

Comité des utilisateurs du Mésocentre

Antoine Mahul
David Grimbichler
Nadia Goué
antoine.mahul@uca.fr
david.grimbichler@uca.fr
nadia.goue@uca.fr

Université Clermont Auvergne, Mésocentre

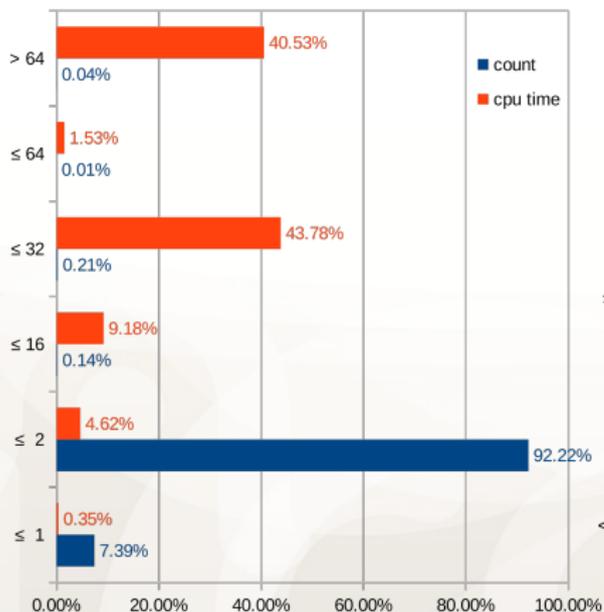
Jeudi 25 janvier 2018

- 1 Cluster de calcul HPC2
- 2 Cluster de stockage Ceph
- 3 Calculateur SMP
- 4 Cloud OpenStack
- 5 Projets

