Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
- TP
- For further with SLURM
- Best practices
- Support
- TP

# Context

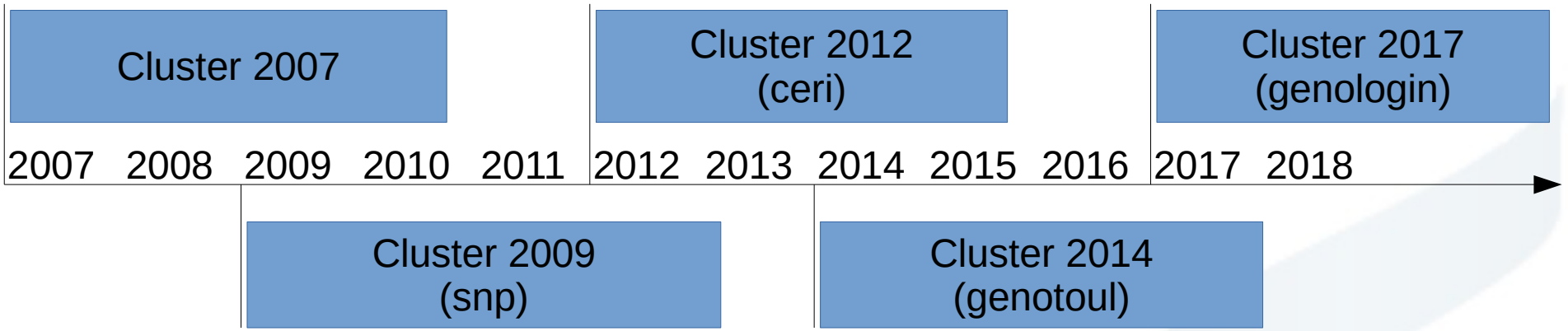
**PRE-REQUISITE : LINUX**

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

**TODAY**

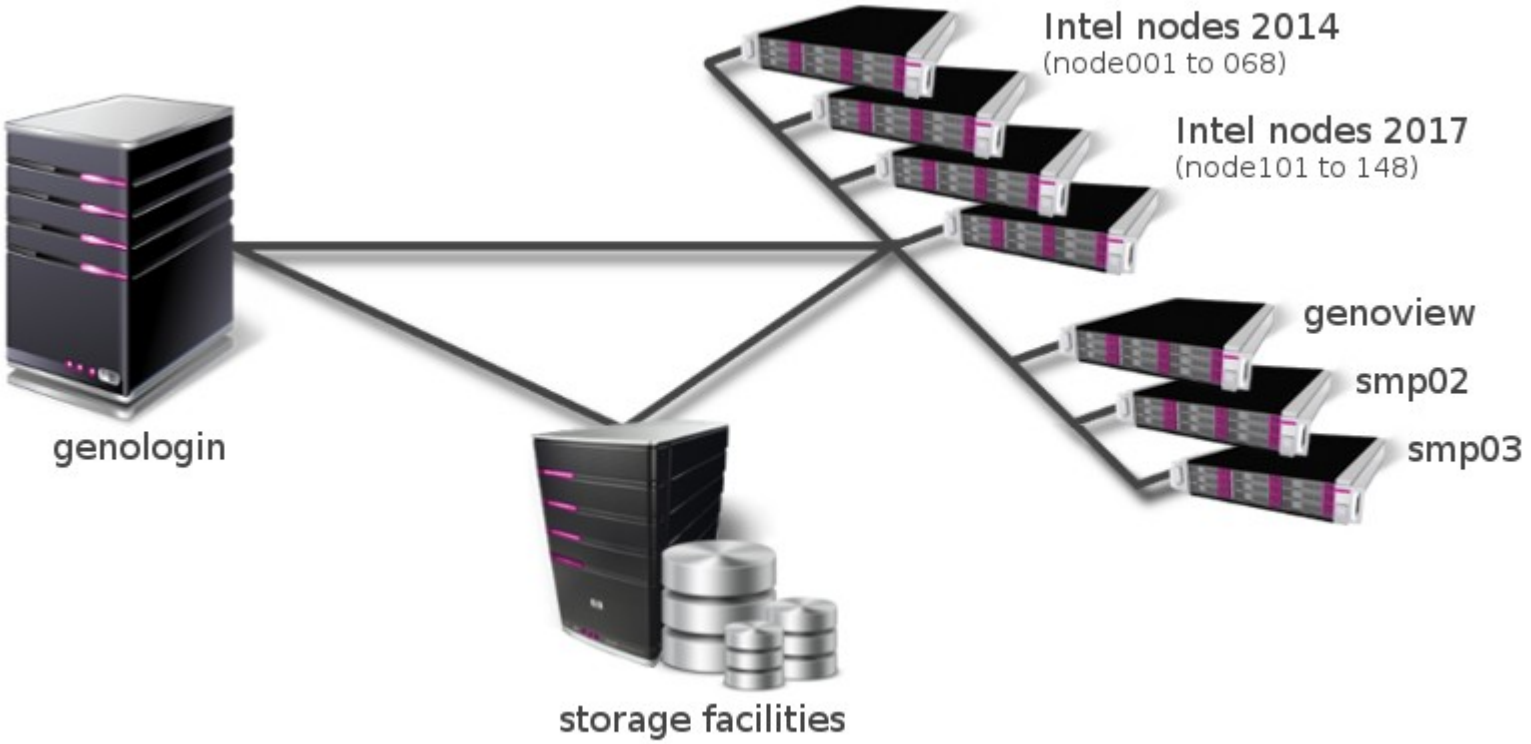
- How to use compute nodes cluster (submit, manage & monitor jobs)
- Objectives : Autonomy, self mastery

