CRIANN

Comité Technique HPC

CRIANN - 9 avril 2024

Comité technique HPC

Ordre du jour

- Attribution / production d’heures
- Bilan d'exploitation : mises à jour, difficultés restantes, organisation des files d’attente
- Bonnes pratiques et sécurité
- MIG sur GPU NVidia et nouveaux services pour l’IA (Jupyter et MLDE)
- Veille technologique sur GPU AMD
- Questions / Réponses

Attribution / production d’heures

Rapport de synthèse

RA 2023

- Disponible en ligne sur le site du CRIANN
  - Rubrique Présentation → Documents
  - https://www.criann.fr

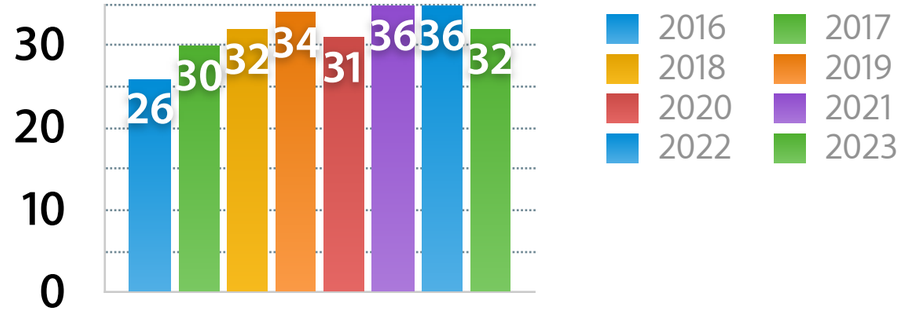
Laboratoires utilisateurs 2023

Activité ESR - Bilan annuel 2023

	Myria	**Austral**	**Total**
Attributions AAP 2023	*26 M.h.c*	*26 M.h.c*	52 M.h.c
Compta système	39 M.h.c		
Compta elapse	45 M.h.c	22 M.h.c	
Production sur GPU	160 k.h.gpu	100 k.h.gpu	

Laboratoires utilisateurs 2023

Activité ESR - calculateur Myria



Laboratoire - UMR	Heures CPU (acct)	Heures GPU	Comptes	Projets
CORIA - UMR 6614	18 277 385	8 565	54	14
LOMC - Le Havre - UMR 6294	5 455 444		12	3
COBRA - Rouen - UMR 6014	3 593 606	1 753	14	11
CIMAP Alençon - UMR 6252	2 925 708		6	4
GPM - Rouen - UMR 6634	1 934 429		14	5
LCS - Caen - UMR 6506	1 311 046	0	1	1
LMRS - UMR 6085	918 501		6	2
M2C - Caen - UMR 6143	818 256		13	5
CRISMAT - Caen - UMR 6508	720 399		5	2
M2C - Rouen - UMR 6143	470 874	2 899	6	5
ICMN - Orleans - UMR 7374	470 093		1	1
CERMN - UNICAEN	424 011	34 496	3	1
LUSAC - Cherbourg	328 346		7	1
IDEES UMR 6266	174 958		2	1
ICMR - Reims - UMR 7312	121 351		3	1
LITIS - Rouen	107 343	48 508	15	11

Laboratoire - UMR	Heures CPU (acct)	Heures GPU	Comptes	Projets
LCT - Paris - UMR 7617	96 886	0	1	1
Chrono-environnement - Besancon - UMR 6249	96 551	10 639	1	1
GREYC - Caen - UMR 6072	93 138	29 834	10	5
LRCS - Amiens - UMR 7314	80 708		3	1
GSMA - Reims - UMR 7331	70 599		1	1
Dynamicure	66 702	2 664	8	1
MEDyC - Reims - UMR CNRS 7369	64 159	10 696	1	1
LISN - UMR 9015	50 925		1	1
LMNO - UMR 6139	33 240	0	3	1
LERN - Rouen - UR 4702	19 885		2	2
GATE - UMR 5824	12 101		1	1
GSA - ENSA Paris Malaquais	6 001		1	1
LASIR - Lille - UMR 8516	4 297		1	1
ESIGELEC - Rouen - IRSEEM	737	323	1	1
LMI - Rouen - FR 3335	265		2	2
HCERES	31	16	1	1

Appel à projets scientifiques 2024

Bilan des attributions et des consommations

- Attributions du 1er AAP 2024 (avec quelques projets au fil de l'eau du 1er trimestre)
  - 58 M.h.c attribuées
  - 92 projets scientifiques

- Production académique sur le 1er trimestre : 27 M.h.c
  - Soit 46.5 % des attributions
  - **Ne pas hésiter à demander une rallonge lors du prochain AAP !**

- Production des industriels : 2M.h.c

Valorisation des heures de calcul

Publications

- Penser à mentionner l'utilisation des moyens de calcul Criann dans les publications
  - *Ce travail a bénéficié des moyens de calcul du mésocentre CRIANN (Centre Régional Informatique et d'Applications Numériques de Normandie).*
  - *Part of this work / The present work / was performed using computing resources of CRIANN (Normandy, France)*

