Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

Training day SLURM cluster

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

Context

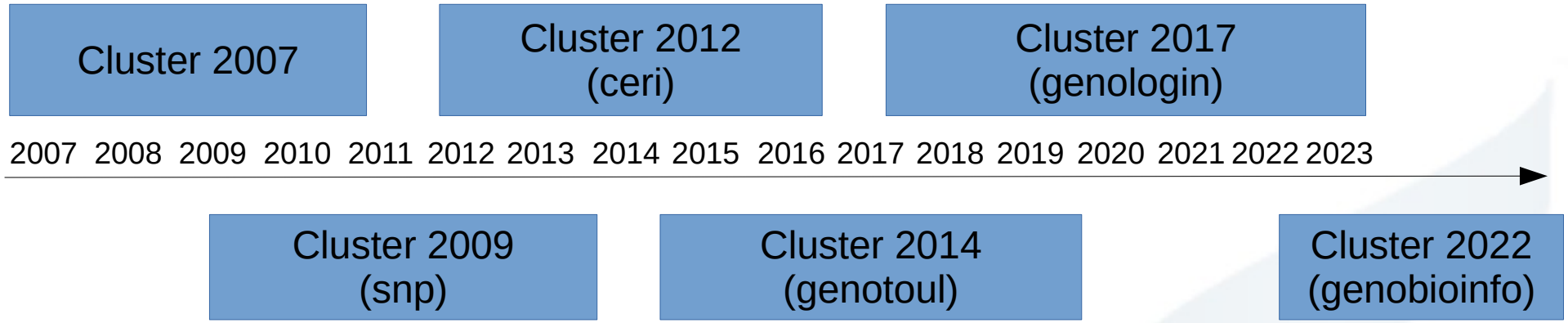
PRE-REQUISITE : LINUX

- 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)

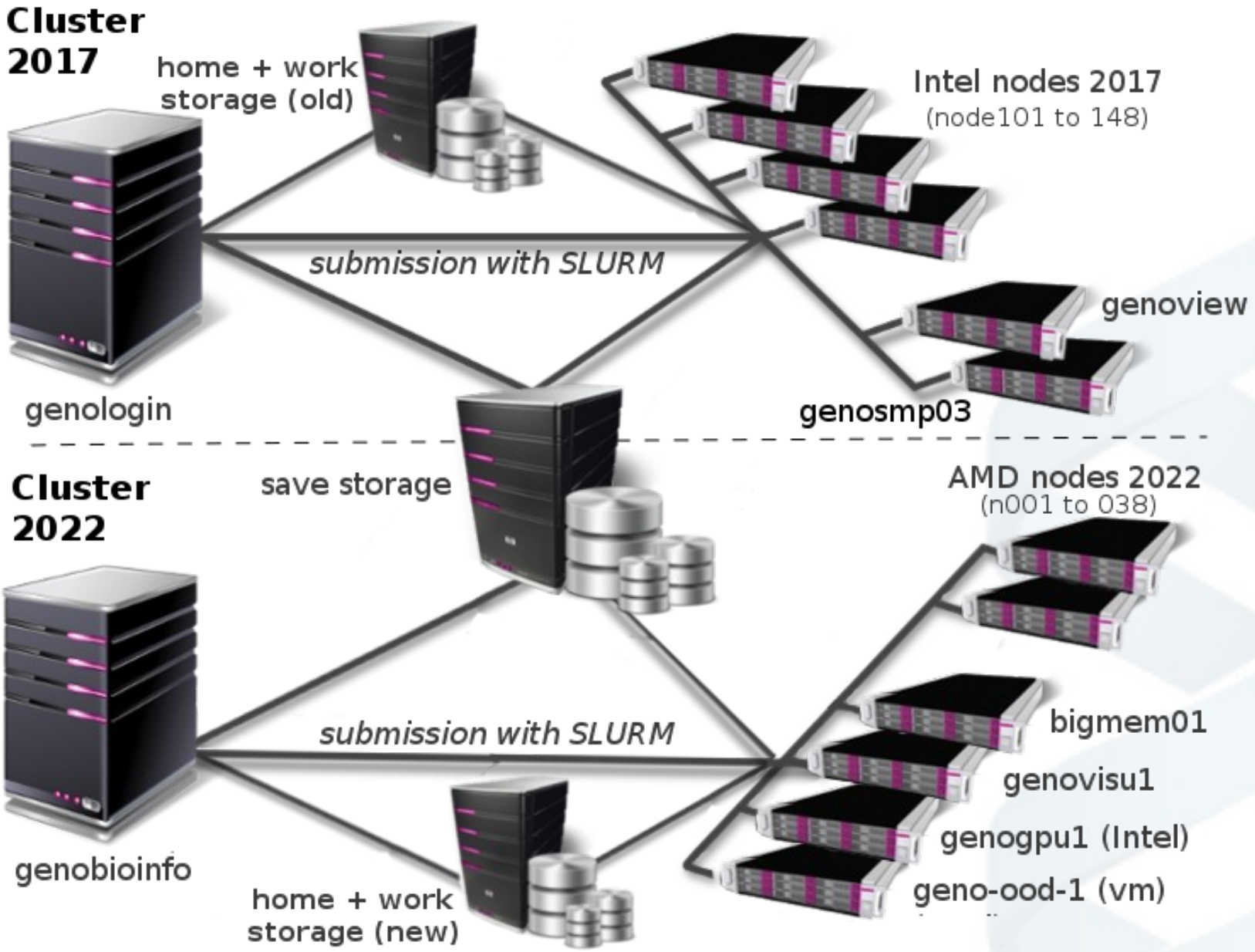
Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