# Context
## renewal strategy




- Overlapping clusters enabling to keep the service active and to renew the machines
- But this time we have changed the job scheduler (from SGE to SLURM)
- Only SLURM at the end on 2018

# Infrastructure



# Infrastructure
## login & compute nodes

## login nodes

- 2 login nodes : genologin1&2 * (32 cores, 128 GB RAM)
- Alias : genologin.toulouse.inra.fr
- Linux based on CentOs-7 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

- 1 visualization node : genoview (32 threads, 128GB, Nvidia K40)
- 68 compute nodes : [001 to 068] * (40 threads, 256GB memory)
- 48 compute nodes : [101 à 148] * (32 threads, 256/512GB memory)
- genosmp02 (48 threads, 1536GB memory, 20TB HD)
- genosmp03 (96 threads, 3072GB memory, 20TB HD)
- Low latency &  high bandwidth interconnection (56GB/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

### CPU / Core / Threads

- Cpu : Central Processing Unit (socket)
- Core : multi-core in a CPU
- Threads : nb of parallel execution into a cpu/core (multi-threading)

# Infrastructure
## User accounts

- Access to the platform: via a command line SSH connection (putty or MobaXterm for Windows)

frontal/login servers: genologin1 & 2

alias for the connection: genologin.toulouse.inra.fr

- Example

ssh <login>@genologin.toulouse.inra.fr

# Infrastructure
## Disk spaces

- All of directories are the same between genologin servers & cluster nodes
- you don’t have to copy anything between cluster nodes & genologin
- Examples :

/home, /save, /work : user directories
/bank : international genomics databanks
/usr/local/bioinfo : Bioinformatics software

# Infrastructure
## User quotas

- **1GB** for ***/home*** directory (configuration files only)

- **250GB (*2)** for ***/save*** directory (permanent data, with replication)

- **1TB** for ***/work*** directory (temporary compute disk space)

Be careful : /work directory might be purged (120 days without access)

- **100,000H** annual **calculation time** (500H for private user)

You could have more time on demand (resource request)

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
- TP
- For further with SLURM
- Best practices
- Support
- TP

