

Résumé

La montée en puissance des ordinateurs personnels offre de nouvelles perspectives à la science. Dans le cadre de la modélisation moléculaire, cette puissance se traduit par la possibilité de traiter des systèmes toujours plus grands, plus précis et donc, de plus en plus proches de la réalité. Depuis la création du projet Beowulf, les clusters de calcul composés d'ordinateurs personnels se sont progressivement imposés face aux supercalculateurs dont l'obsolescence est prématurée et qui, de surcroît, se montrent bien plus onéreux. Tout comme le matériel, les diverses technologies logicielles connaissent un essor considérable qui permet, entre autre, d'aborder les problèmes de manière différente : la création d'une application client serveur peut se faire au moyen de technologies évoluées offrant la possibilité au développeur de se focaliser exclusivement sur la logique métier de l'application en faisant abstraction des divers aspects techniques relatifs aux communications. Parmi ces technologies, on peut citer les services Web à la popularité croissante. Les gestionnaires de clusters les plus répandus tels que PBS ne font pas usage de ces technologies récentes car ils ont été créés avant l'avènement de celles-ci. Si au fil des versions, ces gestionnaires se sont enrichis en termes de fonctionnalités, leur infrastructure de base est restée inchangée pour des raisons de compatibilité.

Les besoins en matière de ressources évoluent à un rythme plus élevé que l'augmentation de vitesse des processeurs : quelque soit la puissance d'une infrastructure de calcul, elle se révélera insuffisante au bout de quelques années. Les besoins ne se contentent pas d'augmenter, ils se diversifient également et peuvent se complexifier : les fonctionnalités offertes par un gestionnaire de clusters peuvent alors se révéler insuffisantes. Des logiciels tels que PBS peuvent déléguer certains aspects de la gestion des tâches à d'autres logiciels tels que MAUI afin d'enrichir certaines fonctionnalités. Toutefois, l'ajout de nouvelles fonctionnalités non prévues, à la base, au sein du gestionnaire demeure complexe, voire impossible.

L'objectif est de développer un concept apportant un maximum de flexibilité pour la création d'un gestionnaire de clusters pouvant s'adapter à des besoins variables. Dans le cadre d'une grappe de calculateurs, il est indispensable qu'une modification quelle qu'elle soit n'implique pas un arrêt des calculs en cours ou un arrêt du service. La solution apportée repose sur l'emploi de modules constitués de composants faiblement couplés pour définir l'ensemble des fonctionnalités du gestionnaire de grappes. A chaque fonctionnalité ou besoin correspond un ou plusieurs modules et composants. Par ailleurs, un module peut être créé au moyen de tout langage permettant la manipulation de données XML-RPC ce qui offre un choix relativement large. L'emploi de divers langages permet, en cas de besoin, de tirer parti des spécificités de certains d'entre eux pour la simplification de certaines tâches telles que la manipulation de bases de données, d'images, de matrices ou autres. Une autre particularité du système est que les modules sont indépendants du gestionnaire ce qui signifie qu'une mise à jour de ce dernier ne rendra pas l'ensemble de ses modules obsolètes.

Grâce à ce concept, étendre les fonctionnalités d'un cluster peut se faire aisément. Certaines opérations répétitives effectuées par les utilisateurs peuvent être regroupées au sein d'un module paramétrable afin de devenir un service. Le travail effectué dans le cadre de cette thèse a permis la réalisation de plusieurs publications ; principalement dans le cadre de la modélisation moléculaire, mais également dans le traitement du signal. L'infrastructure créée est actuellement utilisée pour la résolution de travaux dans divers domaines tels que la création de nouveaux matériaux.