Compte-rendu de la réunion C4 du 03 Juillet 2019 au CINES

Participants C4 :	Participants CINES :
- Sébastien Theetten (CT1) : pb connexion - Julien Bodart (CT2A) : Excusé - Florent Duchaine (CT2B) : Visio - François Sabot (CT3) : Présent - Alain Miniussi (CT4) : Visio - Virginie Grandgirard (CT5) : Présente - Rachel Nuter (CT5) : Visio - Michel Kern (CT6) : Excusé - Nicolas Floquet (CT7) : Présent - Sébastien Le Roux (CT8) : Excusé Participants GENCI : - Jean-Philippe Proux : Visio - Delphine Theodorou : Visio	- Boris Dintrans (directeur du CINES) - Eric Boyer (responsable DCI depuis 01/09/2018) - Gabriel Hautreux (DCI) - Romain Couturat (DCI) - Nicole Audiffren (DCI)

Ordre du jour :

Cette réunion C4 plus courte qu'à l'habitude s'est déroulée en visio conférence de 9h30 à 13h00.
La journée a débuté par une discussion d'une demi-heure entre membres du C4 où nous avons fait le point sur les questions et commentaires des utilisateurs. La matinée s'est ensuite poursuivie par une réunion C4/CINES/GENCI de 10h00 à 13h00. Vous trouverez ci-dessous un compte-rendu des points qu'il nous semble importants de souligner. Pour plus de détails n'hésitez pas à consulter les transparents ci-joints présentés par le CINES et GENCI lors de la réunion.
En premier lieu nous sommes heureux d'accueillir Mr François Sabot comme nouveau membre du C4 en tant que représentant du CT3 (Biologie et Santé). Nous en profitons pour rappeler que nous sommes toujours à la recherche d'un représentant pour le CT10 (Nouvelles applications et applications transverses du calcul intensif) dont le nombre d'utilisateurs devrait grossir avec l'émergence des besoins IA.

Points marquants CINES (présentation N. Audiffren, R. Couturat et G. Hautreux)

1) **Point actualité CINES** par B. Dintrans
   - Le nouveau stockage arrivera fin 2019. Attention cela aura un impact sur le stockage de niveau 2 avec la mise en place de l'espace WORKDIR. Cela concerne tout ce qui n'est pas le SCRATCHDIR.
   - Le budget concernant le renouvellement d'Occigen sera figé en Octobre 2019 pour un choix de la machine qui aura lieu fin 2020. L'installation de la machine se fera pour sa part à la fin du 1er semestre 2021. En effet, il y a une volonté de décaler l'installation de la machine pour tirer parti des meilleures opportunités technologiques sur les processeurs AMD et INTEL.
   - Le CINES va réitérer la proposition d'allocation au fil de l'eau (via les heures directeur) pour la période estivale où Occigen est généralement moins utilisée. Cette opération avait très bien fonctionné l'été dernier. Après une première analyse rapide, une vingtaine de projets A5 pourraient être bloqués prochainement par manque d'heures. Ils vont donc être

directement contactés par le CINES pour vérifier qu'ils pourraient être intéressés. Dans ce cas ils devront faire une demande au fil de l'eau pour une consommation entre le 1er et le 31 août 2019.

C4 : Le C4 demande s'il serait possible d'étendre la période du 15 juillet au 31 août.

Cette proposition est à l'étude. Une décision sera prise très rapidement.

2) Point formation et support

- Trois formations ont eu lieu au CINES depuis la précédente réunion C4 :
  - Formation « Quickstart » pour les nouveaux utilisateurs de la machine Occigen (principalement ceux de l'allocation A6). Cette formation basée sur les bonnes pratiques et les aspects sécurité a été complètement remaniée. Elle sera de nouveau proposée environ 3 semaines après l'ouverture de l'allocation A7. Même si elle est plutôt dédiée nouveaux utilisateurs elle peut très bien intéresser d'anciens car les normes et usages ont évolué.
  - La formation MPI-3 donnée par Antonio Peña (BSC) a été très intéressante et devrait être reconduite.
  - Formation QLM Quantum Learning Machine.

- Le passage à Red Hat 7.6 s'est fait au fil de l'eau sans grande difficulté signalée et sans interruption de service. Le CINES rappelle qu'il ne faut pas oublier la commande « module purge » avant tout chargement de modules pour éviter des conflits.

- Le module Tinker-HP (pour la dynamique moléculaire) a été installé suite à la venue de J-P Piquemal. Il est maintenant disponible pour la communauté avec des exemples (cf https://pubs.rsc.org/en/content/articlelanding/2018/sc/c7sc04531j#!divAbstract pour plus de détails).