# 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:

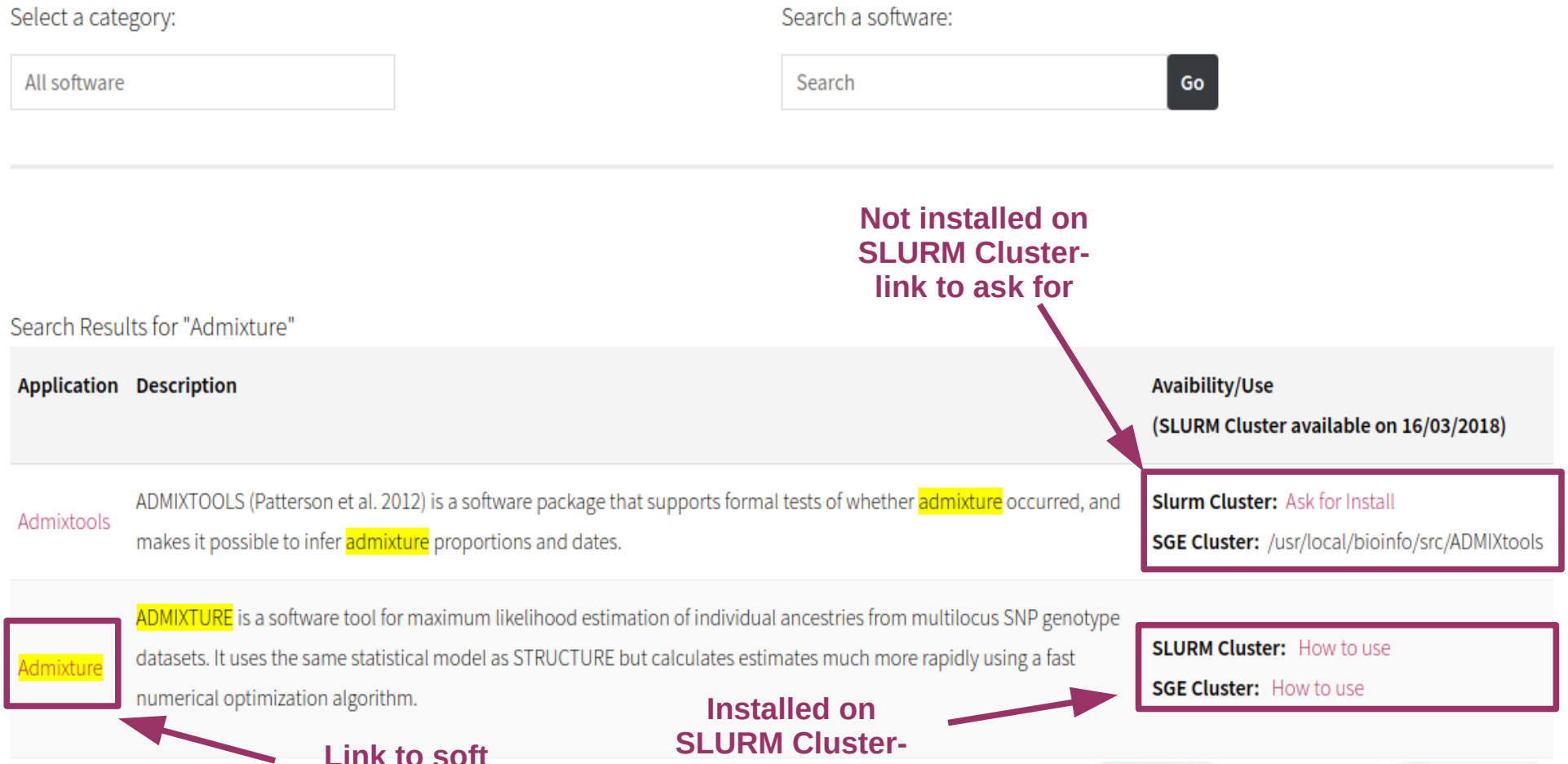
- add command in your PATH
- define specific environment variable
- add path to dependencies
- add path to specific librairies

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/



# Environment
## Search/Find a soft (CLI)

### Installation paths

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

Compilers → /tools/compilers

Libraries → /tools/librairies

Others system tools → /tools/others_tools

Langages (Python, R , Java..) → /tools

Useful scripts → /tools/bin (sarray, squota_cpu, saccount_info…). In user's default PATH.