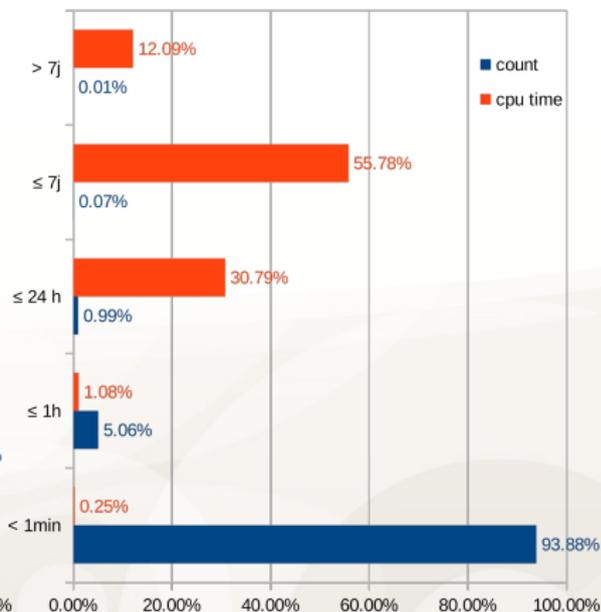
Répartition des jobs

Répartition des jobs en 2016

Répartition des jobs par coeurs



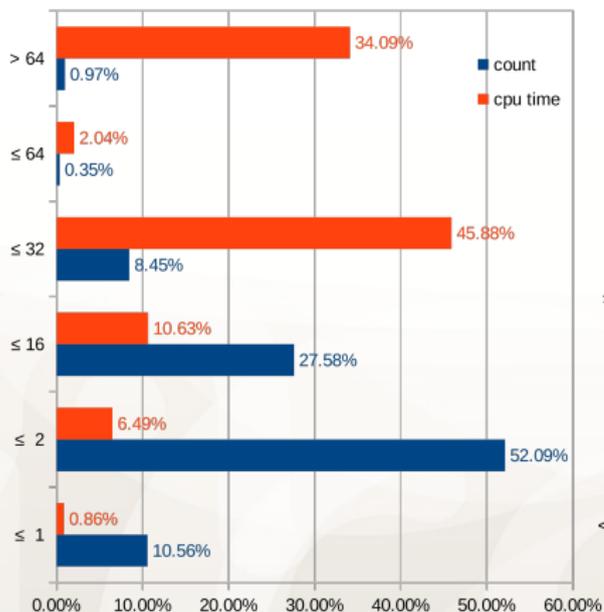
Répartition des jobs par durée



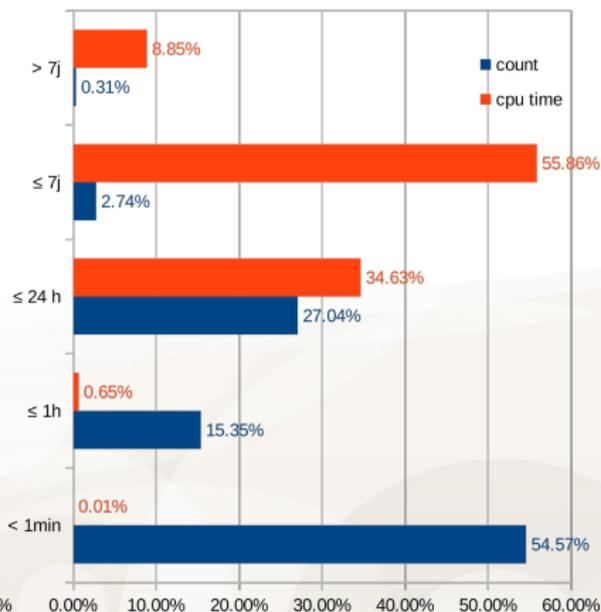
Répartition des jobs

Répartition des jobs en 2017

Répartition des jobs par cœurs

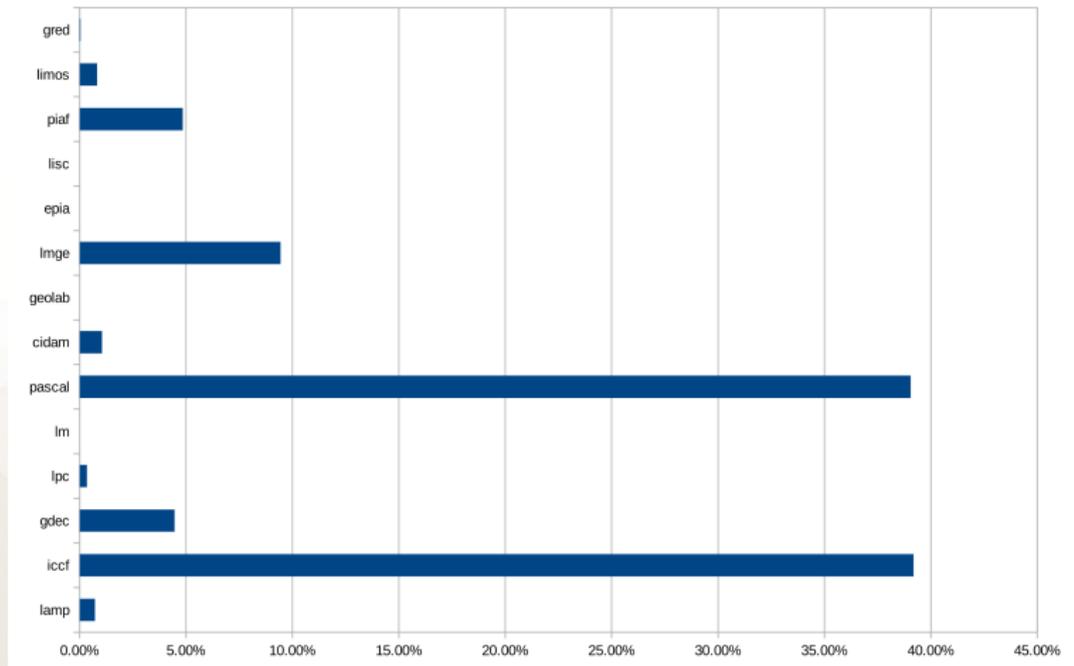


Répartition des jobs par durée

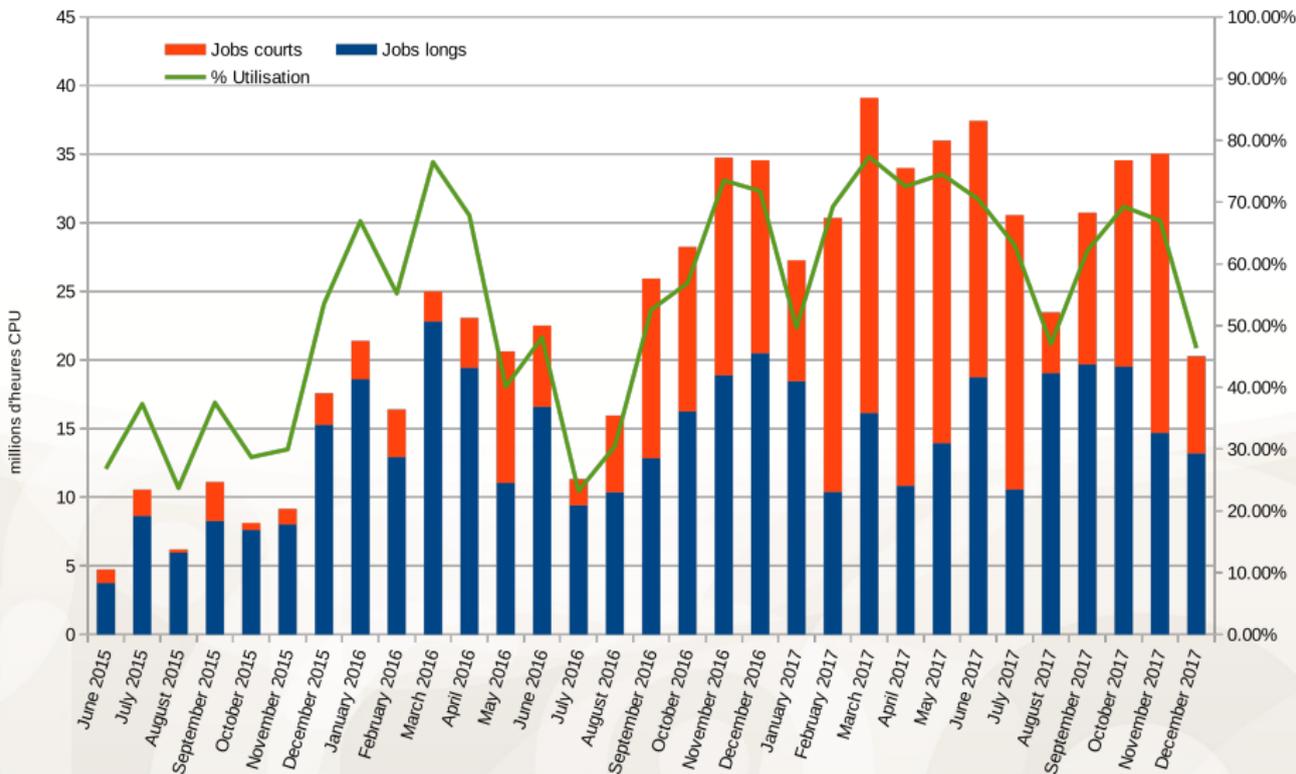