3) Réponses CINES aux questions utilisateurs

a) Problème de dépassement de quota de nombre de fichiers sur le SCRATHDIR et sur le HOMEDIR. Plusieurs utilisateurs se sont plaints de limites incompatibles avec leurs besoins notamment pour l'utilisation de OpenFOAM et de Julia.

CINES : De manière générale ces limites peuvent être augmentées en justifiant le besoin auprès du support CINES.

- Par contre, pour ce qui est d'OpenFOAM la difficulté est qu'il génère un grand nombre de fichiers qui met souvent à mal le système de fichiers /scratch.
  - Le CINES rappelle que le système de fichiers actuel ne peut pas dépasser 50000 opérations de métadonnées par seconde et que 4 fois sur 5 au cours des derniers mois lorsqu'un stress du file système est apparu, il s'est agit d'un stress généré par OpenFOAM.
  - Le **CINES aimerait** donc **sensibiliser les utilisateurs au fait que** cela est particulièrement vrai pour la version 4.0 mais que **des efforts de gestions des I/O ont été faits pour les versions 5.0 et 6.0.** Le CINES encourage donc à ne plus utiliser la version 4.0 et à faire les efforts de paramétrages pour améliorer les I/O dans les versions plus récentes (cf. slides CINES 4 et 5). Le support CINES est prêt à aider les utilisateurs dans ce sens si besoin.

- Le CINES signale qu'une école d'automne CFD OpenFOAM aura lieu du 18 au 20 novembre 2019 sur le campus Paris-Saclay.

- Pour ce qui est de Julia, les équipes support sont tout à fait prêtes à l'installer sur Occigen si les utilisateurs en font la demande. Il est toujours préférable de privilégier une installation commune plutôt que l'installation en local par chaque utilisateur de ses propres logiciels.

b) *Est-ce qu'il serait possible que les utilisateurs déposent sur un site ou un directory d'Occigen les scripts qu'ils utilisent pour tel ou tel programme. Cela serait utile pour les utilisateurs débutants afin de trouver les bons paramètres de parallélisation? Actuellement, il y a un (ou deux) exemples basiques de scripts SLURM minimaux sur le site Web et c'est tout ou bien, je n'ai pas trouvé le bon endroit.*
   - Des exemples de scripts ont été mis en ligne pour un certain nombre de code de chimie. Ils sont accessibles via la commande 'module show' (Ex : 'module show gaussian/G16/B.01').
   - Cette base d'exemples peut bien sûr être augmentée mais pour que les scripts proposés soient pertinents le plus efficace serait d'avoir une personne de la communauté par logiciel pour aider à la validation des scripts.
   - N'hésitez pas à signaler au support le type d'exemples de scripts qu'il vous manquerait.

c) *La technique de container Singularity a le vent en poupe. Pour MésoNH : http://mesonh.aero.obs-mip.fr*
   cf ici http://mesonh.aero.obs-mip.fr/mesonh54/MesonhTEAMFAQ/SINGULARITY
   Nous avons commencé des tests de portabilité et performances des container Singularity+MESONH sur notre Cluster Local au Laboratoire d'Aérologie ainsi qu'au CALMIP (MésoNH centre Toulousain) 2 machines sous SLURM et réseau Infiniband. Est-il envisagé au CINES d'autoriser l'utilisation de Singularity ? Ne serait qu'en beta testeur ...
   Le CINES n'est pas réfractaire à l'idée de faire des essais d'installation de Singularity si les demandes sont bien ciblées. Il faudrait pour cela avoir plus d'informations sur l'intérêt des utilisateurs concernés car il peut y avoir des problèmes de performance ou de modes d'usages des conteneurs inenvisageables pour des raisons liées aux contraintes de sécurité. Une première démarche pourrait consister à faire des tests sur une machine autre qu'Occigen tel que Frioul.

4) Point administration système

- D'Avril à Juin peu de problèmes système à remonter si ce n'est quelques problèmes liés à la saturation des disques Lustre et à SLURM (cf slides CINES pages 6 et 7).

- Attention le système panasas ne sera plus maintenu après fin 2019. Il faudra donc de toute façon changer le HOME. Le CINES espère pouvoir le maintenir jusqu'à l'installation des nouvelles zones de stockage. Les discussions sont en cours avec ATOS pour définir une solution.

