

Le Calcul

■ Machines interactives

- Ressources (CPU + Mémoire) partagées entre tous les utilisateurs
- Utilisation :
 - Développements de programmes
 - Visualisation de résultats
 - Traitements courts nécessitant une interaction :
 - saisie de valeurs d'entrée
 - suivi de résultats

■ Grappe de calcul ou ferme de calcul ou cluster

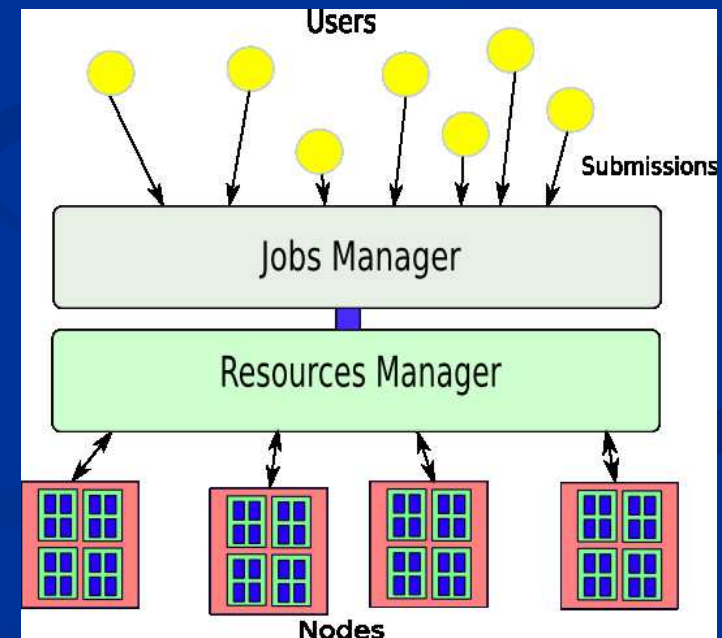
- Ressources allouées à un programme (1 job = au moins 1 CPU)
- Utilisation : Tout traitement ne nécessitant pas de visualisation

■ L'ensemble des serveurs de calcul (interactif et cluster) accède à l'ensemble des espaces de stockage de Climserv

Le Cluster : Définition

Extrait de Wikipédia : On parle de **grappe de serveurs** ou de **ferme de calcul** (**cluster** en anglais) pour désigner des techniques consistant à regrouper plusieurs ordinateurs indépendants (appelés nœuds, **node** en anglais) pour permettre une gestion globale et dépasser les limitations d'un ordinateur pour :

- augmenter la disponibilité
 - faciliter la montée en charge
 - permettre une répartition de la charge
 - faciliter la gestion des ressources (CPU, mémoire, disques, bande passante réseau)
- **Un nœud maître gérant :**
 - Le suivi de l'état des nœuds du cluster (ressources manager)
 - Le gestionnaire de batches
 - L'ordonnanceur de tâche (jobs manager)
 - **Des nœuds de calcul exécutant :**
 - Les jobs soumis par les utilisateurs.



Le Cluster : La configuration

- **Nœuds Exclusifs :**
1 processeur = 1 job
- **Les queues d'exécution :** Quatre files (ou queues) ont été configurées pour la soumission des jobs en fonction de la durée des calculs :

Nom de la file	WallTime Max
- std (défaut)	6h CPU
- day	24 h CPU
- week	168 h CPU
- infini	pas de limite

- **WallTime :** Temps d'exécution total d'un job (CPU Time + IO Time + Idle Time)

Le Cluster : Les outils

- L'interface graphique **xpbs** ou la commande **qstat** permettent de surveiller la répartition des jobs sur chaque file.

The screenshot displays the xpbs1.1.12 graphical interface, which is divided into several sections for monitoring cluster resources and jobs.

HOSTS

Server	Max	Tot	Que	Run	Hld	Mat	Trn	Ext	Status	PEsInUse	Select All
merlin-c.climserv	0	10	0	10	0	0	0	0	Active	-/-	detail Submit..

