



Groupe de travail "Infrastructure du Calcul Intensif"

Animé par Dominique Boutigny (CNRS / IN2P3) et Patrick Mascart (Observatoire Midi-Pyrénées)

Le calcul intensif en général et la future machine Européenne de capacité pétaflopique en particulier, ne peuvent pas être envisagés de manière isolée. La montée en puissance sans précédent des moyens de calcul intensif basés en France est une opportunité pour repenser complètement l'environnement des super-calculateurs et en faire un ensemble cohérent et efficace.

L'utilisation optimale de la pleine puissance des supercalculateurs suppose un passage à l'échelle des applications qui ne pourra se faire que progressivement et en donnant la possibilité aux développeurs de mettre au point et de tester les algorithmes et les codes sur des plateformes aisément accessibles et représentatives des architectures choisies pour la où les machines Pétaflopiques. Les mésocentres, voire certaines machines situées dans des laboratoires peuvent servir de banc de développement, mais ceci implique une véritable politique concertée de déploiement et d'exploitation de ces moyens.

Le passage à l'échelle des applications entraine dans bien des domaines une augmentation exponentielle des données produites. Il est donc nécessaire de prévoir un dimensionnement des capacités de stockage au niveau des machines pétaflopiques et / ou des calculateurs nationaux en rapport avec l'augmentation de la puissance de calcul. De même, la distribution de ces données et leur mise à disposition pour la communauté scientifique suppose le développement de très grandes bases de données. Le modèle des grilles de données, déjà utilisé dans certaine discipline est certainement une voie à explorer. La mise en commun des données est une opportunité à saisir pour structurer les communautés scientifiques et renforcer les synergies.

Dans bien des cas, les applications destinées à tourner sur les architectures de calcul haute performance nécessitent des étapes de préparation et / ou de post-processing qui s'accommodent bien des clusters de PC ou des grilles de calcul. Un rapprochement des communautés HPC, Grilles, mésocentres serait donc très utile et ne pourrait que faire progresser l'état de l'art dans chacun de ces domaines.

L'analyse des données produites par le ou les super-calculateurs va nécessiter des moyens nouveaux et de plus en plus sophistiqués : accès hyper rapide aux données, système de visualisation, analyses interactives et distribuées... Tout ces domaines devront se développer de concert avec la montée en puissance des moyens de calcul.

L'ensemble de ce qui a été discuté ci-dessus, suppose l'existence de réseaux dimensionnés de manière à pouvoir supporter la charge. Une analyse précise des besoins devra être faite tant au niveau national qu'international, avec une implication forte de RENATER.

Enfin, bien que la formation soit explicitement du domaine du groupe de travail : "Métiers du calcul intensif", il faudra veiller à ce que l'infrastructure du calcul intensif se développe de manière cohérente avec les formations mises en place.

Pistes de travail pour le groupe :

- Identification et caractérisation de "cas d'utilisation" pour lesquels l'infrastructure du calcul intensif est primordiale.
- Relations avec les mésocentres
- Relations avec les grilles de calcul et de données
- Outils à développer pour l'analyse des données
- Les réseaux pour le calcul intensif
- Relations infrastructure / formation

Organisation :

1. Constitution du groupe de travail
2. Échanges à travers une liste de diffusion
3. Réunion du groupe en septembre et fin octobre
4. Restitution en novembre

Des réunions téléphoniques intermédiaires pourront éventuellement être organisées