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

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

Infrastructure
login & compute nodes

Compute nodes

- 38 compute nodes : [001 à 038] * (128 cores, 2048GB memory)
- n039 : (128 cores, 4096GB memory)
- GPU node : (64 cores, 1024GB memory, 10TB HD, 4*Nvidia A100/80GB)
- VISU node : (64 cores, 512GB, Nvidia RTX6000)
- 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)

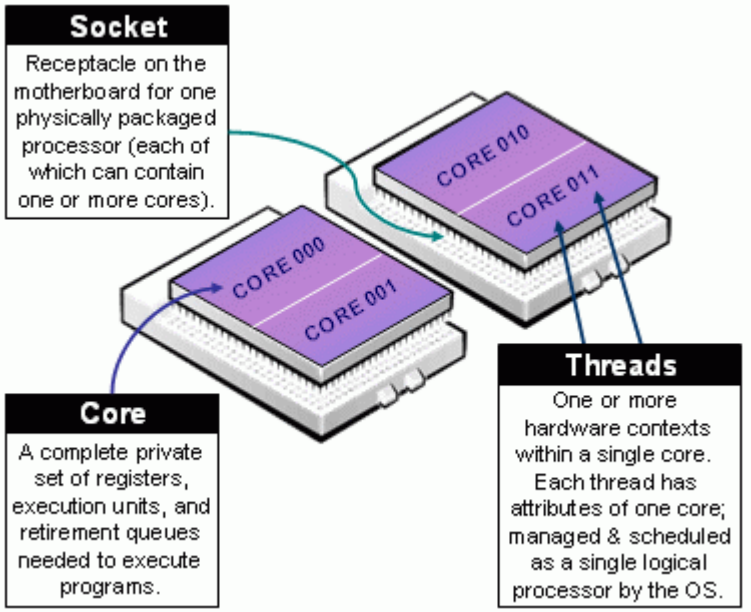
Infrastructure
Vocabulary

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

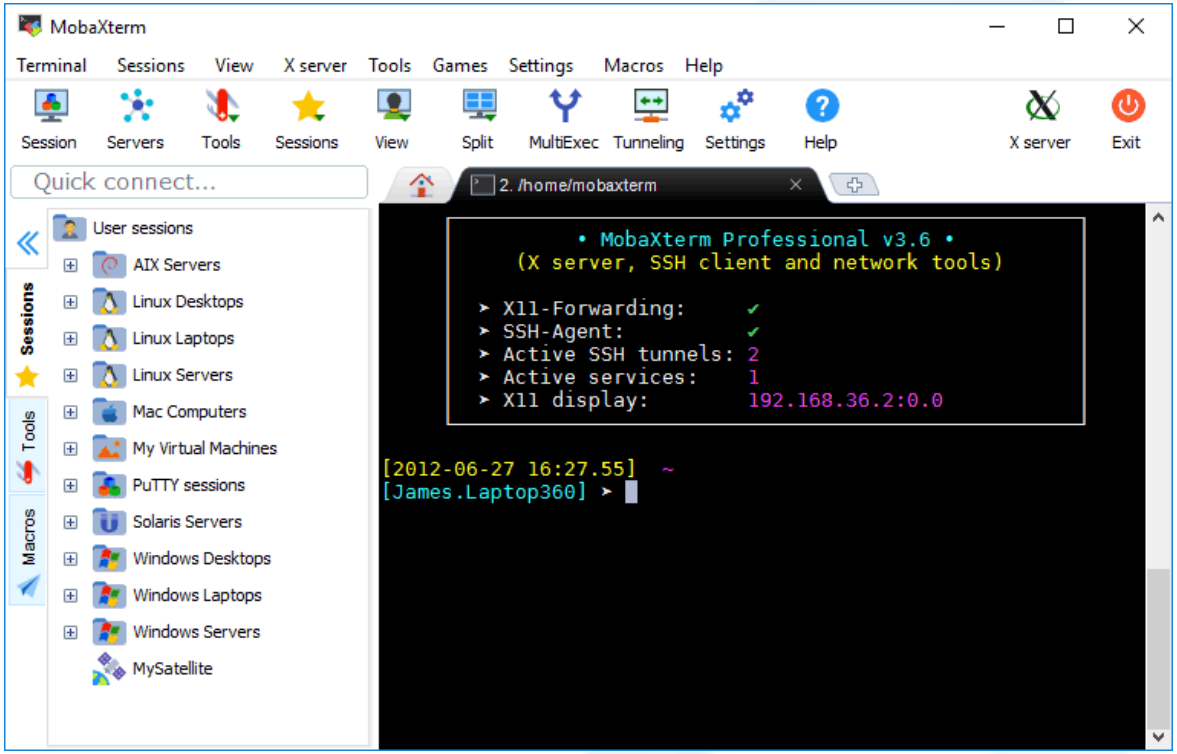


Infrastructure
User accounts

- 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)

TP 1.1

Refresh your linux knowledge:

https://genotoul-bioinfo.pages.mia.inra.fr/linux-cluster/cluster/tp1.1/

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

Training day SLURM cluster

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

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/Find a soft (Web)

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

All software

Application	Description	Availability/Use
3D-DNA	3D de novo assembly (3D-DNA) pipeline.	**Genologin Cluster:** How to use **Genobioinfo Cluster:** How to use
AAF	This is a package for constructing phylogeny without doing alignment or assembly.	**Genologin Cluster:** How to use **Genobioinfo Cluster:** Ask for Install

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

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

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tunning
  - 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 : 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/

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

Training day SLURM cluster

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

Help
Software help

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 (Avaibility/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

Help
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

Help
Other help

- **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)

http://vm-genobiotoul.toulouse.inra.fr/How_to_Softs/ (old cluster)

https://web-genobioinfo.toulouse.inrae.fr/How_to_Softs/ (new cluster)

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

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tunning
  - 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

[BATCH]

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

[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)

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 consummed ressources (time, cpu, ram)

TP 1.2

Submit your first job:

https://genotoul-bioinfo.pages.mia.inra.fr/linux-cluster/cluster/tp1.2/

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tunning
  - 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)

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

SLURM
Directives (2/2)

--cpus-per-task=4	ncpu on the same node
--mail-type=[events]	event notification
--mail-user=[address]	user email
--export=[ALL|NONE|variables]	copy environment
--workdir=[dir_name]	working directory
--wrap=”command”	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)	on demand	1	12h	128
gpuq (GPU)	on demand	1	4 days	128

