



UniGE - DataGrid CERN: le réseau du futur

C'est au Cern qu'en 1989, Tim Berners-Lee a proposé un système d'information décentralisé basé sur l'hypertexte. C'est ce système qui est devenu le web. C'est au Cern toujours que se développe actuellement le DataGrid, nouvelle technologie de gestion de données. Ce projet, financé par l'Union européenne, vise à fournir un très large accès en puissance et stockage aux équipements informatique de différents établissements, qu'aucun de ces derniers ne serait capable d'assumer individuellement. L'objectif du projet est de fournir les ressources nécessaires pour traiter des quantités considérables de données venant des expériences scientifiques dans trois disciplines différentes: physique, biologie et sciences de la terre. Dans ce dernier domaine notamment, il doit permettre de traiter les quelques 500 Gigabytes d'images générées quotidiennement par le satellite ENVISAT. En physique, le DataGrid devrait permettre dès 2007, dans le cadre du projet Egee (European Grid for European E-science), de stocker et traiter les très grandes quantités de données qui seront générées par le LHC (Large Hadron Collider).

Qu'est-ce qu'une grille?

Si le terme de grid a un sens direct en anglais (celui de « grille »), il existe un acronyme francophone plus parlant en l'occurrence: Globalisation des Ressources Informatiques et des Données (GRID). L'illustration la plus parlante des dispositifs de « grid computing » est sans doute le projet Seti@home, qui depuis 1999 distribue les signaux captés par le radio-télescope d'Arecibo (Porto-Rico) sur près de 3,5 millions de PC. Ceci permet d'utiliser la puissance de calcul de ces postes sans en affecter notablement les performances pour leur utilisation régulière.

La métaphore la plus fréquente pour illustrer le fonctionnement du Grid est celle d'un réseau électrique. Ce dernier fournit à son utilisateur les ressources nécessaires selon 3 critères:

- une interface simplifiée et standardisée;
- un système "transparent" (l'utilisateur n'a pas à connaître la complexité du réseau pour l'utiliser)
- la possibilité pour l'utilisateur de faire varier considérablement la charge d'utilisation du réseau sans en affecter les performances.

Sur la base des premiers travaux réalisés par Ian Foster et Carl Kesselman voici plus de dix ans, plusieurs projets cherchent à promouvoir les standards nécessaires à l'émergence d'une grille. Un organisme, le Global Grid Forum (GGF), composé d'intervenants publics et privés (il comprend notamment des représentants du DataGrid et du TeraGrid), cherche à développer les nouvelles normes Grid. Le GGF a notamment mis au point le développement d'un ensemble de logiciels permettant d'adapter des systèmes informatiques variés au Grid, le Globus toolkit. Outre le projet européen du DataGrid, il faut mentionner le projet américain de TeraGrid. À l'UniGE, rien n'a encore été entrepris selon Bastien Chopard (professeur adjoint au CUI), qui suit toutefois avec beaucoup d'intérêt le développement du Datagrid.

Problèmes et perspectives

Si le réseau physique sous-jacent au développement de la GRID est déjà largement constitué, l'infrastructure logicielle à la base de la grille doit encore être largement développée. Cette couche comprend notamment les processus d'identification, de cryptage et de sécurisation des données, l'interface utilisateur, les logiciels de workflow et d'arbitrage des ressources. En l'état actuel des choses, ces faiblesses, dues principalement à la coexistence de standards et systèmes incompatibles, constituent encore un frein à une diffusion plus large du projet, notamment dans le secteur commercial. Dans le domaine informatique, les disparités sont importantes, IBM étant le plus actif (tous ses serveurs sont équipés par défaut du Globus toolkit), suivi par Sun et McAfee (questions de sécurité); le fabricant de processeurs Intel collabore aussi directement avec le Cern sur le Datagrid. Quelques grandes entreprises (Monsanto, Pratt&Whitney, EDF) actives dans des domaines de recherche de pointe ont en outre des applications qui utilisent déjà des technologies Grid. Une compétition existe entre les différents projets de Grid, afin de fournir le premier une grille véritablement opérationnelle. S'il semble démarrer lentement, ce projet, vu son ampleur, suscite en effet un intérêt non-négligeable et pourrait se transformer, à l'instar du web, en un véritable tsunami dans un avenir relativement proche.

Ressources:

DataGrid Project (<http://eu-datagrid.web.cern.ch/eu-datagrid/>)
entretien avec Rosy Mondardini, Cern (Rosy.Mondardini@cern.ch)

Fred Radeff
Université de Genève
Presse Information Publications
Janvier 2004