- Le CINES signale qu'il y a de plus en plus de gros jobs en queue sur OCCIGEN. Ceci est une très bonne chose car cela veut dire que la machine joue parfaitement son rôle de Tier-1. Par contre cela nécessite de vider petit à petit la machine pour pouvoir réserver le nombre de cœurs nécessaires. Les équipes CINES pensent qu'une meilleure utilisation de ce temps entre 2 gros jobs pourrait être fait en paramétrant plus finement le nombre d'heures nécessaires dans les cartes de soumission des jobs en général et des plus petits en particulier. En effet, beaucoup sont paramétrés à 24h00 par défaut. Une partie des petits jobs s'interdisent donc de passer alors qu'ils pourraient le faire sur des durées plus courtes. La mise en place d'un outil qui permettrait d'alerter l'utilisateur d'une trop grande différence entre le temps demandé et le temps réel d'exécution est actuellement à l'étude au CINES.

- Attention, plusieurs actions à venir nécessiteront l'arrêt de la machine : (i) Nettoyage des circuits d'eau des racks Broadwell, (ii) migration vers RHEL7.6 pour la partie admin et (iii) passage en SLURM 17.02. Le maximum sera fait pour grouper les interventions afin que les arrêts soient minimes.

5) Pôle veille technologique et innovation

- Pour rappel le CINES possède plusieurs prototypes au sein de la cellule de veille technologique, à la disposition des utilisateurs intéressés par ces nouvelles technologies :
    - Frioul : 54 Intel KNL BULL Sequana
    - 1 nœud Intel Optane : 12x512GB d'Optane DC + 12x16GB de DDR4
    - 1 nœud Skylake + 4 Nvidia V100
    - Environ 100 Toctets de stockage BeeGFS
- Le CINES a récemment déployé une interface IA en collaboration avec ATOS. Des tests de sécurité sont en cours. Cette plateforme est déjà ouverte aux utilisateurs et tout retour peut permettre de l'améliorer.
- Le nœud Intel Optane est testé en collaboration avec Intel. Il est actuellement possible d'utiliser la mémoire Optane DC soit en mode SSD (auquel cas la DDR4 joue le rôle de la RAM), soit en mode RAM (la DDR4 est alors utilisée comme mémoire cache) ou un mix des 2. La prochaine étape sera d'utiliser cette technologie sur cas des tests plus réalistes. Une collaboration est en cours avec le CEA via le projet de workflows in situ et in transit PaDaWan https://github.com/cea-hpc/pdwfs.

- QLM Workshop : émulateur de simulateur quantique. Un workshop a été organisé autour de cette nouvelle technologie. La QLM (pour Quantum Learning Machine) n'est pas une machine, mais un émulateur de machine quantique permettant de réaliser des tests. Un émulateur de ce type a été utilisé à distance pour les travaux pratiques.

Points marquants GENCI (présentation Jean-Philippe Proux)**:**

1) Actualités sur centres des calculs nationaux autres que CINES

✓ **TGCC** : Seconde phase Joliot-Curie prévue en 2020 :

- Ajout de partition de calculs & post-traitement IA début 2020 :
    - Partition AMD Rome 11,75 PFlop/s : 293 376 cœurs de calcul. Attention ces nœuds n'auront que 2 Goctets de RAM par cœur. Même si le dépeuplement sera possible il faut garder cette information en tête pour les applications exigeant beaucoup de mémoire par cœur.

- Post-traitement / IA 1.13 PFlop/s : 32 nœuds hybrides soit au total 128 GPU. Ceci représentera environ 1/4 de ce qui sera disponible à l'IDRIS pour la partie HPC.
- Période de Grand Challenges (GC) de décembre 2019 à fin février 2020. Par contre les GC n'auront pas lieu sur la totalité de la machine car une partie devra être ouverte dès décembre 2019 pour le call 19 PRACE.
- Des heures pourront être demandées sur cette extension dès la campagne A7 mais sur une période de 8 mois plutôt que 12.

- Ajout partition exploratoire ARM fin 2020 : 60 nœuds de calcul bi-processeur ARM Marwell ThunderX3 de prochaine génération.
  - Période de Grand Challenges (GC) de décembre 2019 à fin avril 2020.
  - Des heures pourront être demandées sur cette extension à partir de l'allocation A8 (mai 2020) pour la partie GPU.

✓ **IDRIS** : La machine Jean Zay est installée dans sa première phase. Il s'agit d'une machine hybride de 14 PFlops (60000 cœurs INTEL CSL + 1044 NVIDIA GPU).