Répartition des jobs

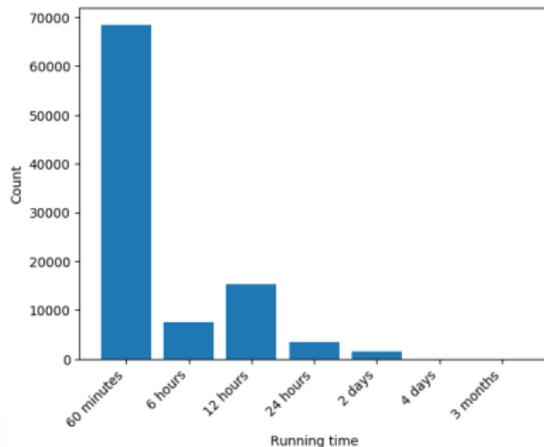
Répartition des jobs par labo en 2017



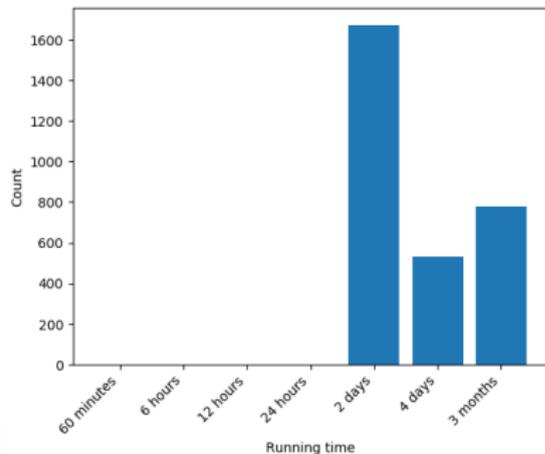
Historique d'utilisation



Profil des jobs



Jobs courts



Jobs longs

- 95 % des jobs sur la file *normal* durent moins de 12 heures
- 74 % des jobs sur la file *long* durent moins de 4 jours

→ changement des durées par défaut des files pour une meilleure rotation ?