Nouvelles règles de comptabilité

Ressources consommées

- Myria : utilisation de la comptabilité CPU du système d'exploitation
  - Indépendante des ressources *réservées*
    - Exemple : 1 calcul en mode exclusif comptabilise uniquement la consommation des cœurs actifs du calcul
- Austral : utilisation de la comptabilité de Slurm
  - *Ressources réservées* x *durée du calcul*
    - Exemple : 1 calcul en mode exclusif sur un nœud fin :
      192 cœurs/serveur X durée du calcul
  - La différence entre les 2 modes est très faible sur des calculs efficaces utilisant tous les cœurs des serveurs
  - Les ressources GPU sont comptabilisées
    - Les % d'utilisation du quota sont encore basés sur la consommation CPU

Suivi des heures consommées

Individuelle et par projet

- La commande ***maconso*** indique votre consommation individuelle
  - L'option -u permet aux responsables et responsables adjoints de suivre la consommation de tous les membres du projet

- Mise en place prochaine
  - Envoi de mail mensuel aux responsables de projet avec le détail par login

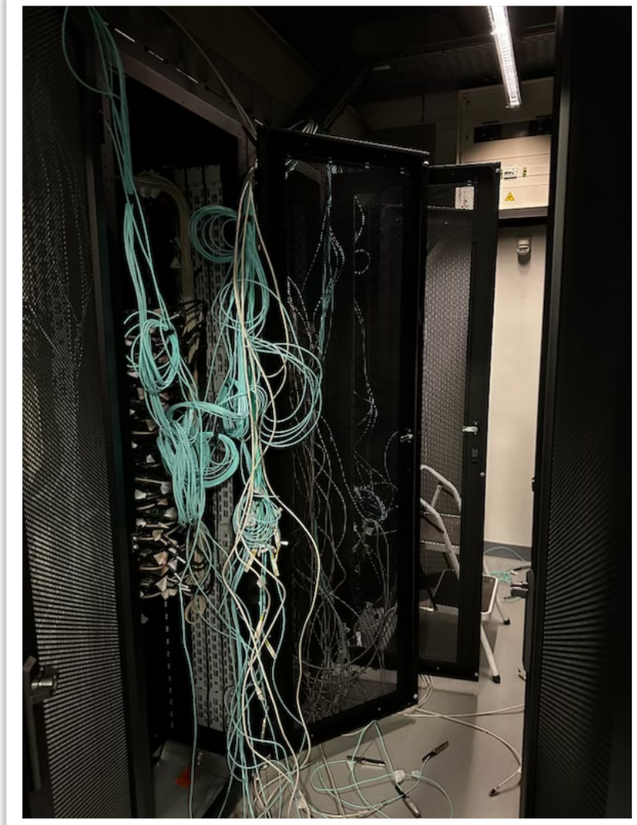
Myria

Myria c’est fini !

Bilan et suites

- Arrêt complet fin février 2024
- Bilan au Criann
  - 7 années de service [2017-2023]
    + quelques mois en 2024
  - 382 M.heures.cœur produites

Myria : démantèlement en cours



CRIANN

Austral

Volet exploitation

Arrêts de production

Depuis octobre 2023

- Du 6 février 16h au 22 février 16h
  - Maintenance MAJ HPCM 1.9
  - Durée 16 jours (vs une semaine prévue initialement !)
- Le 26 mars 2024 de 9h à 18h30
  - Arrêt non programmé (panne électrique)
  - Durée 9h30
  - Difficultés au redémarrage (bug), résolu dans la journée

