ARRAY_TASK_MAX	Job array's maximum ID (index) number
SLURM_ARRAY_TASK_MIN	Job array's minimum ID (index) number
SLURM_ARRAY_TASK_COUNT	total number of tasks in a job array

# SLURM

## --format option

Some commands (like **sacct** and **squeue**) give the possibility to **tune output format** :

### Example :

```
sacct --format=jobid%-13,user%-15,uid,jobname%-15,state%20,exitcode,Derivedexitcode,nodelist% -X
-job 6969
```

JobID	User	UID	JobName	State	ExitCode	DerivedExitCode	NodeList
6969	root	0	toto	COMPLETED	0:0	0:0	node[101-102]

```
squeue --format="%10i %12u %12j %.8M %.8l %.10Q %10P %10q %10r %11v %12T %D %R" -S "T"
```

JOBID	USER	NAME	TIME	TIME_LIM	PRIORITY	PARTITION	QOS	REASON	RESERVATION	STATE	NODES	NODELIST(REASON)
6612	root	bash	16:09	4-00:00:00	1	workq	normal	None	(null)	RUNNING	2	node[101-102]
6542	dgorecki	TurboVNC	1-06:27:44	UNLIMITE	1	interq	normal	None	(null)	RUNNING	1	genoview

# SLURM

## Useful scripts

These useful scripts are already in your default path or **/tools/bin**

- **saccount\_info <login>**: account expiration date and last password change date, primary and secondary Linux group, groups' members, Slurm limits, CPU Time ...
- **sq\_long** or **sq\_debug**: squeue long format
- **sa\_debug**: sacct long format
- **squota\_cpu**: see your CPU time limit.
- **seff <jobid>**: check the efficiency of your terminated job (cpu, memory)

## Job array

<https://genotoul-bioinfo.pages-forge.inrae.fr/linux-cluster/cluster/tp3/>



# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tuning
  - Job array
- Best practices
- Support



### One user = one account

You are responsible of the damage caused by your login.

### Default permissions directories

- **home, save and work** : **R**ead, **W**rite, **eX**ecution for the owner, **R**ead and **eX**ecution for all (drwxr-xr-x)

To change permissions: **chmod** command

### **Cluster is a shared resource, so ... think about the others**

Try to adapt requested resources to your needs

**DO NOT run data treatment on frontal servers:**

#### **Why ?**

- overloading frontal servers slow down everyone
- overloading frontal servers can provoke crash and block everyone
- save time for the administrators to answer support requests

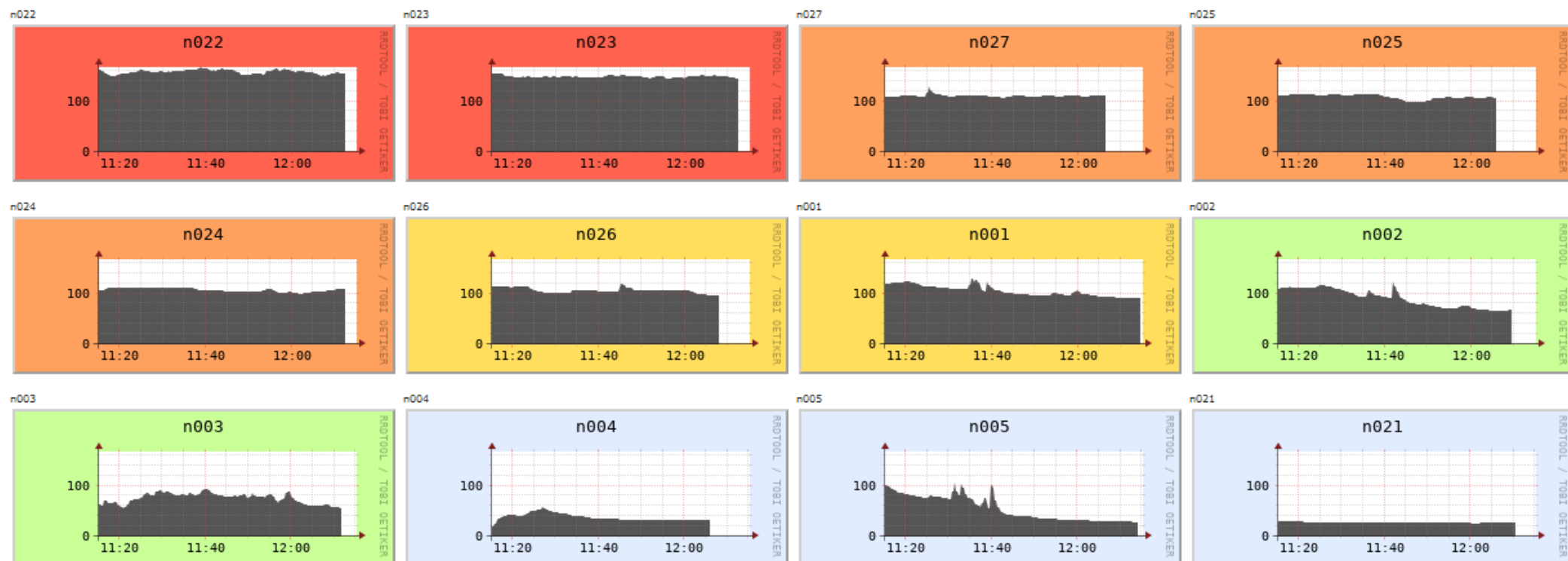
Check your process on frontal servers : **\$ pstree -u <login>**

