

Calcul distribué à grande échelle : un avenir prometteur dans tous les domaines scientifiques

Hélène Renard
LIP, UMR CNRS-INRIA-UCBL 5668,
École normale supérieure de Lyon, France
Helene.Renard@ens-lyon.fr

Rechercher les plus grands nombres premiers, le signal d'une vie dans la voie lactée, ou bien encore trouver la structure spatiale des protéines sont autant d'applications qui demandent une puissance de calcul importante. Cette façon de combiner la puissance d'un nombre important de calculateurs pour résoudre un problème donné s'appelle, en terme informatique, le « calcul distribué ».

Le but du calcul distribué est de répartir le travail entre plusieurs ordinateurs afin de réaliser rapidement des calculs complexes qui prendraient plusieurs mois voire plusieurs années sur une seule machine. Cette idée de parallélisation des calculs provient principalement du milieu de la recherche scientifique où les besoins en puissance de calcul et en traitement de données ont augmenté considérablement. De plus, le prix des supercalculateurs est bien souvent inabordable pour de nombreuses entreprises et laboratoires. La solution en théorie serait d'utiliser toutes les ressources informatiques disponibles dans ces entreprises ou laboratoires puisqu'une étude réalisée par Dataquest¹ a montré que seulement 5 à 20% des ressources totales étaient utilisées. Certains projets, tels que « Globus² » et « Seti@home³ », mettent en pratique le principe d'utilisation de plusieurs milliers de ressources afin d'augmenter la puissance de calcul.

C'est pourquoi, au cours de ma première année de thèse sous la direction d'Yves Robert et Frédéric Vivien, je me suis intéressée au « placement et équilibrage de charge sur grappes hétérogènes ». Le but de notre travail est de répartir des données sur un groupe d'ordinateurs reliés entre eux en tenant compte de la vitesse de calcul de chacun et de la capacité des liens de communication entre ces ordinateurs. Tout ceci dans le but de minimiser le temps total de traitement de nos données.

Nous avons donc mis l'accent sur une modélisation réaliste des communications concurrentes à l'intérieur d'un groupe d'ordinateurs. Mais ce problème a été difficile à résoudre à cause de l'hétérogénéité des vitesses de calcul de chaque ordinateur ainsi que des différentes capacités des liens de communication entre ces ordinateurs. Nous avons cependant trouvé une méthode efficace permettant d'équilibrer au mieux les données sur l'ensemble des ordinateurs afin de minimiser leur temps de traitement. Cela constitue donc une alternative prometteuse à l'usage de super-ordinateurs onéreux.

1. « Les entreprises américaines commencent à s'intéresser au grid », Michel Ktitareff, Les Echos, 04/09/2002

2. <http://wa.globus.org/>

3. <http://setiathome.ssl.berkeley.edu>