

Sciences

PC perquisitionné pour le Décryphon P 70

Communautés

Un clic pour la cause humanitaire P 72

Nouvelle

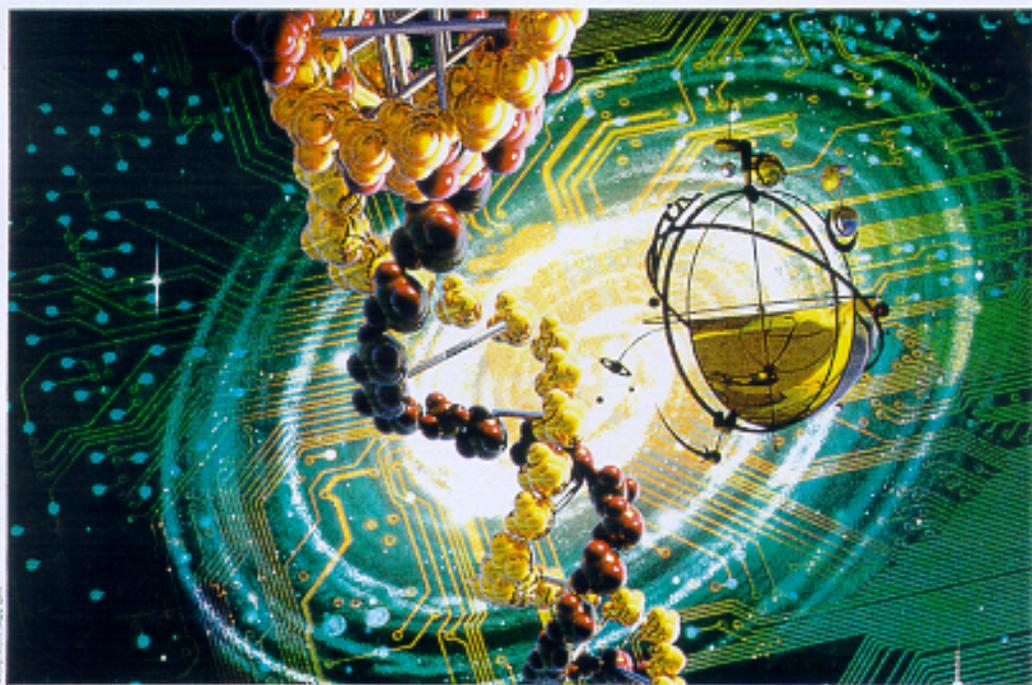
L'ordinateur de papa P 73

Contrepoints

Une ceinture informatique pour sauver nos enfants ! P 74

La recherche a besoin de votre PC

PAR CLAIRE-HÉLÈNE CASTIER



Exploiter la puissance dégagée par l'ensemble des ordinateurs de la planète ? Il ne s'agit pas de science-fiction mais de calcul distribué. Dernière application en la matière, le Décryphon, qui sera lancé officiellement lors du Téléthon, les 7 et 8 décembre prochains.

Et si votre PC participait enfin à de bonnes causes, par exemple la comparaison de plus de 500 000 protéines provenant de nombreux organismes, de la bactérie à l'homme ? C'est en tout cas l'ambitieuse mission que s'est donné le Décryphon, espérant ainsi passer à la vitesse supérieure dans la lutte contre la myopathie en définissant la première cartographie des protéomes (ensemble des protéines d'un organisme). Traduction en langage informatique : plus de 400 Go de données découpées en deux millions de paquets et à traiter en moins de trois mois.

Des projets exploitant les puissances inutilisées de nos ordinateurs (plus de 50 %), on en trouve de plus en plus. Le calcul distribué a en effet le vent en poupe dans le monde de la recherche.

Depuis plus d'un an, "Seti-at-home", du laboratoire des sciences spatiales de l'université de Berkeley, fait aussi appel à la communauté des internautes pour rechercher le signal d'une vie extraterrestre dans la Voie lactée. Plus de 3 millions d'ordinateurs analysent ainsi en parallèle, à l'aide d'un logiciel gratuit de sauvegarde d'écran, les 35 Go de données radiométriques collectées quotidiennement par le radiotélescope géant d'Arecibo.

Le CNRS, lui, participe à deux projets européens, Eurogrid et Data-Grid. Doté d'un budget de 10 millions d'euros sur trois ans, ce dernier devrait être opérationnel d'ici à la fin de 2003 avec pour premier bénéficiaire la physique des

Créer à la demande de lourds systèmes de calcul et d'analyse, à partir des ressources inexploitées de nos ordinateurs, est désormais possible grâce au grid computing. La recherche figure parmi les premiers intéressés.

particules et son dernier jouet, le grand collisionneur de hadrons LHC, à Genève. Autres disciplines concernées : la physique nucléaire pour la simulation d'essais, la biologie pour la lecture du génome humain, la météo avec ses modèles de prévision, la cryptologie et bien d'autres... Sans compter toutes les applications commerciales. La plus grosse entité de R&D privée, qui n'est autre qu'une division d'IBM (6 milliards de dollars par an), a même investi en juin dernier un milliard de dollars dans le grid computing.