Outils récemment installés

- Pour la bioinformatique : exonerate, ruby, finishm, Mothur, PEAR, price, metafast, FastTreeMP, trimmomatics, hmmer, fastx, ITSx, RDPTools, biom-format, IM-Tornado, PIPITS
- Pour la simulation de particules, la physique : MDAnalysis, Geant4, Castem17, openturns
- Divers outils système
- Compilateur Intel 2017 :
 - acquisition de deux jetons pour la compilation sur HPC2
 - Intel Parallel Studio XE 2017 Composer Edition
 - C++, Fortran
 - Demande des utilisateurs, en particulier pour la performance Fortran / OpenMP

Migration des noeuds

- Installation de CentOS 7 sur les noeuds de calcul
- Évolution naturelle de la version du système d'exploitation
- Motivée par l'arrivée du SMP d'emblée en RedHat 7
- Deux noeuds restants en CentOS 6 pour dépanner encore quelques temps les utilisateurs qui rencontrent des problèmes en CentOS 7
- Le frontal hpc2 (la machine de login) est encore en CentOS 6 :
 - migration prochaine
 - assez critique : on ne déploie plus de nouveaux paquets sur CentOS 6...
 - nouvelle architecture master Slurm

Migration des noeuds

- Divers problèmes réseau présents depuis plusieurs mois :
 - Se constatent sur les noeuds via des problèmes NFS, des accès lents, des jobs qui ne se lancent pas ou qui ne fonctionnent pas, ...
 - Profiter du passage en CentOS 7 pour refaire le câblage et la répartition des flux réseau
 - Investigations en cours sur des problèmes matériels au niveau des switchs fabric

Mutualisation

- Simatlab : laboratoire commun ICCF / Michelin (simulation de polymères)
 - Apport de nouvelles ressources mutualisées au Mésocentre
 - 4 nouveaux noeuds de calcul (48 coeurs logiques)
 - 1 nouveau noeud de login (avec GUI)
 - 1 nouveau noeud de stockage partagé pour le projet
 - SLURM : accès privilégié aux noeuds de calcul via partition dédiée
 - Modalités de mutualisation à définir avec l'équipe SIMATLAB
- Discussions en cours pour la mutualisation d'un noeud GPU avec le PIAF optimisé pour GROMACS (simulation dynamique moléculaire)

Taux d'occupation

Cluster Ceph : stockage capacitif données de recherche.

- 58 To de données
- dupliquées 3 fois, donc 174 To d'espace consommé
- 119 To d'espace disponible sur les noeuds en production
- 5 noeuds pas encore en production, 40 To par noeud
- Débit maximum constaté récemment : 800 Mo/s
- 38 comptes utilisateurs de 1 To

Projets stockage en cours

Pour les demandes de stockage > 1 To : mode projet

Nom	Occupation	Durée, type
Barcelona (Institut Pascal)	12 To	3 ans, plusieurs millions de fichiers
MSHArchives (MSH)	5 To	3 ans, scan d'archives
DDRAUN (MSH)	5 To	3 ans, images drones et LIDAR
GDEC	2 To	3 ans, données séquençage (bioinfo)
ANR EUREKA (LMGE)	5 To	3 ans, données de métagénomique (bioinfo)
INRA AgroDataRing (GDEC)	40 To	6 mois

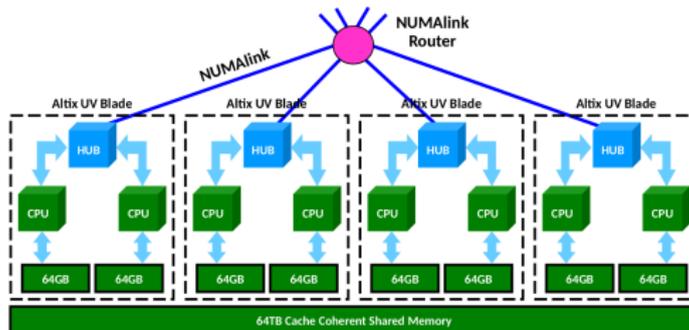
SMP : Calculateur à mémoire partagée

Le calculateur SGI UV 300 (HPE)

- financement CPER AUDACE 2015
- reçu et installé en juin 2017
- mis en place au sein du cluster en septembre 2017
- tests et benchmarks effectués en octobre et novembre 2017
- visible sur le cluster actuellement dans une file *smp* pour tester
- utilisé actuellement par quelques usagers à des fins de tests de performance et de mise au point des jobs
- nom actuel : `hpcsm01...`