- Le taux de candidature aux projets Grands Challenges a été très élevé avec 73 candidatures couvrant 300% des ressources. Après sélection 18 projets sur 41 ont été retenus pour la partie HPC et 12 sur 32 en IA. La période des Grandes Challenges a débuté en Juillet et se poursuivra jusqu'en Septembre avec certains projets pouvant même aller jusqu'en Avril 2020. La machine sera ouverte à l'ensemble des utilisateurs en octobre 2019 pour les allocations A6.
- Les utilisateurs d'ADA et Turing seront également migrés sur Jean Zay en Octobre 2019.
- Attention les utilisateurs ayant obtenu des heures sur Turing via l'allocation A5 ne pourront tourner que 11 mois.
- Comme déjà discuté lors du dernier C4, il existera 2 moyens d'obtenir des ressources sur la machine : (i) le mode standard via les accès réguliers pour le HPC et l'utilisation IA et (ii) un accès dynamique pour les utilisateurs qui développent des algorithmes en IA. Le site eDARI est en cours de refonte pour prendre en compte ce nouveau type d'accès.

✓ **Nouveau mode de régulation :**

- La volonté d'homogénéiser le mode de régulation (sur et sous consommation) des travaux sur les 3 centres a abouti à la décision d'une nouvelle configuration de SLURM. Cette nouvelle configuration sera en cours d'expérimentation durant cet été au TGCC et sera mis en place à l'IDRIS et au CINES dès l'allocation A7 (Novembre 2019).
- Les nouvelles règles seront les suivantes :
  - Priorité en fonction du % des heures mensuelles théoriques glissantes restant à consommer (fairshare).
  - Dépassement de l'attribution mensuelle théorique possible avec une priorité faible (exécution du job si une partie de la machine est libre : émulation des travaux bonus automatique).
  - Blocage de cette surconsommation à 25% sur la période d'allocation.

C4 : Le C4 pense que ce nouveau mode de fonctionnement devrait donner un peu plus de flexibilité aux utilisateurs.

2) Réunion besoins des utilisateurs :

Une réunion a eu lieu le 18 Avril entre les associés, les présidents des comités thématiques, les représentants des COMUT, les équipes support et les responsables de centres. Les points abordés ont été les suivants :

- Difficultés techniques rencontrées par les communautés,
- Adaptation aux nouvelles architectures ou processeurs,
- Manque de ressources humaines : besoin de formations,
- Support primordial des centres et
- Constitution d'une base de benchmarks communs inter-centre constituée de 15 à 20 applications les plus représentatives. Cette base pourra être complétée par la suite. Les porteurs de code devront être réactifs pour pouvoir fournir un ensemble de cas tests représentatifs.

3) Réponses GENCI aux questions utilisateurs :

a) *Le GENCI a décidé d'interdire l'ouverture de compte au CINES pour les utilisateurs ayant une adresse IP non française. Cela nous pose des problèmes dans le cadre des co-tutelles de thèse (doctorant partageant leur thèse entre la France et un autre pays européen). Pourquoi une telle mesure? Quelles raisons? N'est-il pas possible d'assouplir cette règle?*
Par défaut les IP non françaises ne sont plus acceptées pour l'obtention d'heures DARI. Le directeur de centre peut toutefois accepter de le faire suite à une demande justifiée mais en en acceptant la responsabilité. Cependant, la solution préconisée est que le collaborateur étranger passe par une machine française rattachée au laboratoire avec lequel il collabore et qu'il accède aux moyens de calcul via un rebond à partir de cette machine. Ceci permet que la responsabilité d'accord de connexion en France soit prise en charge par le laboratoire français associé et non par les centres de calculs nationaux. Cette mesure ne concerne pas les ouvertures de comptes prévues dans le cadre de projets Européens (PRACE xIP-WP7 ou PCP par exemple).

b) *Concernant les allocations DARI, pour les groupes de recherche ayant plusieurs sous-projets sur une même thématique, vaut-il mieux que chaque porteur de thématique dépose un projet ou bien que les personnes se regroupent pour faire une seule et grosse demande ?*
Pour être regroupé, les projets doivent avoir un lien entre eux. Il ne faut pas les regrouper seulement pour faire un projet plus gros. Il faut penser aux experts qui mettent une note globale sur le projet.

Actions CINES et C4 à prévoir d'ici la prochaine réunion :

- [C4] : Envoyer mail aux utilisateurs pour sensibiliser à l'utilisation d'OpenFOAM version 5.0 ou 6.0 plutôt 4.0. Répertorier les utilisateurs qui auraient des difficultés à changer de version
- [CINES] : Lancer appel à candidature en Septembre pour le renouvellement du C4, le mandat des membres arrivant à échéance fin 2019. L'objectif serait d'avoir la liste des nouveaux membres pour la prochaine réunion du C4 et de faire une réunion qui mixerait anciens et nouveaux arrivants.

Prochaine réunion : Entre le 4 et 15 Novembre en présentiel au CINES