### Commands

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

**module avail <category/soft_name>**: display available versions for a specific application (with category in bioinfo,compiler,mpi or system) (case sensitive)

**search_module <soft_name>**: display available versions for a specific application (case insensitive)

# Environment
## Search examples

**module avail bioinfo/cutadapt**

----------/tools/share/Modules ----------

bioinfo/cutadapt-1.14-python-2.7.2 bioinfo/cutadapt-1.14-python-3.4.3

**module avail -t 2>&1 | grep -i blast**

bioinfo/blast-2.2.26

bioinfo/ncbi-blast-2.2.29+

bioinfo/ncbi-blast-2.6.0+

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
- TP
- For further with SLURM
- Best practices
- Support
- TP

# 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_v0.19.0

**module load bioinfo/Bismark_v0.19.0**

Need bowtie or bowtie2 and samtools, so :

**module load bioinfo/bowtie2-2.3.3.1**

**module load bioinfo/samtools-1.4**

**module load bioinfo/Bismark_v0.19.0**

**which bismark**

/usr/local/bioinfo/src/Bismark/Bismark_v0.19.0/bismark

**bismark --help**

### Use Python-2.7.2

**module load system/Python-2.7.2**

**which python**

/tools/python/2.7.2/bin/python

**python --help**

# Software usage
## Module command

The basic command to use is **module**:

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

**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
- TP
- For further with SLURM
- Best practices
- Support
- TP

# 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: 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
## Software help (2/2)

HOW TO USE ON SLURM CLUSTER

```
SOFT : samtools
------
Site du soft: http://samtools.sourceforge.net
------------

LICENSE:
--------

The MIT/Expat License

See software documentation for more informations.

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

**Software informations**

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

-> Version v0.1.19

module load bioinfo/samtools-0.1.19

-> Version 1.3.1

module load bioinfo/samtools-1.3.1

-> Version v1.4

module load bioinfo/samtools-1.4

or use absolute path
```

**Usage and versions**

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

/usr/local/bioinfo/src/samtools/example_on_cluster

To submit:

sbatch test_v1.4.sh
```

**Example**

# Help
## Help example

### With Admixture on our website Software page


### With Bowtie in command line

**$ ls /usr/local/bioinfo/src/bowtie/**

bowtie-1.2.1.1 bowtie-1.2.1.1-linux-x86_64.zip bowtie2-2.2.9 bowtie2-2.3.3.1
bowtie2-2.3.3.1-linux-x86_64.zip example_on_cluster How_to_use_SLURM_bowtie

**$ ls /usr/local/bioinfo/src/bowtie/example_on_cluster/**

errot.txt example lambda_virus.1.bt2 lambda_virus.2.bt2 lambda_virus.3.bt2
lambda_virus.4.bt2 lambda_virus.rev.1.bt2 lambda_virus.rev.2.bt2 output.txt
test_v2-2.2.9.sh

# Help
## Other help

- **Find "How_to_use_SLURM_<soft_name>" file path**

**$ module help bioinfo/bowtie2-2.2.9**

---------- Module Specific Help for 'bioinfo/bowtie2-2.2.9' ----------

See How_to_use file: /usr/local/bioinfo/src/bowtie/How_to_use_SLURM_bowtie

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

http://vm-genobiotoul.toulouse.inra.fr/How_to_Softs/

- **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
- TP
- For further with SLURM
- Best practices
- Support
- TP

# SLURM
## System evolution

### SLURM

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

### CentOS-7

- Community ENTerprise Operating System
- Supported by IBM Spectrum Scale
- Cgroups (Control Groups) compatible

# SLURM
## Commands (1/2)

## 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)

  For the first time, create your public key as below (onto genologin server)

  $ ssh-keygen

  $cat .ssh/id_rsa.pub >> .ssh/authorized_keys

- **runVisuSession.sh** : submit a TurboVNC / VirtualGL session with the graphical node (interq partition). Just for graphics jobs.

# SLURM
## Commands (2/2)

### Job holding

- **scontrol** hold : job hold
- **scrontol** release : job release

### Job monitoring

- **sinfo** : display nodes, partitions, reservations
- **squeue** : display jobs and state
- **sacct** : display accounting data
- **scontrol show** : get informations on jobs, nodes, partitions
- **sstat** : show status of running jobs
- **sview** : graphical user interface

# SLURM
## Default parameters

- workq partition
- 1 thread
- 2GB RAM memory
- 100,000H annually compute time (more on demand)
- 10,000: max jobs for all users
- 2500: max jobs per user inside the queue
- 2500 : max tasks array per job

# SLURM
## Sample 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 (email address is LDAP account's)
#Purge any previous modules
module purge
#Load the application
module load bioinfo/ncbi-blast-2.2.29+
# My command lines I want to run on the cluster
blastall ...
```