QUEUES Listed By Host(s): merlin-c.climserv

Queue	Max	Tot	Ena	Str	Que	Run	Hld	Mat	Trn	Ext	Type	Server	Select All
std	0	0	yes	yes	0	0	0	0	0	0	Execution	merlin-c.climserv	detail
day	0	1	yes	yes	0	1	0	0	0	0	Execution	merlin-c.climserv	
week	0	6	yes	yes	0	6	0	0	0	0	Execution	merlin-c.climserv	
parallel	0	0	yes	yes	0	0	0	0	0	0	Execution	merlin-c.climserv	
infini	0	3	yes	yes	0	3	0	0	0	0	Execution	merlin-c.climserv	

JOBS Listed By Queue(s): week@merlin-c.climserv

Other Criteria | Select Jobs

Job id	Name	User	PEs	CputUse	HaltUse	S	Queue	Select All	
10562.merlin-c.climserv	...AOP_1.pbs	vcapelle	-	01:01:29	01:01:57	R	week@merlin-c.climserv	merl	detail
10563.merlin-c.climserv	...AOP_2.pbs	vcapelle	-	01:01:25	01:01:56	R	week@merlin-c.climserv	merl	modify..
10564.merlin-c.climserv	...OP_2b.pbs	vcapelle	-	01:00:35	01:01:47	R	week@merlin-c.climserv	merl	delete..
10565.merlin-c.climserv	...OP_3b.pbs	vcapelle	-	01:00:35	01:01:44	R	week@merlin-c.climserv	merl	hold..
10566.merlin-c.climserv	...AOP_3.pbs	vcapelle	-	00:59:04	01:01:38	R	week@merlin-c.climserv	merl	release..
10568.merlin-c.climserv	doBLH1064	morille	-	00:56:59	00:57:51	R	week@merlin-c.climserv	merl	signal..
									msg..
									move..
									order

INFO

```
[02/04/08 17:21:40] '/opt/torque-2.0.0-p8/lib64/xpbs/bin/xpbs_datadump -t 30 merlin-c.climserv' .....  
[02/04/08 17:22:06] 'qstat -Q -f infini@merlin-c.climserv'...done.
```

Soumission de jobs : qsub

■ Qsub en mode interactif

```
X - [ramage@loholt1-c ~]$ qsub -I
qsub: waiting for job 10587,merlin-c,climserv to start
qsub: job 10587,merlin-c,climserv ready

[ramage@merlin5-c ~]$ cd jortime/
[ramage@merlin5-c ~/jortime]$ ./jo
jortime*                jortime-g77*
jortime-32-ifort.exe*   jortime-g95*
jortime-64-ifort.exe*   jortime-ifort*
jortime.exe*           jortime-pgf77*
[ramage@merlin5-c ~/jortime]$ ./jortime.exe
N= 1000
a-1*a et a*a-1
      1.00      1.00
      1.00      1.00
total secondes  14,7747536
user secondes   14,7537575
system secondes 0,0209960006
[ramage@merlin5-c ~/jortime]$ exit
logout

qsub: job 10587,merlin-c,climserv completed
[ramage@loholt1-c ~]$
```

Lancement en mode interactif

Message de login sur le cluster

jortime-pgf90*
jortime.sh*

Exécution d'un job

Sortie

Message de Sortie

Soumission de jobs : IDL et Matlab

```
#!/bin/sh
# Job description
#PBS -N "example"
# Resources used
#PBS -l "nodes=1;ppn=2"
#PBS -l "walltime=10:00"
# Standard error & standard output are merged in example.out
#PBS -j oe
#PBS -o "example.out"
# Sends a mail when the job ends
#PBS -m e
# Use the following command to go in your working directory (default is home)
cd $PBS_O_WORKDIR
# Job Matlab
echo ".r programme.pro" | /opt/rsi/idl/bin/idl
# Job IDL
/opt/matlab/bin/matlab -nojvm < programme.m
```

Le Calcul : Les évolutions

■ Serveurs interactifs :

Exclusion des jobs non interactifs :

→ Limitation du temps CPU à 1h30/2h00 pour tous les jobs

■ Cluster :

- Limitation du nombre de jobs simultanés sur les queues de longue durée
- Mise en place d'une politique de Fairshare : ajustement des priorités en fonction de l'historique d'utilisation des ressources

Nom de la file	WallTime Max	Nombres de jobs simultanés Max (en projet)
- std (défaut)	6h CPU	pas de limite (historique d'utilisation)
- day	24 h CPU	pas de limite (historique d'utilisation)
- week	168 h CPU	10 jobs / utilisateur
- infini	pas de limite	5 jobs / utilisateur