Mise à jour de février 2024 : nouvelles versions

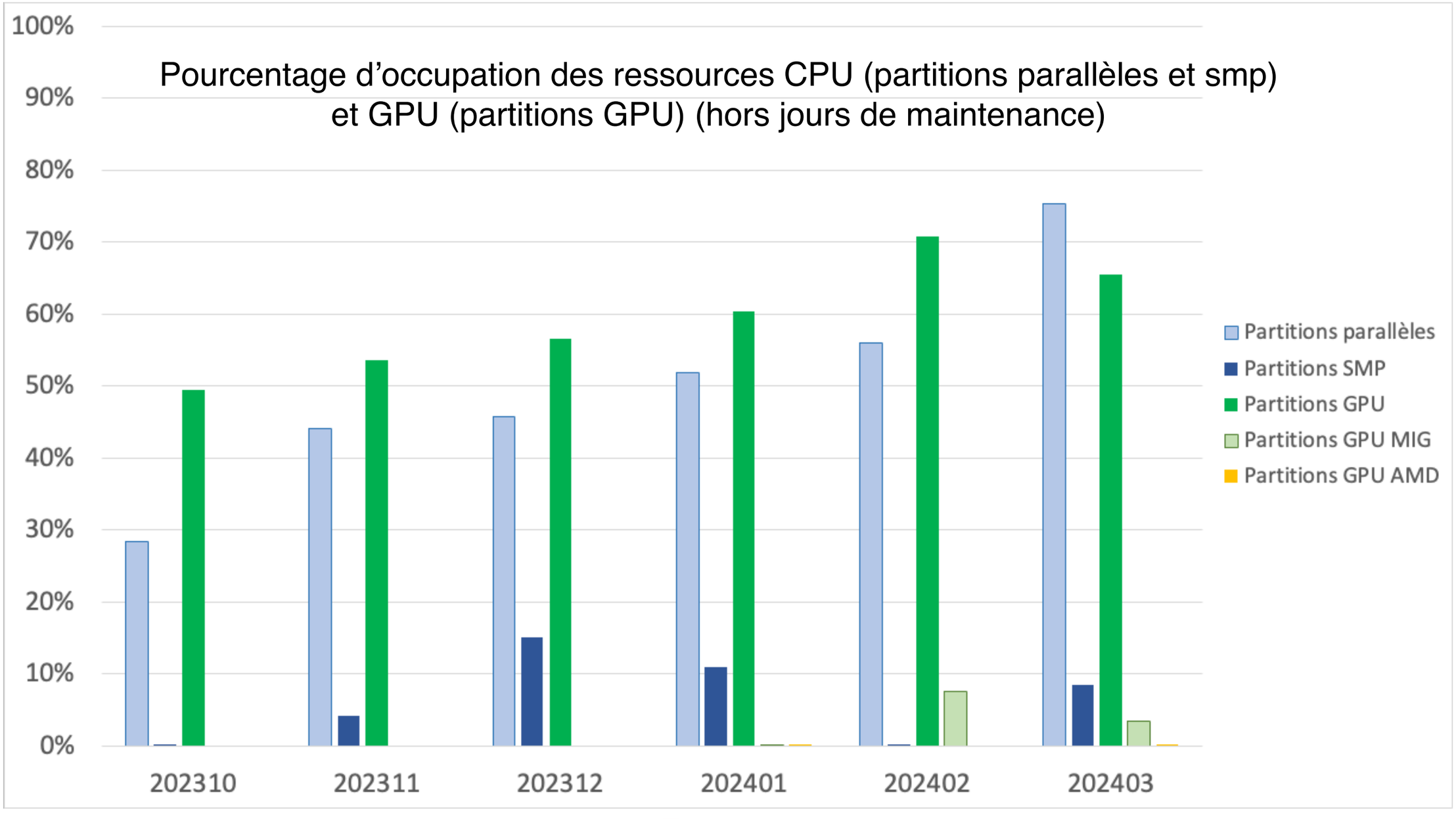
Mise à jour février 2024	Version précédente	Nouvelle version
RedHat	8.6	8.7
Slurm	22.05.10	23.02.7
Slingshot (Réseau rapide)	2.1.0	2.1.1
Nvidia	SDK : 22.7 Driver : 515.65.01 Cuda : 11.7	SDK : 23.3 Driver : 535.154.05 Cuda : 12.2

Maj de février 2024

Bilan

- Quantité de mémoire réduite par serveur (voir documentation)
  - 714 000M (3500M par coeur) pour les noeuds fin
  - 492 000M (7600M par coeur) pour les serveurs équipés de GPU Nvidia
- Nouvelles versions de compilateurs disponibles
  - La version par défaut reste la même
- Nvidia : mise à jour des drivers permettant l'utilisation de Cuda 12.2
- Prochaine mise à jour : cet été
- **Investigations en cours : performance des I/O**
  - Suite au signalement de performances irrégulières par les équipes IA
  - **En cas de performances significativement dégradées, merci de remonter vos observations au support pour aider aux diagnostics (n° job concerné)**

Charge d'occupation d'Austral



Charge d'Austral

Slurm - partage des ressources

- La charge est déjà élevée sur le ressources GPU et calcul parallèle
- Limitations par utilisateur et évolutions prévues
  - Sur GPU A100, actuellement 32 GPU simultanés
    - Passer à 16 GPU simultanés
  - Sur CPU, actuellement 8448 cœurs simultanés (44 nœuds, soit 35% de la capacité parallèle)
    - Différencier les jours de semaine (22 nœuds / 4224c) et le week-end (44 nœuds / 8448c)
- Au niveau utilisateur :
  - Consommer régulièrement
  - Utiliser les heures de week-end : soumettre le vendredi soir

THÉMATIQUE SCIENTIFIQUE	NOM DU LOGICIEL
SIMULATION ATOMISTIQUE ET OUTILS CONNEXES	CHARMM
	GROMACS
	NAMD
	MOLPRO
	VASP
	PSI4
	DALTON
	AMF
	LAMMPS
	QCORE
	ASE
	ICMR-GAUSSIAN
	Quantum Espresso
	VMD
BIOLOGIE	Augustus
	Guppy
	Dorado
MÉCANIQUE DES FLUIDES	Star CCM+
	SWASH
	MODULEF
	DUALPHYSICS
	YADE
	TELEMAC-MASCARET
	OPENFOAM
	FOAM-EXTEND
	CODE_SATURNE
MODÉLISATION ATMOSPHÉRIQUE, CLIMATOLOGIE, OCÉANOGRAPHIE	WRF - WPS
	NCL
	WGRIB
	GEOS
	GDAL
	CDO
	R_TERRA
	NCO
	SIRANE

THÉMATIQUE SCIENTIFIQUE	NOM DU LOGICIEL
MÉCANIQUE, ACOUSTIQUE	Code ASTER
	CAST3M
	HYPERWORKS
	LS-DYNA
	NASTRAN
	SALOME-MECA
MATHÉMATIQUES, STATISTIQUES	FREEFEM ++
	OCTAVE
	SCILAB
	R
	Python/dask
	Python/pandas
MACHINE LEARNING, DEEP LEARNING	PyTorch
	TensorFlow/Keras
	Horovod
	Scikit-learn
	OpenCV
MAILLAGES	GMSH
COUPLEURS	Oasis
	Precice
VISUALISATION	Paraview
	Ferret
	Xmgrace
	Molden
	Ncview