**Any data treatment launched on the frontal servers may be killed**

# Support Cluster monitoring

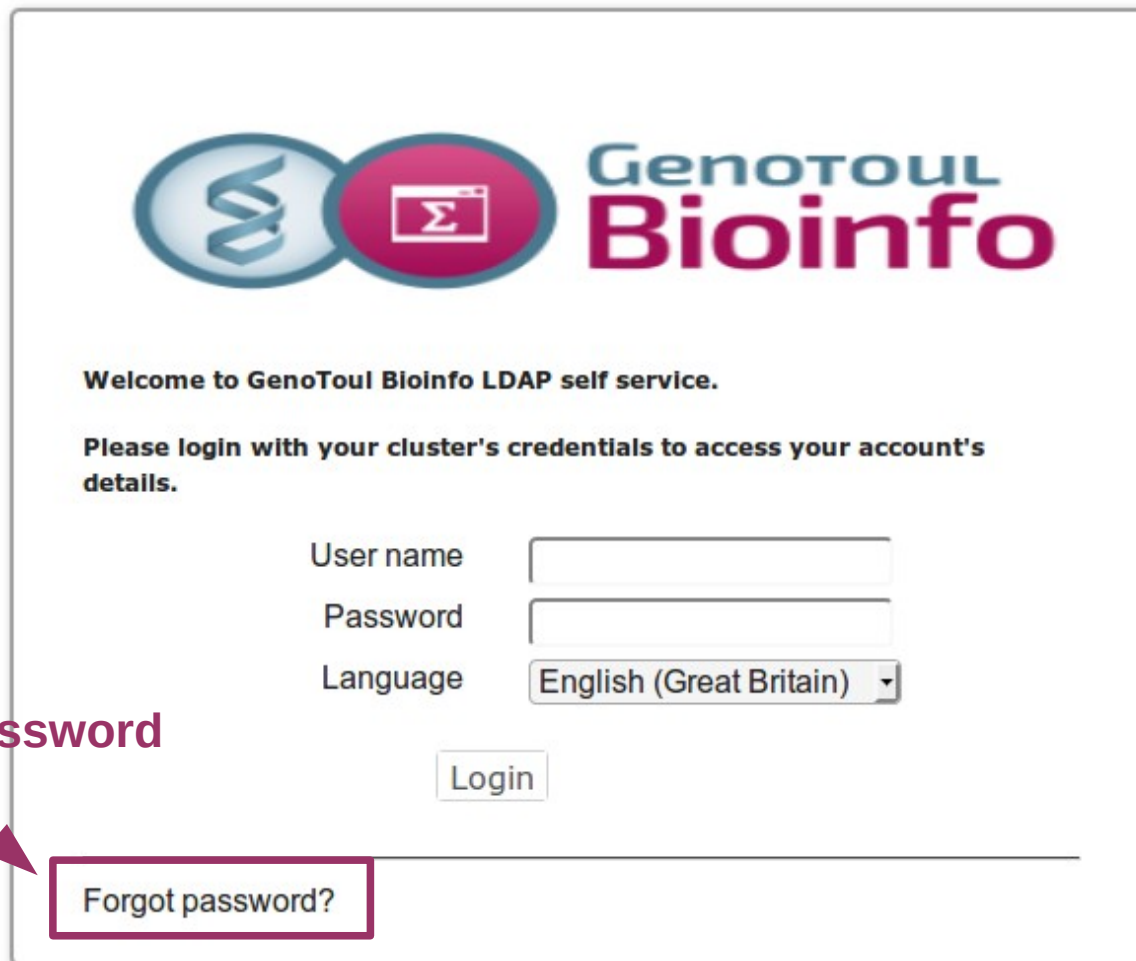
Ganglia → <http://monitoring.bioinfo.genotoul.fr>

