

# Omega

Un super-ordinateur dans un poste de travail

## La problématique

La course à la puissance de calcul est loin d'être terminée. Imagerie, météo, biochimie, simulations diverses sont aujourd'hui des thèmes consommateurs de ressources informatiques.

Chaque unité travaillant sur ces thèmes, à un moment ou à un autre, fera émerger la question de l'organisation d'une puissance de calcul suffisante pour arriver à ses fins.

Par ailleurs, l'informatique se démocratise. Partout, dans toutes les entreprises, les collectivités, les parcs informatiques augmentent, se modernisent. Les postes de travail mis actuellement sur le marché, à des prix que l'on peut qualifier de très bas, sont, du point de vue de leurs capacités de calcul, de véritables bêtes de courses : des processeurs cadencés plus de 3 GHz, de la ram d'au minimum 512 Mo, si ce n'est 1024, des disques durs de plusieurs centaines de Gigaoctets. Et qui plus est, cette puissance est largement sous-utilisée.

Parallèlement, les capacités des réseaux augmentent également de manière spectaculaire. Au niveau des LAN, le gigabit/s est devenu une norme. Des dizaines de réseaux de transport en fibre optiques sont apparus un peu partout, permettant des débits du même ordre que ce que l'on trouve sur un LAN. Aujourd'hui, sans chercher fort longtemps, on peut trouver dans chaque département un point d'accès au réseau internet avec un débit dépassant la centaine de Mo par seconde.

En d'autres termes, la puissance offerte par un cluster de calcul traditionnel, c'est à dire un certain nombre d'ordinateurs, identiques, tous au même endroit, raccordés entre eux en gigabit par seconde, est plus que largement concurrentielle par un système de mutualisation des puissances de calculs inutilisées.

## La Clusterotaxie : la fin du Gigaflop/s

Les clusters de calculs sont traditionnellement classifiés selon l'organisation du flux des données qu'ils proposent :

- clusters à mémoires partagées,
- Single System Image (openMosix, kerrighed),
- SIMD (*Single Instruction, Multiple Data*),
- MIMD (*Multiple Instructions, Multiple Data*).

Traditionnellement encore, on évalue la puissance d'un cluster au nombre d'instructions qu'il est capable d'effectuer *par secondes* exprimé en GFlop/s ou GFlops, *Giga-Floating-Operation-per-second*.

La tradition n'a pas toujours que du bon.

Certes, il est des cas où le nombre d'instructions à la seconde peut être important (calcul temps réel, comme pour l'imagerie médicale en « réalité virtuelle augmentée »). Mais dans l'immense majorité des cas, ce qui compte, c'est le nombre d'instructions effectuées à la minute, à l'heure, voire à la journée. On parlera donc de Téra- ou PétaFlop par heure (Tflop, Pflop .....).

Condor HTC introduit une nouvelle espèce de cluster : les clusters à haut-débit de calcul (*High Throughput Computing*).

Ce qui importe à Condor, c'est le nombre d'opérations effectuées sur un intervalle de temps plus significatif que la seconde, partant du constat que les applications nécessitant des fermes de calcul tournent au moins plusieurs minutes avant de produire un résultat.

En considérant cela, on s'affranchit des architectures chères et potentiellement instables qui ont été la norme du calcul parallèle jusqu'ici.

Condor utilise donc un parc hétérogène (en processeurs, en systèmes d'exploitation, en puissances) de machines. Le ciment est effectué par des serveurs décentralisés en cascade qui distribuent les tâches aux postes de travail selon les besoins de l'application et les moyens des noeuds de calculs.

## Le projet Omega

Omega est un cluster de calcul distribué. Il mutualise actuellement les capacités de calcul de deux postes de l'université de Montpellier II, Une grappe de 4 noeuds, dont un est lui-même un cluster Openmosix de 15 processeurs au CNRS à Toulouse ainsi que trois serveurs de PraKsys. Le projet a été proposé à divers organismes, qu'ils soient organismes de recherches, aménageurs, responsables de pôles de compétitivité. Vu d'aujourd'hui, rien ne s'oppose à ce que la barre des 1000 processeurs soit passée en 2007.

## Subsidiarité et suppléance

Il importe que chacun dispose de sa propre puissance de calcul. Il importe également que si sa propre puissance n'est pas suffisante, il puisse faire appel aux autres.

## Intégrité des résultats

Le cluster juge de l'intégrité du résultat fourni par chacune des entités participantes : postes de travail, groupe, grappe de groupes ... Il élimine les résultats non intègre, quelle qu'en soit la raison. Le premier condor master de la hiérarchie tague la machine ou le groupe d'où provient le résultat non intègre. Une étude statistique peut alors conduire à l'exclure.

Architecture logicielle

*Les machines raccordées*

Tout poste de travail ou serveur fait l'affaire. Les systèmes d'exploitation peuvent être Windows (95, 95, Me, 2000, XP-Pro, ...), MacOS X, Linux (Ubuntu, Debian, Fedora Core, CentOS, RedHat, SuSe,...)

*Les programmes lancés sur le cluster*

Omega supporte les programmes fonctionnant sous Windows, Linux et/ou MacOS-X

## Architecture matérielle

Les puissances de calcul peuvent provenir de :

- serveurs centraux de tout type
- les ordinateurs de bureau du parc existant : architectures x86, x86\_64, PowerPC, ...
- clusters existants (beowulf, MPI, PVM, PBS/Open-PBS, openMOSIX, etc)

Ces machines peuvent se trouver sur des réseaux (WAN ou LAN) Ethernet 10, 100 ou 1000 Mb/s.

# Conclusion

Techniquement, fédérer des puissances de calcul distribuées, indépendamment des matériels et des systèmes d'exploitations ne pose pas de problème. Le travail va plutôt être de disposer d'une interface d'administration solide assurant :

- une connaissance parfaite des capacités de chacun des noeuds de calcul, tant en puissance qu'en débit pour l'atteindre
- une mise à disposition de données statistiques pour que les utilisateurs puisse dimensionner correctement la parallélisation de leur code
- la sécurité des systèmes
- la surveillance de la réalité du principe de subsidiarité, qui consiste à laisser à chaque membre l'utilisation pleine et entière de sa propre puissance de calcul, seules les puissances non utilisées étant mutualisées
- le QOS global permettant à chacun de savoir quel type de code a tourné sur ses machines quand et pendant combien de temps.
- la conservation de la compatibilité actuelle avec le Grid 5000.
- et, bien sûr, un système quasi automatique d'intégration de nouveaux CPUs afin de minimiser le coût d'entrée dans le cluster.

Omega est un projet ambitieux, mais très léger. Il n'implique qu'un tout petit travail par machine raccordée. Il ne nécessite qu'une bonne gestion centralisée des configurations des noeuds et de leurs capacité d'échange. Les outils existent. Il ne demandent plus qu'à être mis en oeuvre !

Pour toute information complémentaire, faites un tour à [omega.praksys.net](http://omega.praksys.net) (en construction) ou [écrivez-nous](#) !

Si vous êtes étudiant et que vous cherchez un stage sur ce thème, consultez notre [offre de stage](#).