Logiciels disponibles sur Austral (mars 2024)

Austral

Logithèque

- Applications scientifiques
  - Installations à la demande, optimisées pour l'architecture
  - Possibilité d'effectuer sa propre installation

Bonnes pratiques et recommandations

Sécurité

Conséquence de la Maj

- Le tunnelling SSH est interdit vers les frontales
  - L'option "Remote SSH" ne fonctionne plus dans le logiciel Visual Studio Code
  - Alternatives :
    - Sécuriser et versioner votre code en utilisant GIT
    - Utiliser le plugin VS Code "module-rsync" (https://github.com/thisboyiscrazy/vscode-rsync) pour pousser les modifications vers le cluster
    - Utiliser les sessions Jupyter Hub
    - Utiliser sshfs pour monter votre Home-dir sur votre poste de travail

Sécurité

Quelques rappels

- Les comptes sont individuels : pas de communication de mot de passe ni d'ajout de clé SSH
- Les partages de données restreintes sont possibles : si besoin, contactez le support
- Prévenir le support quand un membre du projet quitte le laboratoire ou l'entreprise

- Les commandes `***sudo***`,`***apt-get***`, `***yum***` sont réservées à l'équipe CRIANN…

Stockage

Limitation des ressources

- L'effacement des données des anciens jobs sur /dlocal/run n'est pas encore automatisé
- Un quota par utilisateur est appliqué sur l'ensemble de la baie de disque (/home, /dlocal, /soft)
  - En cas de dépassement de quota, les jobs tombent en erreur
  - Surveiller votre utilisation
  - En cas de limitation, nous contacter

Soumission des travaux MPI

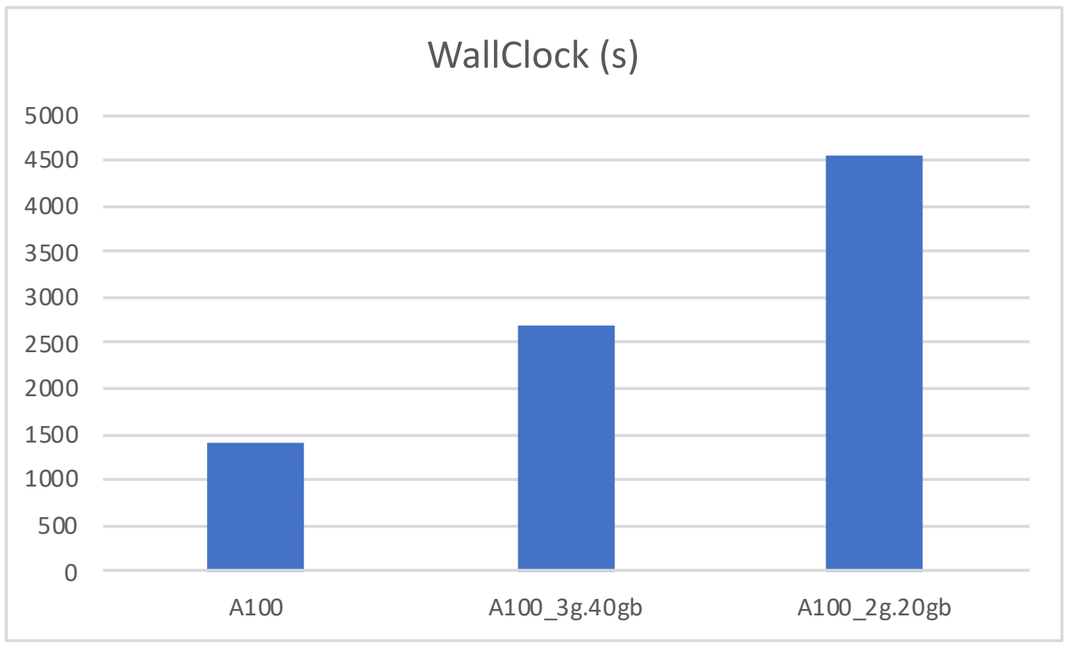
Recommandations (suite à la mise à jour)

- Les cartes de réseau rapide Slingshot ne devant pas être partagées par plus de trois travaux multi-nœuds, cette limite demande
  - Pour les travaux intra-nœud (< 192 cœurs et < 714 GB de mémoire)
    - Appliquer « #SBATCH --nodes 1 »
    - Par ailleurs, un utilisateur peut restreindre le partage d'un nœud de calcul à des travaux lui appartenant : « #SBATCH --exclusive user »
      - https://services.criann.fr/services/hpc/cluster-austral/guide/#usage-exclusif-dun-serveur

- Pour les travaux multi-nœuds (> 192 cœurs ou > 714 GB de mémoire)
  - Appliquer « #SBATCH --exclusive »
  - De plus, si l'algorithme et l'efficacité parallèle du code sur le cas traité le permettent, un nombre de tâches MPI multiple de 192 est pertinent

MIG sur GPU NVidia
Nouveaux services pour l'IA
(Jupyter et MLDE)

	A100-SMX—80GB
Stream Multiprocessors	108 (128 cores / SM)
Tensor Core	432 (4 TC / SM)
Bandwith	2 TB/s
Interconnect	NVLink 600GB/s
Flops DP	9.7 TFlops / 19.5 Tflops TC
Flops SP	19.5 TFlops / 156 Tflops TC
Flops HP	312 Tflops with TC