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
- TP
- For further with SLURM
- Best practices
- Support
- TP

# 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]** | #default LDAP account's |
| **--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 | Priority | Max time | Max threads |
|---|---|---|---|---|
| workq | everyone | 100 | 4 days (96h) | 3072 |
| unlimitq | everyone | 1 | 180 days | 500 |
| interq (runVisusession.sh) | on demand | | 1 day (24h) | 32 |
| smpq | on demand | | 180 days | 96 |
| wflowq | specific software | 200 | 180 days | 3072 |

# SLURM
## Ressources

- It depends on your genotoul linux group : contributors / INRA or REGION / others.
- There are limitations on user + group of users
- It is the same thing for the RAM memory (1 thread = 4GB)

| Partition / max threads | workq (group) | workq (user) | unlimitq (all) | unlimitq (user) |
|---|---|---|---|---|
| contributors | 3072 | 768 | 500 | 125 |
| Inra or region | 2304 | 576 | 500 | 94 |
| others | 768 | 192 | 500 | 31 |

# 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
```

# 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, status of your Linux primary group in Slurm (contributors, inraregion or others), groups' members, some Slurm limitations of your account : cpu and memory limit, 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 job (cpu, memory)

# 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 (+)

| | | |
|---|---|---|
| **$SLURM_JOBID** | **#jobID** | **($JOB_ID)** |
| **$SLURM_SUBMIT_DIR** | **#submit directory** | **($SGE_O_WORKDIR)** |
| **$SLURM_SUBMIT_HOST** | **#submit host** | **($SGE_O_HOST)** |
| **$SLURM_NODELIST** | **#node list** | **($PE_HOSTFILE)** |
| **$SLURM_ARRAY_TASK_ID** | **#job array index** | **($SGE_TASK_ID)** |

| | |
|---|---|
| **$SLURM_NNODES** | **(#SBATCH -N)** |
| **$SLURM_NTASKS** | **(#SBATCH -n)** |
| **$SLURM_NTASKS_PER_NODE** | **(#SBATCH –task-per-node)** |
| **$SLURM_CPUS_PER_TASK** | **(#SBATCH -c)** |

Plateforme Bioinfo Occitanie Toulouse

Genotoul Bioinfo

# Training day SLURM cluster

- Context
- Infrastructure
- Environment