SLURM
Ressources

- There are job limitations on users + group of users
- 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	4992	2000	780	500
INRAe or Occitanie	3888	1024	780	376
Others	1296	250	780	100

TP 2

Submit your first job:

https://genotoul-bioinfo.pages.mia.inra.fr/linux-cluster/cluster/tp2/

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
  - Basics
  - Tunning
  - 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 valued 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

Variable	Correspondance
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
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

Type	Correspondance
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

--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)

TP 3

Submit your first job:

https://genotoul-bioinfo.pages.mia.inra.fr/linux-cluster/cluster/tp3/

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

Training day SLURM cluster

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

Best practices
Security

One user = one account

You are responsible of the damage caused by your login.

Default permissions directories

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

To change permissions: **chmod** command

Best practices
Resources

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

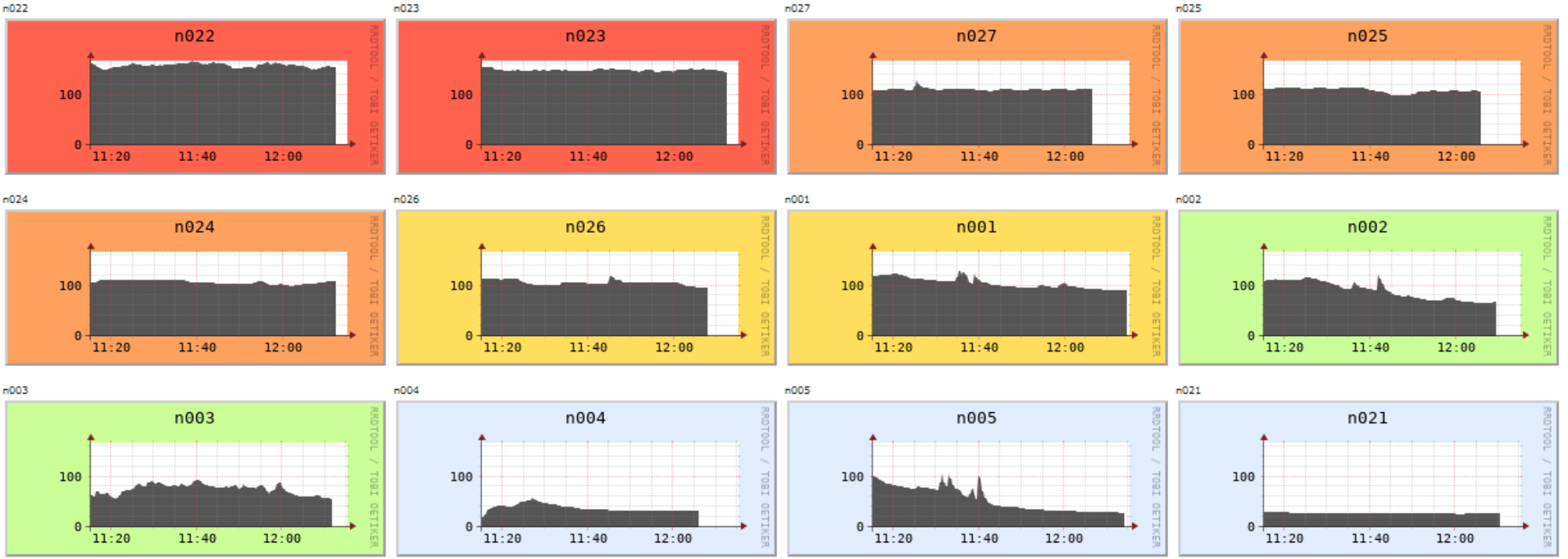
Check your process on frontal servers : **$ pstree -u <login>**