Bench Pytorch (Resnet50)

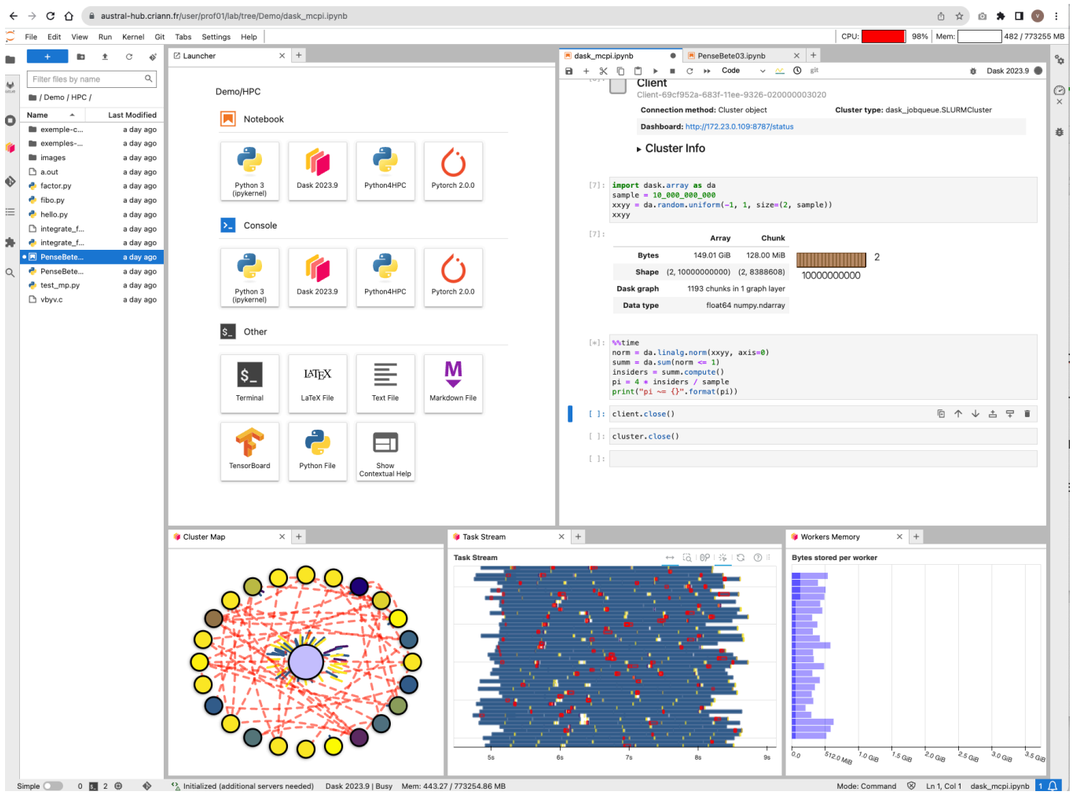
GPU MIG

Multi-Instance GPU

- Découpage 1 GPU Nvidia A100 jusqu'à 7 GPUs distincts
- Augmentation du nombre de devices GPU
- Chaque device bénéficie des tensor cores
- /!\ 1 seul device par processus

Sur Austral

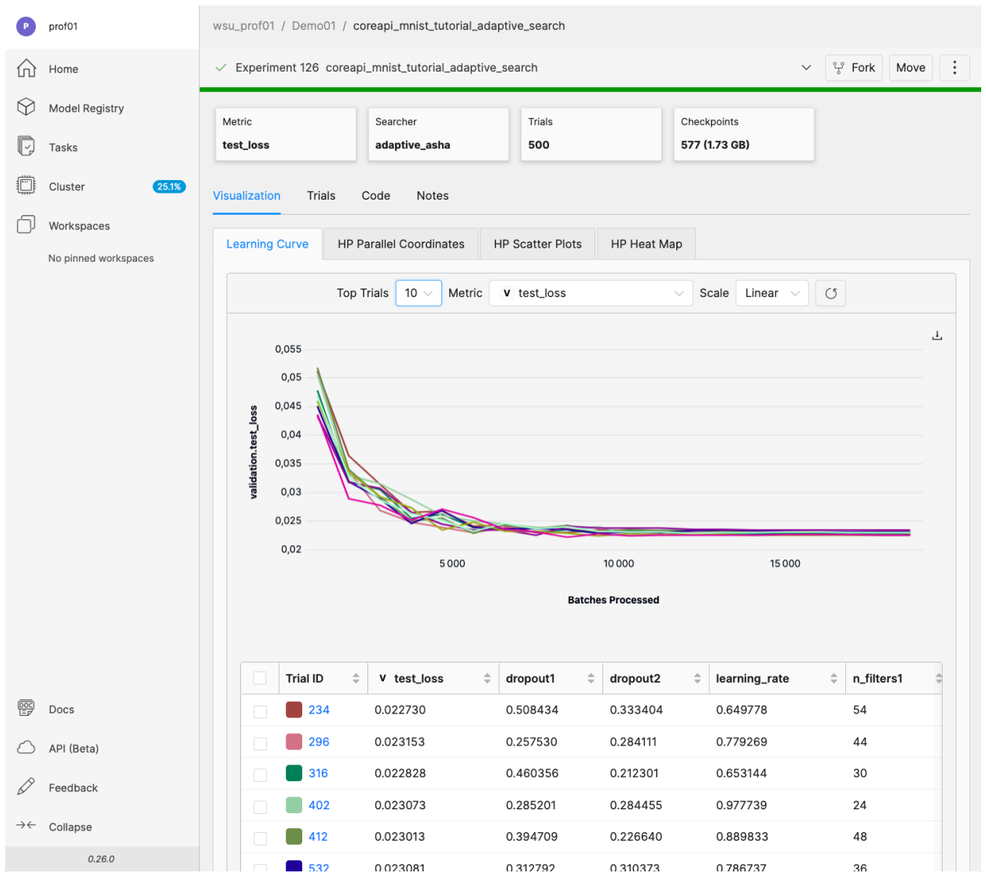
- https://services.criann.fr/services/hpc/cluster-austral/guide/#gpus-mig
- 1 serveur HPDA de 8 GPUs => 31 device au total
  - 10 devices a100_1g.10gb avec 10 GB de Mémoire, 14 SM et 56 TC
  - 17 devices a100_2g.20gb avec 20 GB de Mémoire, 28 SM et 108 TC
  - 4 devices a100_3g.40gb avec 40 GB de Mémoire, 42 SM et 164 TC
- Objectif : plus de devices pour les petits travaux
- Partition `hpda_mig`