(or our website : Resources/Monitoring)



## Account information and password change

**Self Service** → <http://selfservice.bioinfo.genotoul.fr>



The screenshot shows the login interface for the Genotoul Bioinfo LDAP self service. At the top, there is the Genotoul Bioinfo logo. Below the logo, the text reads: "Welcome to GenoToul Bioinfo LDAP self service." and "Please login with your cluster's credentials to access your account's details." The login form includes three input fields: "User name", "Password", and "Language" (a dropdown menu currently set to "English (Great Britain)"). A "Login" button is positioned below the password field. At the bottom of the form, there is a link labeled "Forgot password?".

**Change your password  
(every year)**



Forgot password?

- **Bioinfo genotoul website :**

<http://bioinfo.genotoul.fr/>

- **Bioinfo Genotoul Chart**

<http://bioinfo.genotoul.fr/wp-content/uploads/ChartPFBioinfoGenoToul.pdf>

- **FAQ**

<http://bioinfo.genotoul.fr/index.php/faq/>

- **Support**

Mail: [support.bioinfo.genotoul@inrae.fr](mailto:support.bioinfo.genotoul@inrae.fr)

Fill form (best for us): <http://bioinfo.genotoul.fr/index.php/ask-for/support/>

<https://vm-genoword.toulouse.inrae.fr/FAQ/tutorials/>

- **Conda** – How to use conda on the cluster?
- **R** – How to use R on the cluster?
- **Rstudio (OOD)** – How to use Rstudio with Open On Demand?
- **Jupyter Notebook (OOD)** – How to use Jupyter Notebook with Open On Demand?
- **Linux Desktop (OOD)** – How to use Linux Desktop (VM) with Open On Demand?

# End of Presentation

<https://sondages.inrae.fr/index.php/84236?lang=fr>

**Thanks for your attention !**



# Further informations

## More SLURM directives (+)

**--depend=[state:job\_id]**

**--odelist=[nodes]**

**--array=[array\_spec]**

**--begin=[datetime]**

**--exclusive or shared**

**#partition name (-hold\_jid)**

**#host preference (-l hostname)**

**#job arrays (-t )**

**#begin time (-a)**

**#resource sharing (-l exclusive)**

# Further informations

## SLURM variables (+)

<code>\$SLURM_JOBID</code>	the job ID
<code>\$SLURM_SUBMIT_DIR</code>	submit directory
<code>\$SLURM_SUBMIT_HOST</code>	submit host
<code>\$SLURM_ARRAY_TASK_ID</code>	job array index
<code>\$SLURM_CPUS_PER_TASK</code>	Number of cpu per task/job ( <code>#SBATCH -c</code> ), default 1

and others...

# SLURM

## Job dependencies

**sbatch -d | --dependency=<dependency\_list>**

Defer the start of this job until the specified dependencies have been satisfied completed.

<dependency\_list> is on the form <type :jobID[:jobID][,type :jobID[:jobID]]>

Example :

```
sbatch --dependency=afterok:6265 HELLO.job
```

after	this job can begin execution after the specified jobs have begun execution
afterany	this job can begin execution after the specified jobs have terminated
afterok	This job can begin execution after the specified jobs have successfully executed (ran to completion with an exit code of zero)
afternotok	This job can begin execution after the specified jobs have terminated in some failed state (non-zero exit code, node failure, timed out, etc)

# SLURM

## Sample MPI sbatch script

```
# !/bin/bash  
#SBATCH -J mpi_job           #job name  
#SBATCH --nodes=2           #2 different nodes  
#SBATCH --tasks-per-node=4  #4 tasks per node  
#SBATCH --cpus-per-task=2   #2 cpu per task  
#SBATCH --time=00:10:00     #job time limit  
  
cd $SLURM_SUBMIT_DIR  
module purge  
module load compiler/intel-2018.0.128 mpi/openmpi-1.8.8-intel2018.0.128  
mpirun -n $SLURM_NTASKS -npernode $SLURM_NTASKS_PER_NODE ./hello_world
```

# Environment

## Search/Find a soft (CLI)

### Installation paths

Bioinfo -> **`/usr/local/bioinfo/src/`**

Compilers, languages, others → **`/tools`**

Useful scripts → **`/tools/bin`** (user's default PATH)

### Commands

**`module avail`**: display all available software installed on the cluster

**`module_search <soft_name>`**: display available versions for a specific application