- Software usage
- Help section
- SLURM
- TP
- For further with SLURM
- Best practices
- Support
- TP

# Best pratices
## 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

# Best pratices
## Resources

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

- try to adapt requested resources to your needs.

**- DO NOT run treatments on frontal servers:**

**Why ?**

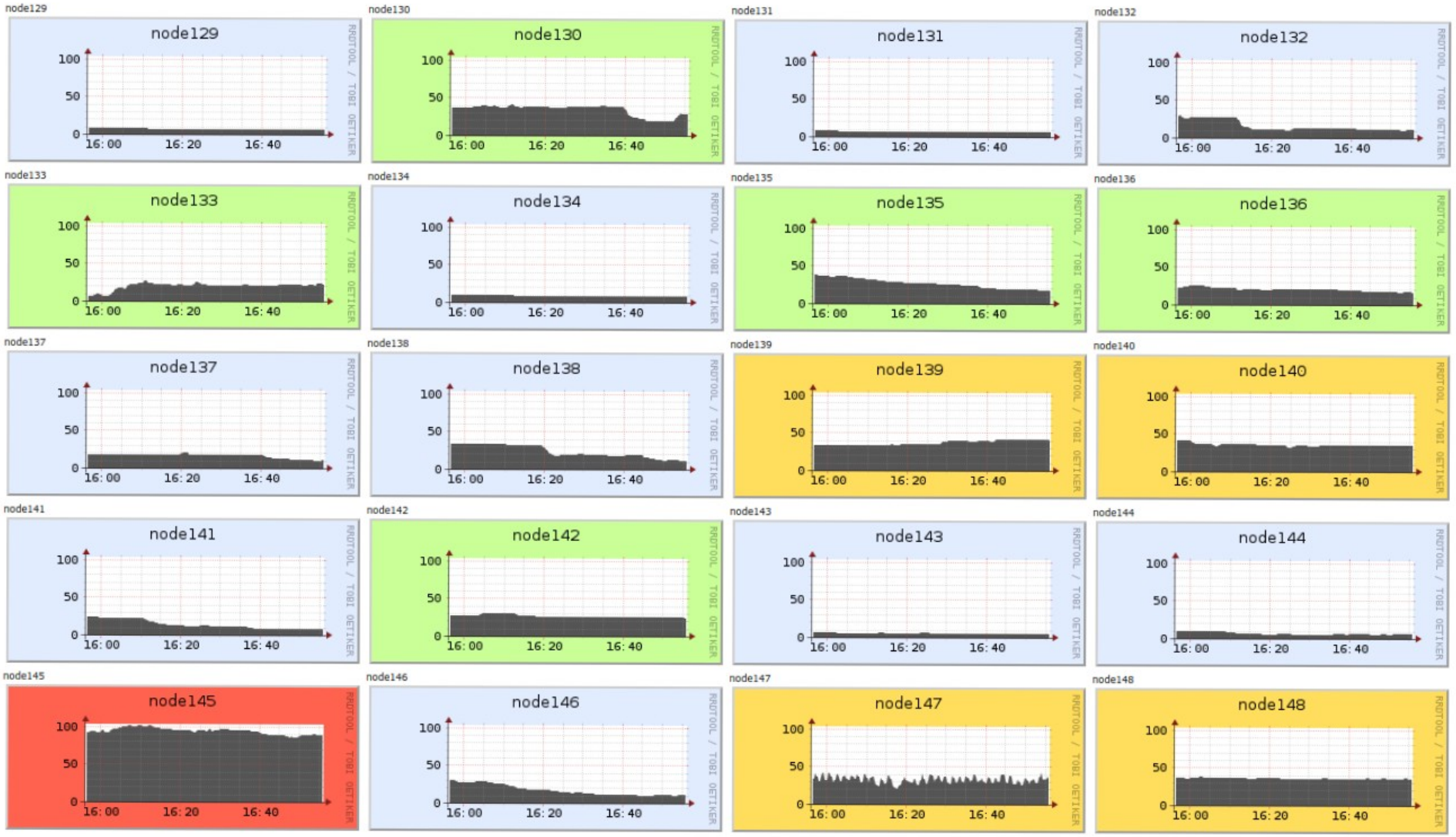
- overloading frontal servers slow down everyone.
- overloading frontal servers can crash frontal servers and block everyone.
- more time for the administrators to answer support requests.

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

**Any treatment launched on the servers "genologin" will be immediately killed by the system administrators**

# Support
## Cluster monitoring

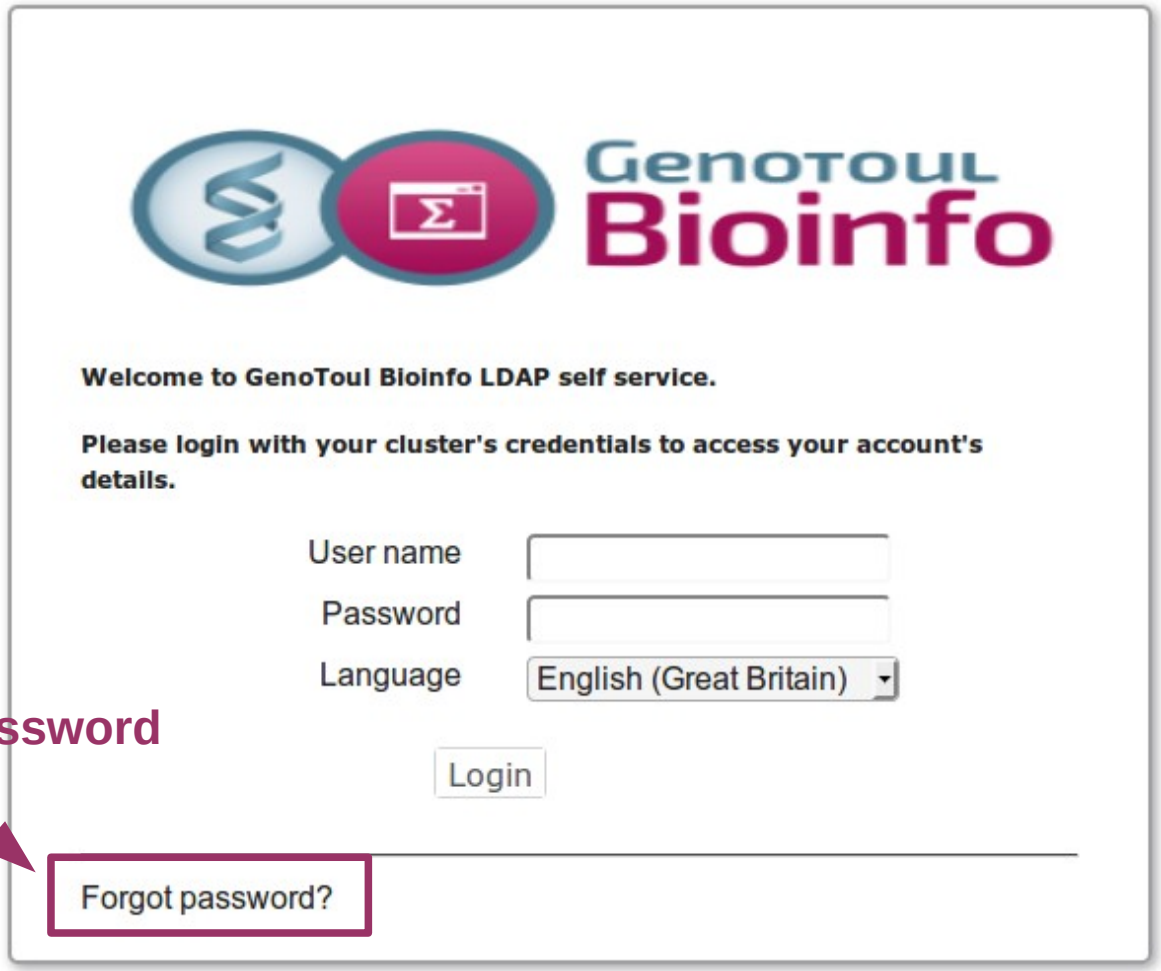
**Ganglia** → http://monitoring.bioinfo.genotoul.fr



# 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@inra.fr

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

# End of Presentation

**Thanks for your attention !**