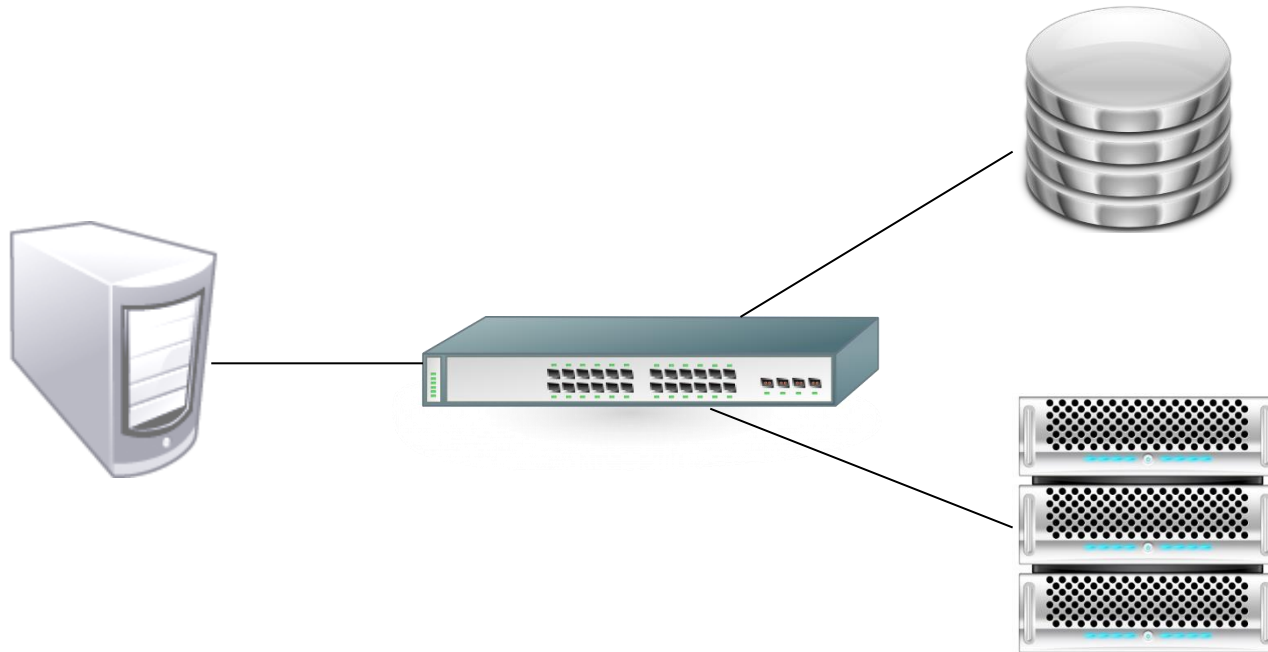


# Présentation des équipements HPC@LR

Comment soumettre son premier job

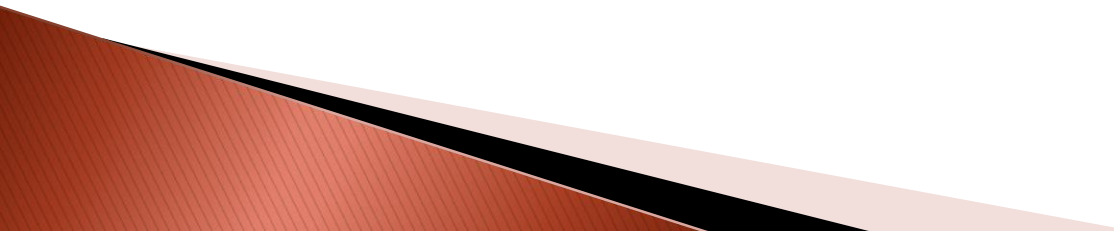
# Qu'est ce qu'un cluster



# Le cluster muse

- ▶ 308 nœuds de calcul Dell PowerEdge C6320
  - Bi processeur Intel Xeon E5-2680 v4 2,4 Ghz
  - 128 Go RAM par nœuds
  - 280 Tflops Linpack
- ▶ 1 Po de stockage rapide
- ▶ 300 To de stockage pérenne
- ▶ Réseau d'interconnexion Intel OmniPath 100 Gb/s