Nouveaux services

Jupyterhub

- https://services.criann.fr/services/hpc/cluster-austral/guide/jupyter/
- Application web de développement
  - Jupyter notebook, tensorboard, dask dashboard, …
  - connexion
    - https://austral-hub.criann.fr
    - login & password identique à ceux de la connexion ssh
  - accès aux environnements disponibles sur Austral
  - Accès à une partie des ressources d'Austral
    - via Slurm => pas nécessairement immédiat
  - Penser à clore le serveur jupyterlab



Nouveaux services

MLDE - Machine Learning Development Environment

- Environnement collaboratif HPE pour les travaux d'IA
  - TensorFlow, PyTorch
  - Distribution des tâches sur les ressources de calcul via Slurm
  - Suivi des entrainements
  - Optimisation des hyperparamètres
  - Entrainement distribué
  - CLI et interface web pour la gestion interactive des travaux
  - https://austral-mlde.criann.fr

Veille technologique sur GPU AMD

GPU AMD sur Austral : 2 serveurs

4 GPU MI210, 2 x (32 cores) AMD Milan CPU, 256 GB RAM

- MI210
  - 22,6 TFlops FP64 (x2 avec «Matrix cores»), 1,6 TB/s bande passante mémoire
  - 64 GB HBM2
- Pilote et toolkit AMD ROCM 5.7.1
- Technologies de programmation supportées
  - HIP (analogue CUDA) en C/C++, compilateurs AMD ou Cray
  - Directives « OpenMP target » en C/C++/FORTRAN, compilateurs AMD ou Cray
  - Directives OpenACC en FORTRAN, compilateur Cray

- Partition Slurm : amdgpu
- Demande d'informations : support@criann.fr

GPU AMD

Veille technologique

Code source C, Stream benchmark

OpenMP target

Copy

```
#pragma omp target teams distribute parallel for
      for (j=0; j<STREAM_ARRAY_SIZE; j++)
          c[j] = a[j];
#endif
```

Scale

```
#pragma omp target teams distribute parallel for
      for (j=0; j<STREAM_ARRAY_SIZE; j++)
          b[j] = scalar*c[j];
#endif
```

Add

```
#pragma omp target teams distribute parallel for
      for (j=0; j<STREAM_ARRAY_SIZE; j++)
          c[j] = a[j] + b[j];
#endif
```

Triad

```
#pragma omp target teams distribute parallel for
      for (j=0; j<STREAM_ARRAY_SIZE; j++)
          a[j] = b[j] + scalar*c[j];
#endif
```

