

June 16, 2016

LCPQ CLUSTER

David Sanchez
Anthony Scemama



lpqsv26

- **LCPQ production server.**



History

- **Phase 1 (April 2015)**
 - A new cluster is born.
 - sv3 integrated.
- **Phase 2 (September 2015)**
 - sv7 integrated, end of sv9.
 - Moonshot arrives.
 - 10 GbE everywhere.
- **Phase 3 (Q1'2016)**
 - sv6 integration.
 - End of Lustre, rise of BeeGFS.



Phase 4 - Q2'2016 - Cluster update

- **Power outage on July 26th**
- **Shutdown on July 25 for software update :**
 - Centos 6.8
 - Last BeeGFS revision
 - SLURM 16.05



LCPQ cluster's today



System software

- **GNU/Linux Centos 6.7.**
- **Linux kernel 2.6.32.**
- **Rocks clusters 6.2.**
- **Slurm 14.11.**



Storage

- **Home directory : 50GB/user. Backed up (soon).**
- **NAS : /mnt/nas/\$USER : ~15 TB.**
- **TMPDIR : /mnt/beegfs/\$USER : 32 TB.**
- **To use a private local disk, add :**
`#SBATCH --gres=ioperso`



Network

- **Large part of the cluster run only on Gb/10Gb ethernet.**
- **ex-sv6 will stay on infiniband.**



Modules

- **No module loaded by default.**
- **Available modules :**
 - module avail
- **Load a module :**
 - module load my_module
- **Unload module :**
 - module unload any_module



Modules

- **Available modules today :**

- intel14
- intel16
- compilers/gcc-4.8.4
- compilers/gcc-4.9.2
- compilers/gcc-5.1.0
- gcc-5.2.0
- gcc-5.3.0
- gcc/6.1.0
- openmpi-i386 (1.8)
- openmpi-x86_64 (1.8)
- openmpi-moonshot (1.8)
- g09/d01 (gaussian)
- openmpi/openmpi-1.10.2-ifort16.3-int64 (work with dirac)



Software collections

- **If you need python 2.7 and/or 3.3**

- `scl enable python27 bash`
- `scl enable python33 bash`

- **How to use in a job**

`source /opt/rh/python27/enable`

`source /opt/rh/python33/enable`



SLURM commands

- **Submit a job :**

`sbatch myscript`

- **Cancel a job :**

`scancel jobid`

- **Watch partition/node status :**

`sinfo`

- **How many cpus are allocated on a node :**

`scontrol show node compute-X-Y`



Partitions

- **#SBATCH -p my_partition**

- CLUSTER (default)
- amd128
- xeonv1
- xeonv2
- xeonv3
- xeonv1_mono
- xeonv2_mono
- xeonv3_mono
- sv6
- sv6_ssd
- moonshot



SLURM script example

```
#!/bin/bash
#SBATCH --job-name=my_task    # job name on squeue
#SBATCH --nodes=4             # I want 4 nodes
#SBATCH -c 16                 # I need 16 cpus per node
#SBATCH --mem-per-cpu=2048    # 2Gb wanted per core
#SBATCH -t 48:00:00           # for 2 days (wall time, not
CPU time)
#SBATCH -p xeonv3             # run on xeonv3 partition

module load intel16
./my_exec
```



SLURM - sbatch

- **How many nodes :**
#SBATCH -N or --nodes
- **How many cpus :**
#SBATCH -c or --cpus-per-task
- **How many tasks :**
#SBATCH -n or --ntasks
- **To reserve an entire node :**
#SBATCH --exclusive
- **To share a reserved node :**
#SBATCH -s or --share



SLURM - sbatch

- **Time duration**

#SBATCH -t or --time = days-HH:MM:SS

- **To validate a script without calculation :**

#SBATCH --test-only



SLURM – Interactive job

- **You can use « salloc »**
- **Same syntax as sbatch**

```
salloc -p moonshot -N 4 -c 8
```

4 nodes with 8 cpus per node on moonshot partition

- **To run the same command on multiple nodes : srun**

```
srun hostname -s
```

- **Default : take 1 cpu and all node memory (if exclusive)**



SLURM - scripts examples

- **Share your submission scripts !**

Here :

/share/apps/common/Examples_slurm/



Partitions

- **CLUSTER** shares nodes with **xeonv1**



Usefull commands

- **List empty nodes :**
 - lcpq-idle
- **Watch partition mix use :**
 - lcpq-mix partition
- **Get storage stats :**
 - lcpq-stockage
- **Get compute reservation :**
 - scontrol show node compute-X-Y



More infos

- [**http://www.lcpq.ups-tlse.fr/cluster/**](http://www.lcpq.ups-tlse.fr/cluster/)
- [**http://lpqsv26.ups-tlse.fr/ganglia/**](http://lpqsv26.ups-tlse.fr/ganglia/)



Questions ?