SMP : Présentation de l'architecture

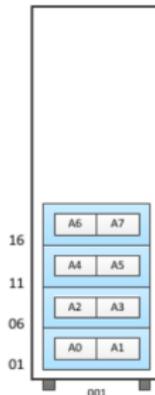
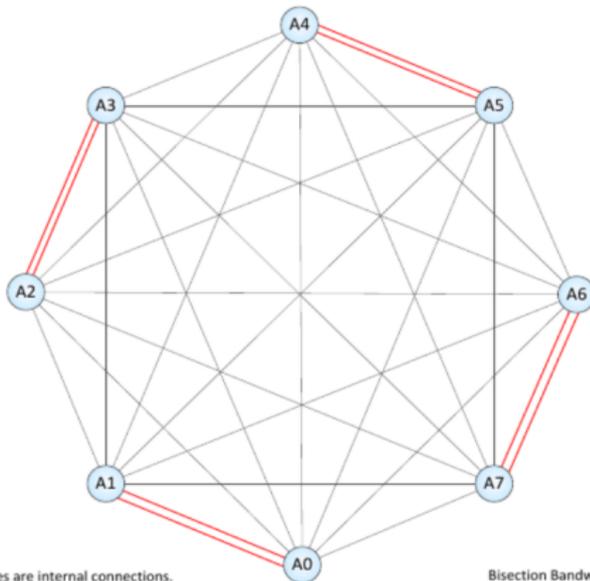
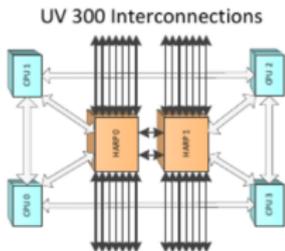
- 4 chassis de 4 sockets Intel Xeon E7-8890 v4 (Broadwell) 2,2GHz (turbo 3,4GHz)
- 768 coeurs logiques (384 coeurs physiques)
- 12 To de RAM (3 To par chassis)
- 60 To de disque en direct attach (data persistentes / scratch)
- Numalink7 : 14 GT/s (7.47 GB/s), low latency < 1 μ s



SMP : Architecture full mesh

Important: All Black lines represent two cables.

Parts:
Qty 4 UV 300 chassis
Qty 48 NL7 cables



Red lines are internal connections.

Bisection Bandwidth: 14.94 GB/s/processor
32 links are cut.

There are four open NL7 ports on each UV 300 chassis.

SMP : Proposition politique d'utilisation

■ Principes

- 1. privilégier les codes : OpenMP / MPI / BigMem
- 2. ne pas laisser le SMP inutilisé...

■ Actuellement dans SLURM :

- partition `normal` : 36 noeuds, jobs < 24h
- partition `long` : 20 noeuds, jobs < 7j
- partition `debug` : 1 noeud dédié, jobs < 1h

■ Proposition d'intégration du SMP dans SLURM :

- partition `smp` dédiée accessible à la demande uniquement (via groupe)
- intégration à la partition `normal` avec une priorité moins élevée ?
- préemption des jobs `normal` par les jobs `smp` ?
- uniquement jobs `smp` d'une certaine taille (CPU, RAM) ?

Cloud OpenStack : Matériel

- Financement CPER 2015 Audace
- Cloud IaaS pour le traitement de grands volumes de données & HTC
- Matériel livré fin décembre 2017 :
 - cluster ceph de 9 noeuds de 96 To pour le stockage en mode bloc
 - 40 noeuds compute (1600 coeurs logiques) pour l'execution VM
 - 2 noeuds redondants pour Neutron (services réseaux)
 - 6 serveurs controleurs (ceph, API, dashboard, authentification...)
 - switchs interco 10GbE

Cloud OpenStack : Avancement

- Collaboration étroite avec l'équipe LPC
- Dec 2017 : formation France Grilles sur Openstack Neutron (réseau) au Mésocentre
- Installation physique en cours
- Test de fabric IP 10GbE en cours
- Prestation d'interconnexion fibre en commande
- Q1 2018 :
 - infra de déploiement OS
 - cluster ceph dédié au stockage bloc
 - déploiement des premiers services openstack

Projets : ConnecSens

- Collaboration avec projet ConnecSens et Axe transverse Instruments CAP2025
- Instrumentation des premiers sites ConnecSens au printemps 2018
- Déploiement d'une plateforme d'expérimentation LoRaWAN sur le campus
- Déploiement des premiers services au mésocentre pour valider la reception des données LoRaWAN
- Infra d'hébergement CPER 2016 :
 - Hébergement virtualisé plutôt orienté services / data
 - 2 ESX + stockage (30 To)
- Collaboration projet Cloud environnemental (Axe 1 CAP2025)

Projets : Galaxy

Galaxy : plateforme web de traitement de données bioinformatiques

- Actuellement plusieurs instances de tests au Mésocentre
- Demande forte de la communauté bioinfo sur cet outil
- Aspect métier très important dans la gestion de l'outil
- Nov. 2017 : recrutement d'une IGR Bioinfo au Mésocentre, mutualisée pour toute la communauté bioinfo
- Jan. 2018 : formation Administrateur Galaxy organisée par ELIXIR
- ~ Mars 2018 : ouverture de la plateforme galaxy AuBi/Mésocentre aux premiers utilisateurs