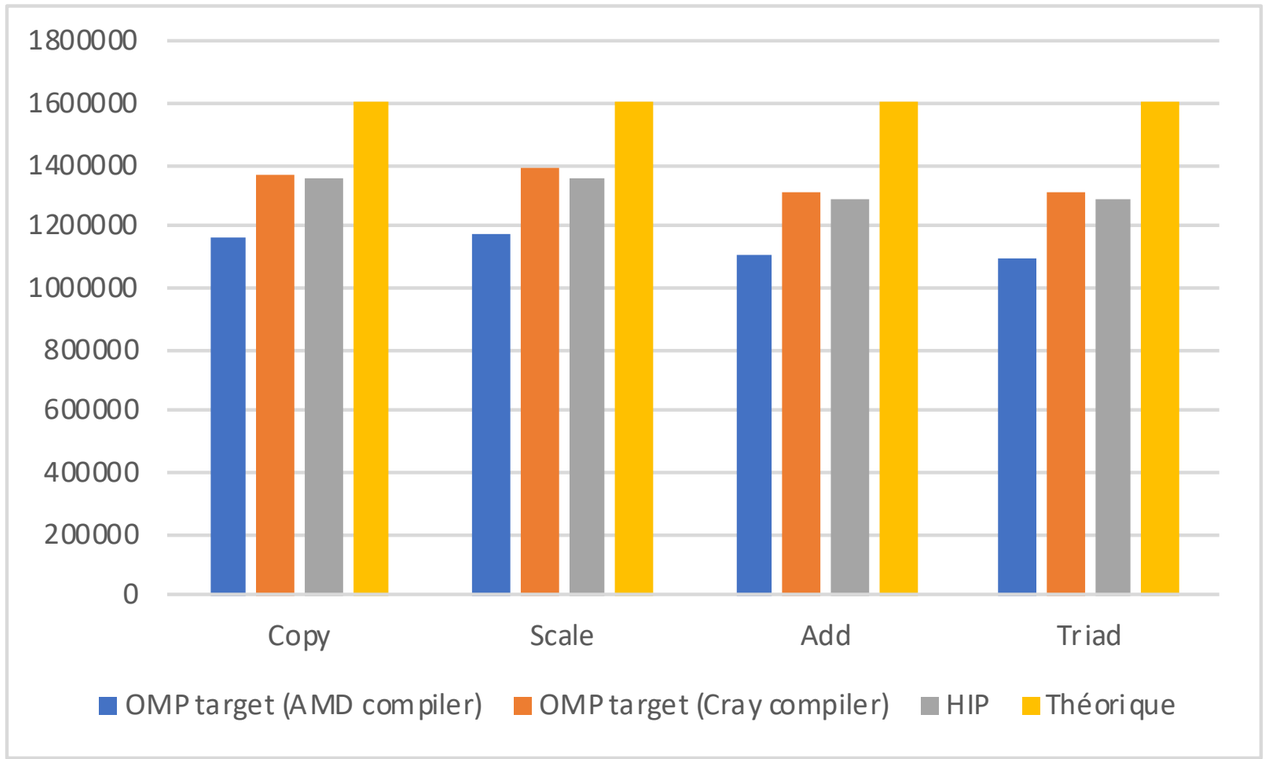
Programmé par HIP (cas Triad ci-dessous)

```
hipLaunchKernelGGL (triad_on_device, numBlocks_x, blockSize_x, 0, 0,
                    Nelem_x, scalar, a_d, b_d, c_d);

hipDeviceSynchronize();
```

```
__global__ void triad_on_device (const int Nelem_x,
                const STREAM_TYPE scalar, STREAM_TYPE * __restrict__ a,
                STREAM_TYPE * __restrict__ b, STREAM_TYPE * __restrict__ c)
{
  int j = blockDim.x * blockIdx.x + threadIdx.x;

  if (j > Nelem_x) return;

  a[j] = b[j] + scalar*c[j];

}
```

GPU AMD : stream (code C) sur MI210

Bande passante mémoire (MB/s)



- Niveau du compilateur AMD et efforts en cours, pour OpenMP target
  - https://sc23.conference-program.com/presentation/?id=ws_waccpd106&sess=sess444

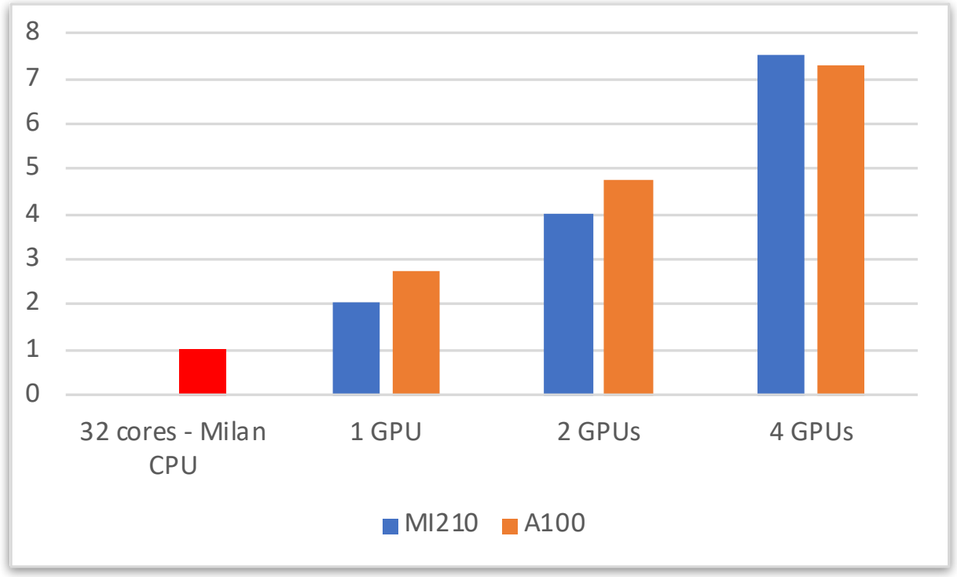
GPU AMD

Noyaux Poisson, directives, compilateur Cray

Noyau Poisson 2D (Jacobi, 4096^2)