Pourquoi un tel engouement ? Apparu aux États-Unis dans les années 90

pour réaliser des calculs très lourds, le calcul distribué semble à ce jour la seule réponse aux besoins des scientifiques en capacités de stockage et de calcul.

Certes, les capacités des principaux centres de calcul mis à la disposition des équipes de recherche sont parfois renforcées, par exemple l'installation à l'Iris (Institut du développement et des ressources en informatique) d'un supercalculateur d'une puissance de 1,3 T-flops (1 300 milliards d'opérations par seconde) et d'une capacité de stockage de 900 Go. Mais ce genre d'initiative reste rare et l'addition difficile à digérer (plus de 8,3 millions d'euros).

Aujourd'hui, le développement de réseaux à très haut débit en France (VTHD) et en Europe (GEANT) permet

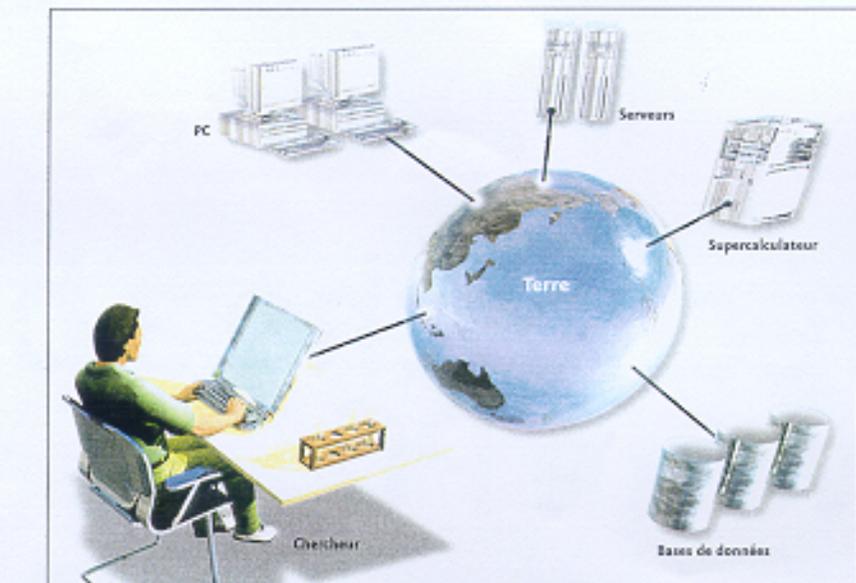
Suite page 70 ➤

► Suite de la page 69

de relier les grands centres de calcul et de découpler les ressources disponibles en fonction des besoins du moment, suivant ainsi le schéma de fonctionnement d'une centrale électrique.

Reste forcément quelques obstacles pour transformer le rêve en réalité et généraliser ce type d'opérations. A commencer par l'hétérogénéité des moyens impliqués tant au niveau des nœuds de la grille (centres de calcul, ordinateurs de laboratoires...) qu'à celui du réseau qui les relie. Ce qui manque, c'est toute l'infrastructure logicielle sur laquelle repose la grille : identification et sécurité sur la grille, logiciels d'arbitrage de ressources, de suivi d'applications, de garantie de qualité, etc. Plusieurs consortiums scientifiques se sont formés en Europe et aux États-Unis pour concevoir ces couches. Globus, la communauté de programmeurs créée par les pères fondateurs du grid computing, Carl Kesselman et Ian Foster, plâche sur les standards ouvrant l'accès à la grille de calcul distribuée comme à tout ordinateur relié à la Toile.

En France, le ministre de la Recherche, Roger-Gérard Schwarzenberg, a lancé en février dernier l'Action concertée incitative (ACI) "Globalisation des ressources informatiques et des données". Dotée d'un budget de 2,3 millions d'euros pour 2001 (reconduit en 2002 avec le même montant), cette ACI vise à financer les équipes de recherche publique travaillant sur le grid computing. "Nous avons retenu des projets d'animation, des projets logiciels proposant des mo-



dèles de grid computing et des projets pluridisciplinaires dans les domaines de la chimie, de la physique, des sciences de la Terre, de l'astronomie, de la linguistique...", explique Michel Cosnard, directeur de l'ACI.

Pas question en effet de laisser la recherche française à la traîne. Des équipes travaillent ainsi notamment à l'élaboration de modèles de calcul distribués différents de Globus. Une autre approche, qui sera développée dans le cadre de l'ACI Grid, consiste à combiner les modèles de calcul parallèle (à base de tâches et de messages) et ceux du calcul distribué (à base d'objets dis-

tribus Java et de composants logiciels Corba). Quant aux projets de calcul global tels que le Seti, ils exploitent les capacités de calcul des ordinateurs volontaires (souvent des PC) pour des applications à parallélisme trivial (sans communications entre les PC participants). L'usage de ces systèmes reste précaire et l'accès aux ressources limité : c'est le système de calcul global qui lance et gère les exécutions. Ces systèmes sont donc très fermés et les participants n'ont que la possibilité d'offrir leurs ressources de calcul.