# Les différents types de calcul

- ▶ Calcul distribué
  - ▶ Calcul en mémoire partagée
  - ▶ Calcul réparti
  
  - ▶ Historiquement C, C++, Fortran.
  - ▶ Aujourd'hui Python et R
- 

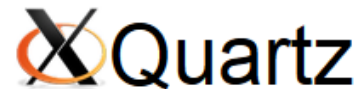
# Connexion ssh

- ▶ Depuis Linux

- [chapis@local ~]\$ ssh -X chapuis@muse-login.hpc-lr.univ-montp2.fr

- ▶ Depuis MacOS X

Pas d'interface X11 par défaut

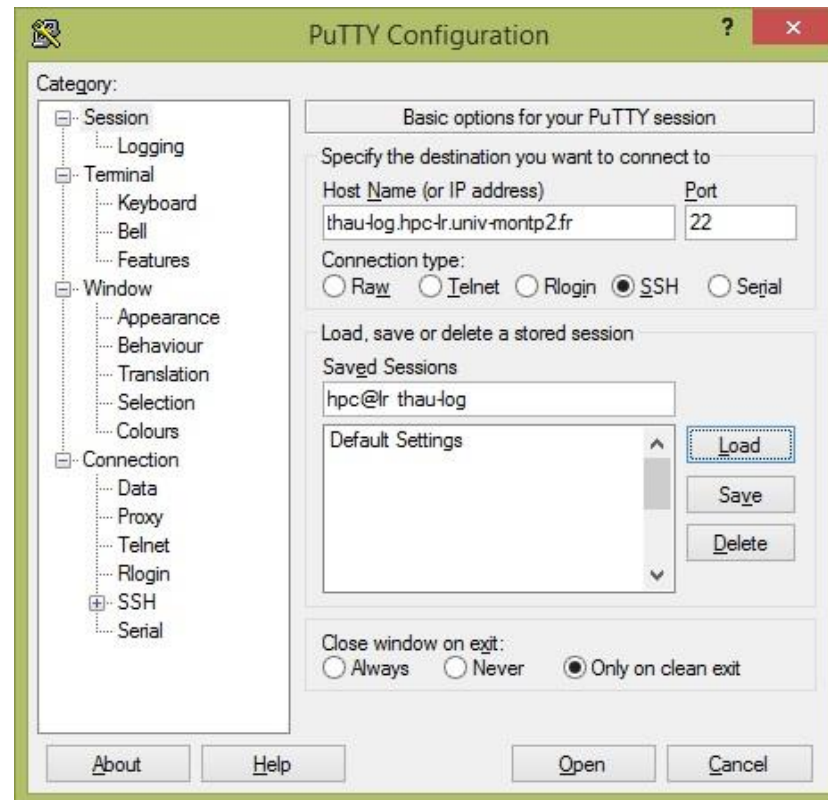


<http://xquartz.macosforge.org/landing/>

# Connexion ssh depuis Windows

## PuTTY

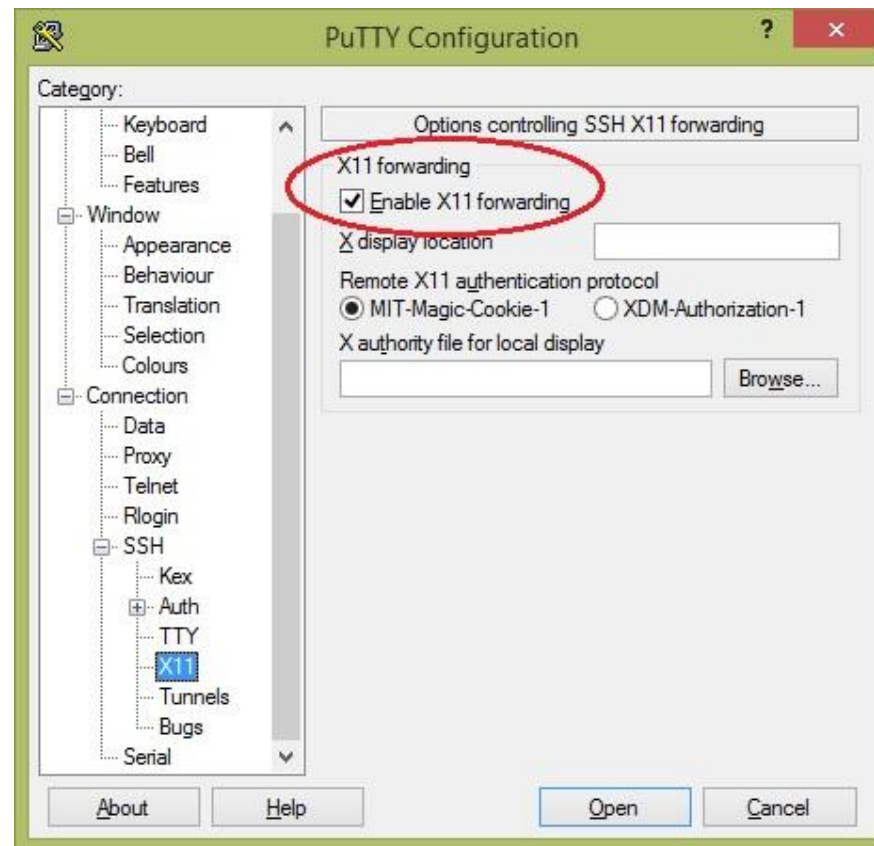
<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>