OpenACC (FORTRAN) + MPI

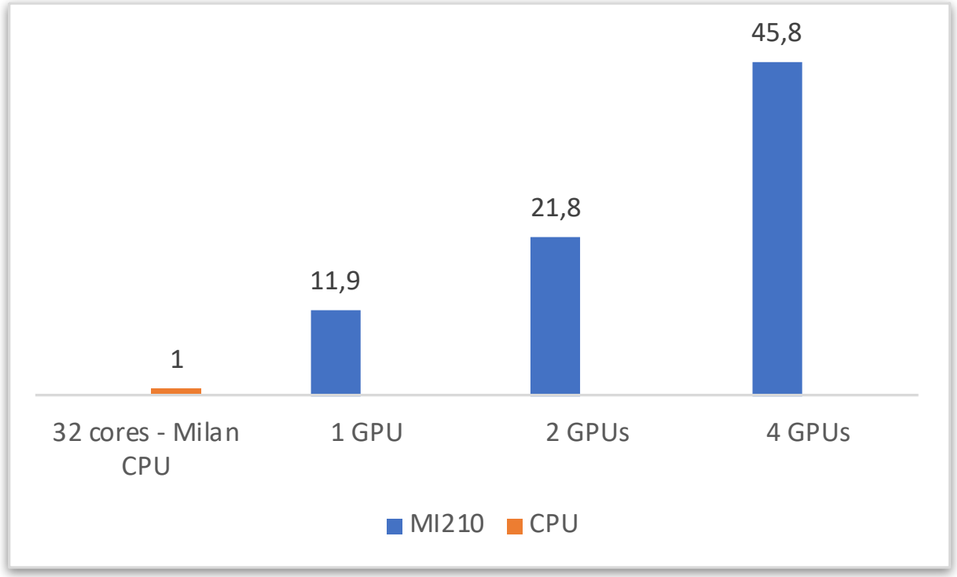
Accélération relativement
à un CPU (32 cores-Milan)



Noyau Poisson 3D (Jacobi, 512^3)

OpenMP target (C) + MPI

Accélération relativement
à un CPU (32 cores-Milan)



Agenda

Formations & actualités diverses

Site / machine	Descriptif	Dispo.
Calmip Toulouse *Turpan*	15 nœuds, interconnect 2xHDR, stockage NFS Total 1200 cœurs ARM 3Ghz et 30 GPU Nvidia A100 80 Go	disponible
Criann Rouen *Boreale*	9 nœuds vectoriels x 8 VE NEC SX Aurora Tsubasa 20B + 2 Intel 16 cœurs 2.9 GHz Interconnect IB 200 Gbps – Stockage GxFS 510 To utile	disponible
Romeo Reims *Juliet*	3 nœuds IA x 8 GPU A100 Nvidia 80 Go + 2 AMD EPYC 7663 56 cœurs 2 GHz Interconnect IB 200 Gbps + 10 Gb eth	disponible
Strasbourg	3 nœuds GPU AMD x 10 GPU AMD Mi210 64 Go + 2 AMD EPYC 7643 48 cœurs 2.3 GHz IB 100 Gb + 25 GbE Stockage 212.8 Tio nets SATA	à venir
Glicid Nantes *Phileas*	32 nœuds x 2 CPU Intel SPR 48 cœurs 2.1 GHz Interconnect IB 100 Gb + 25 Gb eth Stockage GPFS 285 To	à venir
Lille *Zen*	72 nœuds x 2 AMD EPYC Genoa 9534 64 cœurs 2.45 GHz OmniPath 100 Gb + ethernet 10 Gb Stockage BeeGFS 1 Po utile	à venir
Marseille	GPU H100	à venir
Grenoble	Openstack	à venir

MesoNET

Avancées du projet

- De nouvelles machines sont mises en production progressivement
- Documentation
  - https://www.mesonet.fr/documentation/user-documentation/
- Demande de ressources
  - https://acces.mesonet.fr/gramc-meso/
  - Y compris pour l'enseignement

Le jeudis Grey Scott - webinaires de présentation des sujets traités lors de l'école (disponibles en replay)

https://indico.in2p3.fr/event/30939/page/3642-les-jeudis-gray-scott

Formations

À venir sur 1er semestre

- Programmation parallèle avec MPI
  - 15 & 16 mai (1,5 jours)
- Python pour le HPC
  - 13 & 14 juin (2 jours)
- Inscriptions sur le site
  - https://indico.criann.fr/category/3/
- À venir : CPE, MLDE
  - Nous contacter si vous êtes intéressé
- Dans le cadre de EuroCC (CC-FR) le Criann sera partenaire de l'école d'été Gray Scott
  - Annonce prochaine des modalités

Agenda

- JCAD 2024, Bordeaux, du 4 au 6 novembre
  - https://jcad2024.sciencesconf.org/?lang=fr
  - Appel à contributions (deadline 29 mai)
    - https://jcad2024.sciencesconf.org/resource/page/id/2

- Prochain Comité Technique
  - Jeudi 20 juin envisagé

Le plateau de calcul intensif du Criann est cofinancé par la Région Normandie, l'État français et l'Union européenne (Fonds Feder).
MesoNET bénéficie d'un financement de l'Agence nationale de la recherche au titre des Investissements d'avenir.
Le Centre de Compétence EuroCC français est cofinancé par l'Union européenne et par l'État français.
Le réseau régional Syvik est cofinancé par la Région Normandie et par l'Union européenne (fonds Feder).
Le fonctionnement du Criann bénéficie du soutien de la Région Normandie.

RÉGION NORMANDIE

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

UNION EUROPÉENNE
Fonds européen de développement régional

CRIANN

Centre Régional Informatique et d'Applications Numériques de Normandie
www.criann.fr