Any data treatment launched on the frontal servers may be killed

Support
Cluster monitoring

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

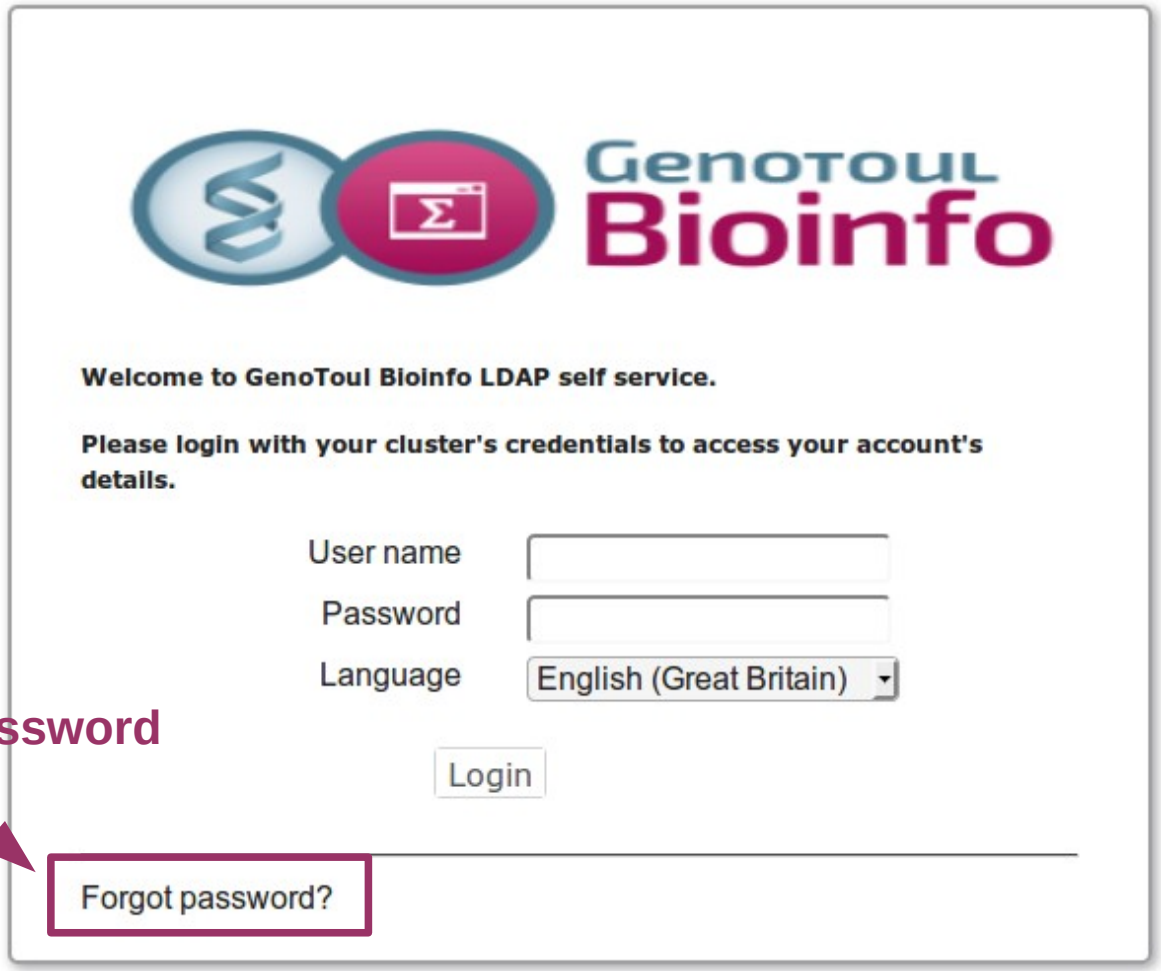
(or our website : Resources/Monitoring)



Support

Account information and password change

Self Service → http://selfservice.bioinfo.genotoul.fr



Change your password (every year)

Support

- **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

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

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]	**#partition name**	**(-hold_jid)**
--nodelist=[nodes]	**#host preference**	**(-l hostname)**
--array=[array_spec]	**#job arrays**	**(-t)**
--begin=[datetime]	**#begin time**	**(-a)**
--exclusive or shared	**#resource sharing**	**(-l exclusive)**

Further informations
SLURM variables (+)

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

and others...

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
```