D'un autre côté, les systèmes pair à pair tels que Napster, Gnutella, Freenet, proposent d'exploiter les ressources de stockage de PC volontaires. L'accès est généralement autorisé à tous les PC participants.

Cependant, aucun système pair à pair ne propose à ce jour une infrastructure permettant de lancer des calculs sur les PC participants. Michel Cosnard cite également une approche de plus haut niveau se basant sur le concept de FAH, dans lequel des clients accèdent à des ressources logicielles (et non plus matérielles) à distance. "L'objectif : permettre l'accès à distance à des machines à forte puissance de calcul et à grande capacité de mémoire, sans forcément être un expert du parallélisme et sans avoir à installer sur sa machine des logiciels compliqués, aux interfaces hermétiques", ajoute Michel Cosnard. Le modèle de location d'applications va peut-être enfin trouver des clients! ●

tribus Java et de composants logiciels Corba). Quant aux projets de calcul global tels que le Seti, ils exploitent les capacités de calcul des ordinateurs volontaires (souvent des PC) pour des applications à parallélisme trivial (sans communications entre les PC participants). L'usage de ces systèmes reste précaire et l'accès aux ressources limité : c'est le système de calcul global qui lance et gère les exécutions. Ces systèmes sont donc très fermés et les participants n'ont que la possibilité d'offrir leurs ressources de calcul.

D'un autre côté, les systèmes pair à pair tels que Napster, Gnutella, Freenet, proposent d'exploiter les ressources de stockage de PC volontaires. L'accès est généralement autorisé à tous les PC participants.

Cependant, aucun système pair à pair ne propose à ce jour une infrastructure permettant de lancer des calculs sur les PC participants. Michel Cosnard cite également une approche de plus haut niveau se basant sur le concept de FAH, dans lequel des clients accèdent à des ressources logicielles (et non plus matérielles) à distance. "L'objectif : permettre l'accès à distance à des machines à forte puissance de calcul et à grande capacité de mémoire, sans forcément être un expert du parallélisme et sans avoir à installer sur sa machine des logiciels compliqués, aux interfaces hermétiques", ajoute Michel Cosnard. Le modèle de location d'applications va peut-être enfin trouver des clients! ●

Références

Ian Foster et Carl Kesselman, *The Grid* : www.mtp.com/books_catalog/1-55860-475-8.asp
Présentation de l'ACI Grid : www.recherche.gouv.fr/recherche/ac/defaut.htm
Partenaires du Décryphon : www.fr.ibm.com
www.genomining.com
www.afm.genethon.fr

PC PERQUISITIONNÉ POUR LE DÉCRYPTHON

Au moins 100 000 machines moulant des petits calculs, en tâche de fond, à raison de deux heures par jour. C'est ce qu'espèrent les organisateurs du Décryphon, l'AFM (Association française contre les myopathies), Genomining et IBM (pour la maîtrise d'œuvre), en mobilisant les internautes. Il suffit dans un premier temps de s'inscrire sur le site telethon.fr en indiquant son adresse e-mail et le nombre de PC que l'on souhaite faire participer au Décryphon. L'adresse sera stockée dans une base de données jusqu'au démarrage du projet (février 2002). Les participants recevront alors un mail leur indiquant comment télécharger gratuitement l'utilitaire effectuant les calculs. Pour réaliser ces opérations et les répartir sur des milliers d'ordinateurs, un serveur IBM découpera les informations (plus de 400 Go de données) en très petits volumes de données (200 Ko). Les résultats seront renvoyés au serveur à l'occasion d'une connexion Internet



Le site du téléthon, à la pointe de la technologie, pour mieux lutter contre la maladie.

traditionnelle. Selon IBM, les calculs devraient prendre moins de trois mois grâce à une puissance de calcul virtuelle proche des 40 à 50 M-flops, soit quarante fois celle de Deep Blue, l'ordinateur qui a battu le champion d'échecs Kasparov. L'expérience prendrait environ 1 170 ans si elle n'utilisait qu'un seul PC standard. Autre moyen d'action pour la lutte contre la maladie : effectuer des dons en ligne sur le site du Téléthon.

ILS ONT DIT

Eric Molinié,
président de
l'AFM

"Parce que nous sommes une association de malades, notre objectif est de faire sauter les godets d'étranglement pour aller plus vite et plus loin vers les thérapies. Après les premières cartes du génome, réalisées à Génethon, au début des années 90, cette cartographie des protéomes est une nouvelle étape sur la route des thérapies."

Dominique Cerruti,
directeur d'IBM
Global Services



"Le grid computing est au cœur de nos préoccupations. Par notre participation au Décryphon, nous souhaitons mieux faire connaître le savoir-faire d'IBM en matière de services informatiques, notamment pour le déploiement et la mise en œuvre de cette nouvelle technologie."

William Saurin,
président de la
société
Genomining

"Grâce à l'ensemble de ces contributions, nous nous apprêtons à mettre en œuvre l'un des plus grands calculs jamais réalisés en biologie."

www.webim.com