# Connexion ssh depuis Windows



<http://www.straightrunning.com/XmingNotes/>



# Environment modules

Permet de renseigner les variables d'environnement selon vos besoins (PATH, LD\_LIBRARY\_PATH...)

```
chapuis@muse-login01:~  
[chapuis@muse-login01 ~]$ module av  
----- /usr/share/Modules/modulefiles -----  
dot          module-git  module-info  modules      null          use.own  
----- /trinity/shared/modulefiles/modulegroups -----  
cv-admin     cv-advanced  cv-local     cv-standard  
----- /trinity/shared/modulefiles/cv-standard -----  
R/3.3.1                gcc/6.1.0                hdf5/openmpi/icc16/1.8.17          intel/mpi/64/2016.3.210          mvapich2/psm2/icc16/2.2rc2  
blas/3.6.0             gdb/7.11                 hwloc/1.11.2                     intel/runtime/32/2016.3.210     numpy/py27/1.11.2  
boost/1.61.0(default)  git/2.9.3               intel/advisor/32/2016.3.210       intel/runtime/64/2016.3.210     openblas/0.2.18(default)  
boost/icc16/1.61.0     hdf5/1.8.17             intel/advisor/64/2016.3.210       intel/tbb/32/2016.3.210        openmpi/2.0.1  
boost/impi/icc16/1.61.0 hdf5/gcc49/1.8.17(default) intel/clock/64/2016.3.210        intel/tbb/64/2016.3.210        openmpi/icc16/2.0.1  
boost/mvapich2/1.61.0  hdf5/gcc53/1.8.17       intel/clock/mic/2016.3.210        intel/vtune/32/2016.3.210      openmpi/psm2/2.0.1  
boost/mvapich2/icc16/1.61.0 hdf5/gcc61/1.8.17       intel/compiler/32/2016.3.210      intel/vtune/64/2016.3.210      openmpi/psm2/gcc49/2.0.1(default)  
boost/openmpi/1.61.0   hdf5/icc16/1.8.17       intel/compiler/64/2016.3.210      lapack/3.6.1                   openmpi/psm2/gcc53/2.0.1  
boost/openmpi/icc16/1.61.0 hdf5/impi/icc16/1.8.17  intel/daal/32/2016.3.210          matplotlib/py27/1.5.3          openmpi/psm2/gcc61/2.0.1  
cmake/3.6.0(default)   hdf5/mvapich2/1.8.17    intel/daal/64/2016.3.210          mellanox/efa/2.5                openmpi/psm2/icc16/2.0.1  
fftw2/2.1.5(default)  hdf5/mvapich2/gcc49/1.8.17 intel/inspector/32/2016.3.210     mellanox/hcoll/3.5             python/2.7.12(default)  
fftw3/3.3.5(default)  hdf5/mvapich2/gcc53/1.8.17 intel/inspector/64/2016.3.210     mellanox/mxm/3.4                qt/gcc/4.8.6  
fftw3/mvapich2/3.3.5  hdf5/mvapich2/gcc61/1.8.17 intel/ipp/32/2016.3.210           mvapich2/2.2rc2                 scalapack/mvapich2/2.0.2  
fftw3/openmpi/3.3.5   hdf5/mvapich2/icc16/1.8.17 intel/ipp/64/2016.3.210           mvapich2/icc16/2.2rc2           scalapack/openmpi/2.0.2  
fftw3-mvapich2/mvapich2/3.3.5 hdf5/openmpi/1.8.17    intel/itac/64/2016.3.210          mvapich2/psm2/2.2rc2           scilab/5.5.2  
fftw3-openmpi/openmpi/3.3.5 hdf5/openmpi/gcc49/1.8.17 intel/mkl/32/2016.3.210          mvapich2/psm2/gcc49/2.2rc2(default) scipy/py27/0.18.1  
gcc/4.9.3(default)    hdf5/openmpi/gcc53/1.8.17 intel/mkl/64/2016.3.210          mvapich2/psm2/gcc53/2.2rc2     valgrind/3.11.0  
gcc/5.3.0             hdf5/openmpi/gcc61/1.8.17 intel/mkl/mic/2016.3.210          mvapich2/psm2/gcc61/2.2rc2  
[chapuis@muse-login01 ~]$
```

```
[chapuis@muse-login ~]$ echo $MODULEPATH  
/usr/share/Modules/modulefiles:/etc/modulefiles:/trinity/shared/  
modulefiles/modulegroups:/trinity/shared/modulefiles/
```



# Environment modules

- ▶ Format de la commande:
  - module <paramètre> <*nom\_du\_module*>
  - avail : liste des modules disponibles
  - show : affiche les informations d'un module
  - add / load : charge un module
  - list : liste des modules chargés
  - rm / unload : décharge un module
  - purge : décharge tous les modules
- <http://modules.sourceforge.net/man/module.html>

# Le gestionnaire SLURM

Simple Linux Utility for Resource Management

- ▶ gestionnaire de ressources
- ▶ Les +
  - Open source
  - Très répandu
  - Intègre des outils de monitoring



# Le gestionnaire SLURM

SLURM	LoadLeveler	PBS / Torque
sbatch	lsubmit	qsub
squeue	llq	qstat
scancel	llcancel	qdel

# Le gestionnaire SLURM

- ▶ sbatch, srun
- ▶ squeue
- ▶ scancel <num\_job>
  
- ▶ sinfo (donne des informations sur les nœuds et sur les partitions)
- ▶ scontrol show jobid <num\_job>
- ▶ sacct, sreport

Format long	Format court	Description
--job-name	-J	Nom du job
--account		Nom du projet
--nodes	-N	Nombre de nœuds
--ntasks	-n	Nombre de tâches au total
--ntasks-per-node		Nombre de tâches par nœud (doit correspondre au nombre total de tâches divisé par le nombre de nœuds)
--ntasks-per-core		Nombre de tâches par cœurs
--partition	-p	Groupe de machines sur lequel le job va tourner

Format long	Format court	Description
--mem		Mémoire réservée par nœud en MegaBytes? (Sans cette option la totalité de la mémoire du nœud est allouée)
--time	-t	Durée maximum d'exécution du job (Passé ce délai, le job est automatiquement arrêté)
--output	-o	Nom du fichier de sorties. Par défaut slurm-%j.out ou '%j' est le numéro du job
--error	-e	Nom du fichier d'erreur. Par défaut slurm-%j.err ou '%j' est le numero du job
--mail-user		Adresse mail où envoyer les notifications

- ▶ [chapuis@thau-log thau]\$ srun -N 2 -n 2 -p defq hostname
- ▶ [chapuis@thau-log thau]\$ sbatch ./job-thau.sh

```
#!/bin/sh
#SBATCH --job-name=test
#SBATCH -N 2
#SBATCH -n 6
#SBATCH --account=projet
#SBATCH --ntasks-per-node=3
#SBATCH --ntasks-per-core=1
#SBATCH --partition=defq
```

```
echo "Running on: $SLURM_NODELIST"
mpirun ./prog_f90.exe
```

# Le gestionnaire SLURM

- ▶ sbatch, srun
- ▶ squeue
- ▶ scancel <num\_job>
  
- ▶ sinfo (donne des informations sur les nœuds et sur les partitions)
- ▶ scontrol
- ▶ sacct, sreport



- ▶ [chapis@muse-login ~]\$ scontrol show jobid <num\_job>
- ▶ [chapis@muse-login ~]\$ sreport cluster AccountUtilizationByUser user=\$LOGNAME -t hour start=2015-01-01T00:00:00
- ▶ [chapis@muse-login ~]\$ sacct -u \$LOGNAME